Journal of the Royal Statistical Society Series A: Statistics in Society, 2023, **186**, 217–240
https://doi.org/10.1093/jrsssa/qnac001
Advance access publication 2 February 2023

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Original Article

Group- and individual-based approaches to health inequality: towards an integration

Iñaki Permanyer[1,2], Isaac Sasson[3] and Francisco Villavicencio[4,5,6]

[1]Centre for Demographic Studies (CED), Universitat Autònoma de Barcelona, Bellaterra, Spain
[2]ICREA, Passeig Lluís Companys 23, Barcelona, Spain
[3]Department of Sociology and Anthropology, Tel Aviv University, Tel Aviv, Israel
[4]Department of Economic, Financial and Actuarial Mathematics, University of Barcelona, Barcelona, Spain
[5]Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense, Denmark
[6]Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

Address for correspondence: Iñaki Permanyer, Carrer de Ca n'Altayó, Edifici E2, Universitat Autònoma de Barcelona, 08193 Bellaterra/Barcelona, Spain. Email: ipermanyer@ced.uab.es

Abstract

When assessing health inequalities, should one compare health outcomes across predetermined groups (e.g., race, ethnicity, socioeconomic status), or across individuals? Group-based approaches comparing group-specific means do not account for intra-group heterogeneity. Yet, traditional approaches based on additive decompositions splitting total inequality in its within- and between-group components fail to elucidate the groups' relative performance. Here, we develop a third approach based on pairwise comparisons to evaluatenot only the variability that might exist across individuals within and between groups, but also the relative performance of the different groups vis-a-vis each other—thus integrating both perspectives into a coherent framework.

Keywords: inequality decomposition, life table, lifespan inequality, mortality differentials, social groups

Introduction

The study of health inequalities has become a prominent issue in global research and policy agendas since the early 2000s. Average health attainments invariably mask the heterogeneity in the underlying distribution of health, and there is a widespread agreement on the need to go beyond the 'study of means' when studying countries' overall performance. When assessing health inequalities, the choice of the basic unit of analysis is essential: Should one compare health outcomes across predetermined groups (classified, for instance, by race, ethnicity, religion, or socioeconomic status), or across individuals? Around the early 2000s, there was a brief but intense debate on whether health differences should be based on comparisons among groups or individuals. The proponents of the group-based approach claimed that individual-based approaches to inequality 'remove equity and human rights from the public health monitoring agenda' because of their exclusive focus on individuals—rather than on the groups to which they belong (Braveman et al., 2001:678). Ignoring such characteristics, they argued, one would be unable to identify whether some groups were disadvantaged vis-a-vis the others (Braveman et al., 2000, 2001). On the contrary, supporters of the individual-based approach argued that focusing exclusively on between-group differences and ignoring the variability existing within groups would miss an important part of the story. Instead, they suggested investigating the inter-individual variations

Received: October 26, 2021. **Accepted:** October 28, 2022

in health outcomes first and then trying to determine how much of such variability would be explained by social class, race, or any other grouping one might be interested in (Gakidou et al., 2000; Murray, 2001).

Almost two decades later, no consensus seems to have been reached and current approaches fail to provide answers that are satisfactory to both sides simultaneously. On the one hand, there are countless studies comparing mean health outcomes (say, life expectancy) across groups that do not incorporate the heterogeneity that might exist within those groups into the analysis. On the other hand, the individual-based approach has gained considerable traction during the last years (Edwards & Tuljapurkar, 2005; Gakidou et al., 2000; Gakidou & King, 2002; Murray, 2001; van Raalte et al., 2018), but still fails to integrate the demands/requirements of the group-based approach (namely: to account for the performance of the different groups vis-a-vis each other). State-of-the-art methods relying on additively decomposable measures (like the variance or the Theil index) inform users on whether within-group inequality is larger or smaller than between-group inequality (Gakidou & King, 2002; Permanyer et al., 2018; Permanyer & Scholl, 2019; Seaman et al., 2019; van Raalte et al., 2012), but tell us nothing about groups' relative performance (e.g., 'Are the rich performing much better than the poor?').

In this paper, we develop a new approach to decompose a well-known class of individual-based health inequality measures that, at the same time, inform users about the groups' relative performances—thus integrating the requirements of both sides of the debate into a coherent framework. We focus on a very specific health outcome: length of life. Being alive is a precondition for any social phenomenon we might be interested in, so inequality in length of life is 'the most fundamental of all inequalities' (van Raalte et al., 2018:1002). Longevity is a crude but extremely useful, non-intrusive, and easy-to-measure health outcome that is collected worldwide on a regular basis. While the study of average longevity (i.e., life expectancy) has attracted considerable attention in demography and other social sciences for a long time, in recent years we have witnessed an upsurge of interest to look beyond the means and study the levels, trends and determinants of lifespan inequality (see, among others, Aburto et al., 2020; Colchero et al., 2016; Edwards, 2011; Edwards & Tuljapurkar, 2005; Engelman et al., 2010; Gillespie et al., 2014, Nau & Firebaugh, 2012; Sasson, 2016; Seligman et al., 2016; Smits & Monden, 2009; van Raalte et al., 2014; van Raalte et al., 2018; van Raalte & Caswell, 2013; Vaupel et al., 2011; Wilmoth & Horiuchi, 1999).

To illustrate the usefulness of our approach, we apply it to racial and ethnic inequalities in length-of-life distributions in the United States. The elimination of racial health disparities is a prominent and pressing issue in the United States (Wrigley-Field, 2020), and it has been one of the main goals in its public health policies. Black–White disparities in mortality are well documented in the literature, and so are their underlying causes of death (Firebaugh et al., 2014a; Harper et al., 2012; Hummer & Chinn, 2011). However, with few exceptions (e.g., Firebaugh et al., 2014b; Gillespie et al., 2014), past studies have focused primarily on differences in life expectancy between these two populations. With our new methodological approach, we capture distributional differences in lifespans among various population groups and quantify precisely by how much an average person from one group (White) outlives a person from another group (Black) and vice versa. We apply this method to Black and White lifespan distributions from 1970 to 2018. Information about Hispanics is also incorporated in the analysis from 2006 onwards. Even though the illustrations and empirical results presented in this paper are all based on the length of life, this new framework can be adapted to the study of other health outcome variables that are measured on a ratio scale, like length of healthy life, frailty, height, weight, blood pressure, or grip strength.

Background and motivation

There are several approaches to assess the extent of inequality in distributions of individual-based health-related outcomes in populations that are partitioned across G social groups. A very popular one is to generate group-specific averages $\mu_1, \mu_2, \ldots, \mu_G$ to compare the relative performance of the different groups vis-a-vis each other. In this way, we can easily assess whether (and to what extent) group A is performing, on average, better or worse than group B. Unfortunately, such approach ignores the variability that might exist within the different groups—which, depending on the health outcome we are working with, can be potentially large.

An alternative to the group-based perspective is to assess the influence that the population partition has on the overall levels of inter-individual inequality. In this regard, the literature on income inequality provides useful examples of indicators that weigh the variation occurring within groups vis-a-vis the variation across groups (Silber, 1999). The so-called 'additively decomposable inequality measures' can be decomposed as

$$I = I_b + I_w \tag{1}$$

where I measures overall inter-individual inequality, I_b is the inequality that would be observed if all individuals in each group (g) attained the same value (μ_g), and I_w is the average level of within-group inequality. Thus, additively decomposable inequality indices can be nicely broken down into two clearly interpretable components: (a) the between-group component (which is obtained after suppressing within-group variation by assuming all members in each group take a common value: the group mean) and (b) the within-group component (which is a weighted average of the inequality within each group). As shown by Shorrocks (1980), the class of inequality measures that are additively decomposable is quite restricted, though. For the case of relative inequality, it only comprises the one-parameter family of generalized entropy measures, such as the Theil index and the mean log deviation.[1] For absolute inequality measures, the variance is the only one that is additively decomposable. Most of these indicators have been used to study lifespan inequality (see, for instance, Edwards, 2011; Permanyer et al., 2018; Smits & Monden, 2009; van Raalte et al., 2012). In Appendix A, we explicitly show the additive decomposition formulae of these indicators when a population is partitioned in several groups.

While the additive decomposition approach can tell us whether differences between groups are more or less prominent than differences within groups, it does not provide any information about the *relative performance of the different groups vis-a-vis each other*—which is typically among the most pressing issues one is interested in when inspecting inequalities. Indeed, the notion of 'inequality' is inextricably linked to the idea that some groups are advantaged while others are disadvantaged. Usually, this kind of 'relative performance' information is inferred from the values and relative position of the group-specific means (i.e., the corresponding life expectancies)—which has been traditionally linked to the group-based approach. Yet, inferring groups' relative performance based solely on their corresponding life expectancies disregards the heterogeneity that might exist within those groups, thus missing potentially relevant information. This is illustrated in the two hypothetical scenarios comparing the age-at-death distributions for populations A and B shown in Figure 1. In the first scenario, there is almost no variability in the corresponding lifespan distributions, so ranking the two populations based on their corresponding life expectancies at birth (denoted by e_0^A and e_0^B, respectively) is not particularly misleading. In the second scenario, both e_0^A and e_0^B are kept unchanged, but there is much greater variability in population A. Now, while individuals in population B, *on average*, tend to live longer lives than those in A, a non-negligible set of individuals in population A outlive by several years *all* individuals from population B. And yet, the between-group component of standard additively decomposable inequality measures like the Theil index or the variance would give the same result in both scenarios. In such circumstances, assessing the relative performance of the two populations vis-a-vis each other solely based on the values of e_0^A and e_0^B misses an important part of the story.

In this paper, we suggest a new approach to decompose overall inequality when the population we are studying is partitioned across socially relevant groups. The basic ideas our new method builds upon are simple, and differences to existing inequality measures pertain mainly to the computation of the between-group component. Specifically, rather than assuming that individuals' attainments can be meaningfully represented by the average attainment of their group, the approach we adopt requires that the level of inequality in a given society be based on making all possible *pairwise comparisons* among individuals. In this way, a natural decomposition ensues in which overall inequality can be broken down into two clearly interpretable parts: the standard within-group component (consisting of a weighted average of inequalities within groups) and a new between-group component containing the average distances among all pairs of groups. In

[1] While the Atkinson index of inequality (Atkinson 1970) is not additively decomposable, it is a non-linear and monotonic transformation of the class of Generalized Entropy measures (see Shorrocks 1980).

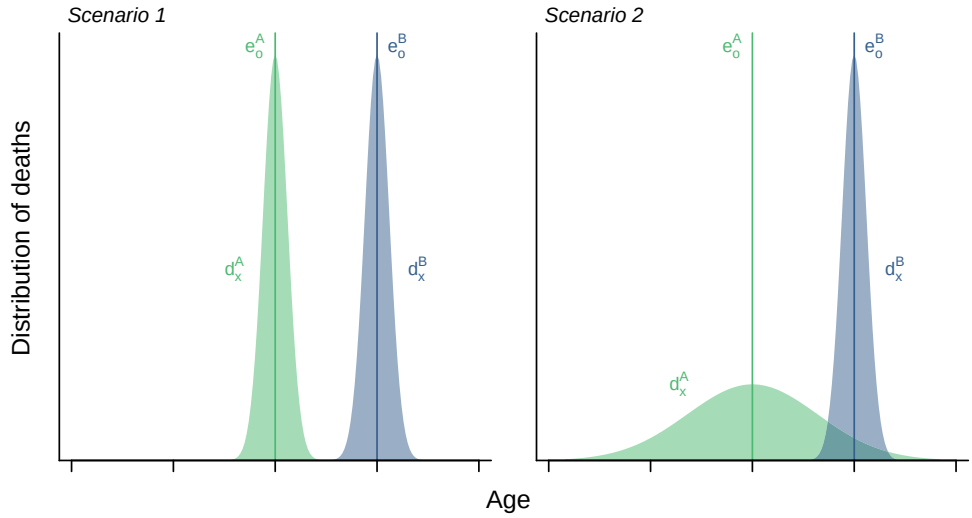

Figure 1. Two scenarios with two hypothetical age-at-death distributions d_x^A and d_x^B from two populations *A* and *B*. In both scenarios, the corresponding means of the distributions (e_0^A and e_0^B, respectively) are the same.

addition, such distances can be further decomposed into two additional components: the part of the distance involving gaps in favour of one group and the part involving gaps in the opposite direction. Thus, we are able to assess not only 'how far apart' any two groups are, but also their corresponding relative position, that is, whether one group is more or less disadvantaged with respect to the other, and vice versa. This kind of information cannot be inferred from the comparison of group-based life expectancies alone.

There is a long tradition of defining inequality measures based on all possible pairwise comparisons. The well-known Gini coefficient and other conceptually related measures (which have often been used in lifespan inequality analysis, e.g., Aburto et al., 2022; Shkolnikov et al., 2003; Smits & Monden, 2009) are constructed exactly in that way. The novelty here is not on the inequality index itself but in the proposed decomposition and the insights it can generate for a better understanding of health inequality across and within social groups.[2] Unlike current existing approaches, this new framework can explicitly assess how different groups perform vis-a-vis each other—thus integrating comparisons across individuals *and* social groups.

A new approach to decompose lifespan variation

We present a new decomposition approach that, most notably, applies to the absolute and relative Gini coefficients. The Gini coefficient was originally proposed to measure income and wealth inequality, and is one of the most popular indices in the social sciences. As thoroughly discussed by Yitzhaki & Schechtman (2013), there are several equivalent ways to define the Gini coefficient, which might depend on the context and the application. In this paper, we define it using life table notation. Life tables have been extensively used by demographers, biologists, epidemiologists, and actuarial scientists to describe the mortality and survival experience of a population (Preston et al., 2001). The literature on the measurement of lifespan inequality is based on the study of the age-at-death distributions described by such life tables, rather than by crude death counts by age, because the latter are affected by the age structure of the population(s) one is working with (e.g., crude death counts might be higher in low-mortality populations owing to the larger shares of individuals surviving to older ages, where the risk of mortality is inevitably higher). These problems are sidestepped with the use of life tables, which render comparable populations with different age structures.

[2] As discussed in further detail in Section 6, the Gini index is not additively decomposable as described in (1). However, it is amenable to other decomposition approaches, like the one proposed in this paper.

In life tables, ℓ_0 denotes the initial population at age 0, which is usually set to 1 or 100,000 and referred to as the radix. Each life table has associated an age-at-death distribution for all the reported ages, commonly denoted by $\boldsymbol{d} = \{d_0, \cdots, d_\omega\}$, where d_x is the number of individuals who die at age x and ω is the maximum possible age.[3] Note that $\sum_{x=0}^{\omega} d_x = \ell_0$, so $\boldsymbol{d}$ can be thought as the density function of the age-at-death distribution. For additional details on life tables, see for instance Preston et al., (2001).

Using this life table notation, the *absolute* Gini coefficient can be defined as

$$\Delta(\boldsymbol{d}) := \frac{1}{2\ell_0^2} \sum_{a=0}^{\omega} \sum_{b=0}^{\omega} d_a d_b |a-b| \,. \tag{2}$$

This measure of inequality considers all possible pairs of ages and calculates the average similarity/dissimilarity among them—obtaining what is oftentimes referred to as the 'average inter-individual difference' (or AID). A related measure is the *relative* Gini coefficient of the life table, given by

$$\mathcal{G}(\boldsymbol{d}) := \frac{1}{2\ell_0^2 e_0} \sum_{a=0}^{\omega} \sum_{b=0}^{\omega} d_a d_b |a-b| = \frac{\Delta(\boldsymbol{d})}{e_0}, \tag{3}$$

where e_0 is the life expectancy at birth. These two measures both have a clear interpretation: The absolute Gini coefficient measures *half* the expected age-at-death difference between two randomly chosen individuals, whereas the relative Gini coefficient is defined as the ratio between the absolute Gini and the life expectancy associated to $\boldsymbol{d}$ (that is, it puts the AID in relation to the mean of the age-at-death distribution). In the remaining of this section, we present decompositions of $\Delta(\boldsymbol{d})$, which are later extended to $\mathcal{G}(\boldsymbol{d})$ in Appendix B.

Decompositions: the two-group case

Assume a population that is partitioned in two groups, for instance females (F) and males (M), with respective population sizes n_F and n_M, such that the total population is $n = n_F + n_M$. Given the group-specific age-at-death distributions $\boldsymbol{d}_F = \{d_0^F, \cdots, d_\omega^F\}$ and $\boldsymbol{d}_M = \{d_0^M, \cdots, d_\omega^M\}$, it is easy to show (see details in Appendix C) that the overall lifespan inequality can be broken down as

$$\Delta(\boldsymbol{d}) = \mathcal{S}_F I_W^F + \mathcal{S}_M I_W^M + \mathcal{S}_{FM} I_B^{FM}, \tag{4}$$

where

$$I_W^F := \Delta(\boldsymbol{d}_F)$$

$$I_W^M := \Delta(\boldsymbol{d}_M)$$

$$I_B^{FM} := \frac{1}{2\ell_0^2} \sum_{a=0}^{\omega} \sum_{b=0}^{\omega} d_a^F d_b^M \cdot |a-b|$$

$$\mathcal{S}_F := \left(\frac{n_F}{n}\right)^2$$

$$\mathcal{S}_M := \left(\frac{n_M}{n}\right)^2$$

$$\mathcal{S}_{FM} := 2\frac{n_F n_M}{n^2}$$

[3] In this paper, we show results for ages ranging from 0 to ω, the minimum and maximum possible ages, respectively. The decomposition formulas we propose do not change if those bounds are changed to a and b, respectively (with $0 \le a < b \le \omega$).

Equation (4) has an intuitive interpretation: It decomposes overall inequality in a *within-group* ($\mathcal{S}_F I^F_W + \mathcal{S}_M I^M_W$) and a *between-group* ($\mathcal{S}_{FM} I^{FM}_B$) component. The within-group component is the weighted sum of the inequality within each of the groups (i.e., women and men), and the between-group takes into account the age-at-death differences between women and men. More specifically, I^F_W and I^M_W are the pairwise comparisons as defined in (4) within each group, whereas I^{FM}_B measures *half* of the expected difference in years of life lived between a randomly chosen woman and a randomly chosen man. The latter is a simple way of measuring how 'close' or 'far apart' the two age-at-death distributions are.[4] The coefficients $\mathcal{S}_F$, $\mathcal{S}_M$ and $\mathcal{S}_{FM}$ are the population weights of each component, satisfying $\mathcal{S}_F + \mathcal{S}_M + \mathcal{S}_{FM} = 1$.

Decompositions: the multiple-group case

Assume now that the population we are studying is partitioned in $G \geq 2$ groups. For each population subgroup g we have an age-at-death distribution $\boldsymbol{d}_g = \{d^g_0, \ \cdots, \ d^g_\omega\}$. The population size of each group is denoted as n_g. Analogously to the two-group case, $\Delta(\boldsymbol{d})$ can be decomposed as

$$\Delta(\boldsymbol{d}) = \Delta_W(\boldsymbol{d}) + \Delta_B(\boldsymbol{d}) = \sum_{g=1}^{G} \mathcal{S}_g I^g_W + \sum_{g=2}^{G} \sum_{h=1}^{g-1} \mathcal{S}_{gh} I^{gh}_B, \tag{5}$$

where

$$I^g_W := \Delta(\boldsymbol{d}_g)$$

$$I^{gh}_B := \frac{1}{2\ell^2_0} \sum_{a=0}^{\omega} \sum_{b=0}^{\omega} d^g_a d^h_b \cdot |a-b|$$

$$\mathcal{S}_g := \left(\frac{n_g}{n}\right)^2$$

$$\mathcal{S}_{gh} := 2\,\frac{n_g n_h}{n^2}$$

Equation (5) breaks down total inequality (as measured by the absolute Gini) in a within-group ($\Delta_W(\boldsymbol{d})$) and a between-group ($\Delta_B(\boldsymbol{d})$) component. Like before, the within-group component is a population-weighted average of the terms I^g_W, which measure the extent of lifespan inequality within each of the groups. The between-group component is a population-weighted average of the terms I^{gh}_B, which measure half the expected age-at-death difference between two randomly chosen individuals, one from group g and another from group h. Once again, these terms should be interpreted as functions measuring the distances that exist among pairs of length-of-life distributions.

Breaking down inequality among groups

The distance function across group pairs I^{gh}_B in the between-group component of (5) can be further decomposed into two additional sub-components. Typically, the age-at-death differences $|a-b|$ included in these terms sometimes go in favour of one group (say, g), and sometimes in favour of the other (h). Putting together the age-at-death differences to the advantage of one group on one side and the differences to the advantage of the other group on the other, we naturally obtain the following decomposition:

$$I^{gh}_B = A_{hg} + A_{gh}, \tag{6}$$

[4] There are other and even simpler ways to define distances between length-of-life distributions. For instance, to measure the distance between the corresponding life expectancies at birth: $|e^F_0 - e^M_0|$. Yet, none of them would fit into the decomposition shown in (4).

where

$$A_{hg} = \frac{1}{2\ell_0^2} \sum_{a=0}^{\omega} d_a^g \left(\sum_{b=a}^{\omega} (b-a) \cdot d_b^h \right) \tag{7}$$

and

$$A_{gh} = \frac{1}{2\ell_0^2} \sum_{b=0}^{\omega} d_b^h \left(\sum_{a=b}^{\omega} (a-b) \cdot d_a^g \right). \tag{8}$$

The term A_{hg} contains all age-at-death differences in favour of group h, that is: the cases when the randomly chosen individual from group h has a longer lifespan than the randomly chosen individual from group g. When the randomly chosen individual in group g outlives the one from h, their contribution to A_{hg} is 0. The opposite happens with A_{gh}. Thus, A_{hg} measures the 'average advantage in length of life for those individuals of group h outliving those of g'. This natural decomposition quantifies not only whether two groups are 'close' or 'far apart' from each other, but also the extent to which one of them is systematically more advantaged than the other. In other words, it speaks about the *groups' relative position*.

To illustrate how the new A_{hg}, A_{gh} terms behave, in Figure 2 we show some hypothetical examples. In panel A, one has two age-at-death distributions that are identical (i.e., $\boldsymbol{d}_g = \boldsymbol{d}_h$). When this happens, then $A_{hg} = A_{gh} = I_B^{gh}/2$, and none of the two groups can be considered to be disadvantaged with respect to the other. In real-world examples, though, age-at-death distributions are not identical, and some of them might represent a better state of affairs than others. In panels B and C, we show examples where the two age-at-death distributions ($\boldsymbol{d}_g$ and $\boldsymbol{d}_h$) partially overlap, but the individuals of group h tend to live longer than those of g. In these scenarios, $0 < A_{gh} < A_{hg}$, with higher values of A_{hg} indicating a better state of affairs for group h vis-a-vis group g. In the limit, when the two age-at-death distributions do not overlap (say, *all* individuals in group h live longer than *all* individuals in group g), then $A_{hg} = I_B^{gh}$ and $A_{gh} = 0$. This has been illustrated in panel D. Importantly, the A_{hg} and A_{gh} terms are sensitive to the differences in longevity among all individual pairs from groups g and h. In panel D, the age-at-death distributions are further apart, so the value of A_{hg} is higher than the one observed in panel C (where the two distributions are closer to each other).

Plugging the new terms of (7) and (8) into (5) yields

$$\Delta(\boldsymbol{d}) = \sum_{g=1}^{G} \mathcal{S}_g I_W^g + \sum_{g=2}^{G} \sum_{h=1}^{g-1} \mathcal{S}_{gh} (A_{gh} + A_{hg}). \tag{9}$$

This is one of the key inequality decompositions introduced in this paper. On the one hand, it contains terms describing the extent of inequality within each group ($\mathcal{S}_g I_W^g$). On the other hand, it contains terms describing not only the distances across groups, but also their relative performance vis-a-vis each other ($\mathcal{S}_{gh}(A_{gh} + A_{hg})$). Equation (9) nicely shows how total inter-individual variability ($\Delta(\boldsymbol{d})$) can be broken down into clearly interpretable pieces explaining not only how heterogeneous the different subgroups are (the within-group component), but also how these groups fare against each other (the between-group component).

Using matrix notation

In some cases, it is useful to represent all the terms in (9) using a more compact matrix notation. For that purpose, we define the *distance matrix*

$$\mathcal{M}_D := \begin{pmatrix} I_W^1 & I_B^{12} & \cdots & I_B^{1G} \\ I_B^{21} & I_W^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & I_B^{(G-1)G} \\ I_B^{G1} & \cdots & I_B^{G(G-1)} & I_W^G \end{pmatrix}, \tag{10}$$

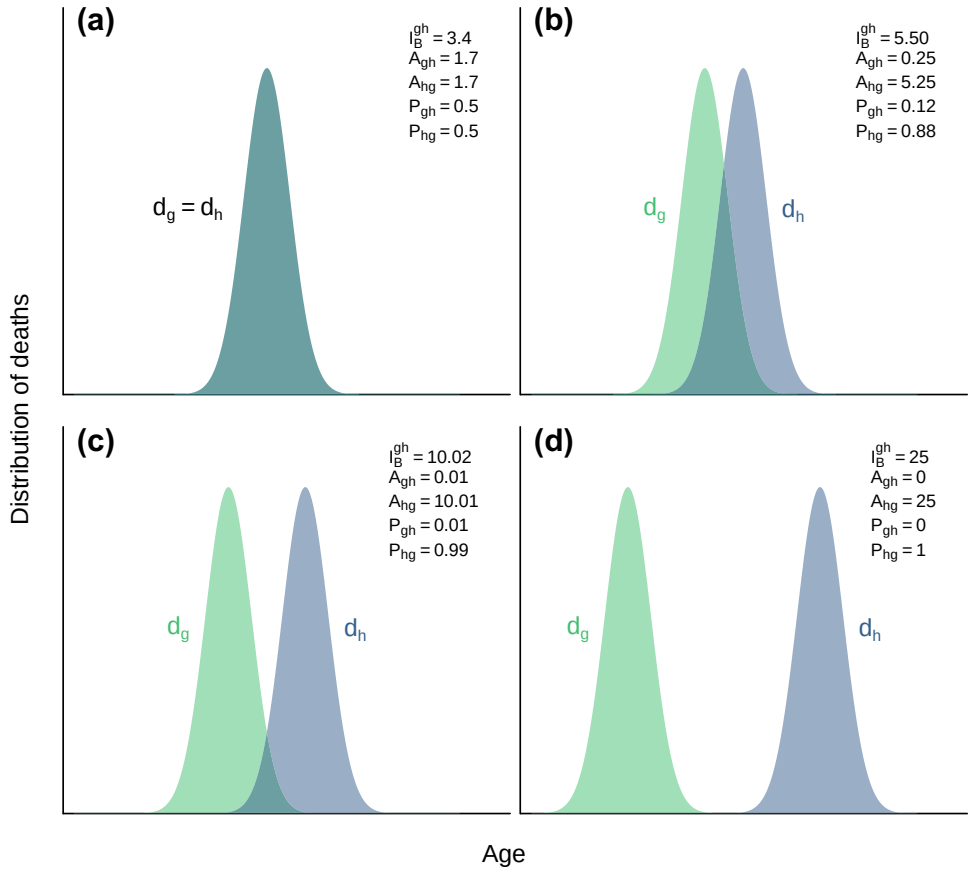

Fig. 2. Four scenarios (a–d) with two hypothetical age-at-death distributions d_x^g and d_x^h from two population subgroups g and h. In each scenario, we report the values of I_B^{gh}, A_{gh}, A_{hg}, P_{gh}, P_{hg} (see (6)–(8), (13), (14)).

which has the within-group comparison values I_W^g in the diagonal, and the between-group comparison values I_B^{gh} off the diagonal. A generic member of the distance matrix in row g and column h can thus be interpreted as 'half the expected difference in length-of-life between a randomly chosen individual from group g and a randomly chosen individual from group h'. Similarly, we define the *advantage matrix*

$$\mathcal{M}_A := \begin{pmatrix} I_W^1/2 & A_{12} & \cdots & A_{1G} \\ A_{21} & I_W^2/2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_{(G-1)G} \\ A_{G1} & \cdots & A_{G(G-1)} & I_W^G/2 \end{pmatrix} \tag{11}$$

containing (half) the within-group inequality values in the diagonal, and the 'advantage terms' A_{gh} off the diagonal. A generic member of the advantage matrix in row g and column h can thus be interpreted as 'the average advantage in length of life for those individuals from group g with respect to those from group h'.

From (6), we have that $I_B^{gh} = A_{hg} + A_{gh}$ for all pairs (g, h), which yields

$$\mathcal{M}_D = \mathcal{M}_A + \mathcal{M}_A',$$

where $\mathcal{M}_A'$ is the transposed matrix of $\mathcal{M}_A$. While both $\mathcal{M}_D$ and $\mathcal{M}_A$ are $G \times G$ matrices with positive entries, the former is symmetrical, but the latter is not. Moreover, it is possible to retrieve the

distance matrix $\mathcal{M}_D$ from the advantage matrix $\mathcal{M}_A$, but not vice versa. Using this notation, the decomposition in (9) can be re-written more compactly as

$$\Delta(\boldsymbol{d}) = \boldsymbol{p}' \times \mathcal{M}_D \times \boldsymbol{p} = \boldsymbol{p}' \times (\mathcal{M}_A + \mathcal{M}_A') \times \boldsymbol{p}, \tag{12}$$

where $\boldsymbol{p} = (p_1, \ldots, p_G)$ is the (column) vector of population shares across groups (i.e., $p_g = n_g/n$), and $\boldsymbol{p}'$ denotes its transposed.

Further extensions

In this section, we present an extension of the decomposition approach introduced in (5) and (9) with potentially many empirical applications. We show how the 'pairwise comparison approach' introduced in this paper can be used to measure the probability that individuals from a given group outlive individuals from other groups.

Over- and under-performance probability measures

The between-group component of both the absolute or relative Gini coefficients is based on the expected age-at-death differences across individuals belonging to two different groups taken at random. The pairwise comparisons-based approach can also be used to measure a similar yet fundamentally different question, like assessing the probability that individuals belonging to one group live longer than the individuals of another group. That is, rather than looking at 'how many additional years (on average) are individuals from one group expected to live with respect to individuals from the other group', we can simply ask how likely is it that individuals from one group outlive those from the other group. To answer this question, for any pair of groups $g, h \in \{1, \cdots, G\}$, we define the probability that a randomly chosen individual from group g lives longer than a randomly chosen individual from group h as

$$P_{gh} := \frac{1}{\ell_0^2} \sum_{b=0}^{\omega-1} d_b^h \left(\sum_{a=b+1}^{\omega} d_a^g \right), \tag{13}$$

and the probability that the individual in group h outlives the one in group g as

$$P_{hg} := \frac{1}{\ell_0^2} \sum_{b=0}^{\omega-1} d_b^g \left(\sum_{a=b+1}^{\omega} d_a^h \right). \tag{14}$$

Therefore, whenever $P_{gh} > P_{hg}$, individuals from group g are expected to live longer than those from group h. Note that (13) and (14) are formally equivalent to the terms A_{gh} and A_{hg} defined in (7) and (8)). The only difference is that here the age-at-death gap ($a - b$ or $b - a$) has been replaced by a value of 1 whenever one age is bigger than the other. Thus, the terms included in (13) and (14) are only sensitive to the *relative position* of longevity outcomes from groups g and h, but not to the distance that might exist among them. To illustrate this point, consider the panels C and D from Figure 2. Since in both panels one has that virtually all individuals in group h outlive those of group g, then $P_{hg} \approx 1$ and $P_{gh} \approx 0$ in both cases, even if the gaps in longevity outcomes increases when moving from C to D—a change that is captured by the terms A_{gh} and A_{hg}.

Interestingly, the P_{gh} and P_{hg} terms coincide with the outsurvival probability indicator φ introduced by Vaupel et al. (2021),—which is presented in a continuous setting and restricted to two groups only. Since there is the (relatively small) possibility of ties, the probabilities P_{gh} and P_{hg} do not add up to one. For this to happen, we must add the probability of ties to P_{gh} and P_{hg}, say

$$P_{gh} + P_{hg} + \frac{1}{\ell_0^2} \sum_{a=0}^{\omega} d_a^h d_a^g = 1. \tag{15}$$

The last term is typically very small in relative terms, so $P_{gh} + P_{hg} \approx 1$. Based on these definitions,

and mimicking the advantage matrix $\mathcal{M}_A$ form (11), we can define the *performance probability* matrix

$$\mathcal{M}_P := \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1G} \\ P_{21} & P_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & P_{(G-1)G} \\ P_{G1} & \cdots & P_{G(G-1)} & P_{GG} \end{pmatrix},$$

where P_{gh} and P_{hg} are as in (13) and (14), and P_{gg} is the probability that a randomly chosen individual from group g lives longer than a randomly chosen individual from the same group. Adapting (15) to the case where $g = h$, yields

$$P_{gg} = \frac{1}{2}\left(1 - \frac{1}{\ell_0^2}\sum_{a=0}^{\omega}(d_a^g)^2\right) \approx \frac{1}{2},$$

provided that the summatory inside the parenthesis is relatively small. Like $\mathcal{M}_A$, $\mathcal{M}_P$ is a $G \times G$ asymmetrical matrix. Its elements indicate the probabilities that the different groups in which the population is partitioned outperform one another.

Application: Length-of-life inequality in the United States, 1970–2018

In this section, we illustrate the usefulness of our approach by looking at the evolution of lifespan inequality in the United States and its decomposition across racial groups since 1970, and by Hispanic origin from 2006 onward.[5] First, we focus our attention on the US Black and White population regardless of ethnicity, treating women and men separately. This means that the *overall population* in these cases consists only of US Blacks and Whites, thus ignoring what happens to other racial groups. Next, we widen the focus to incorporate ethnicity into the analysis, comparing Hispanics (of any race) with non-Hispanic Whites and non-Hispanic Blacks, so that the overall population consists of three groups. We carried out all our analyses using the open-source statistical software R (version 4.1.1) (R Core Team, 2021).

Life expectancy among Black Americans is several years lower compared with their White counterparts. This gap, however, has varied widely over time—peaking at 7.1 years in 1989 and decreasing steadily since to 3.5 years in 2017 (Arias & Xu, 2019). The Black mortality disadvantage diminished during this time due to reduction in mortality levels (HIV, unintentional injuries) or delayed disease onset (heart disease, cancer, stroke) in this group (Firebaugh et al., 2014a; Harper et al., 2012). The lifespans of Black Americans are not only shorter on average, but also more varied compared with Whites.

The Black disadvantage in lifespan variability is largely attributed to preventable causes of death including heart disease, homicide, HIV/AIDS, and diabetes (Firebaugh et al., 2014b). It is important to note, however, that Blacks do not exhibit higher mortality from all causes of death. The opioid epidemic, for example, led to far greater loss of life years among Whites (Sasson & Hayward, 2019). Firearm and COVID-19 deaths, on the other hand, are significantly more prevalent among Blacks than Whites (Sasson & Hayward, 2019; Wrigley-Field, 2020). It is, therefore, imperative that we understand the evolution of racial inequality in age-at-death distributions, and the intricate ways in which members of each group outlive members of the other.

Ethnicity is yet another important source of heterogeneity in US mortality, in addition to race. Hispanics in particular are a growing population group with unique mortality patterns, which will

[5] In distinguishing between race and ethnicity, we follow Directive 15 of the Office of Management and Budget (OMB) concerning Race and Ethnic Standards for Federal Statistics and Administrative Reporting. According to the directive, US census and vital statistics data are reported using separate questions regarding race (Black, White, American Indian or Alaskan Native, Asian or Pacific Islander) and ethnicity (Hispanic origin/Not of Hispanic origin). While the two classifications are not mutually exclusive, US life tables since 2006 are reported for Hispanics as a whole yet for non-Hispanics by race. Thus, our three-group analysis is divided into non-Hispanic Blacks, non-Hispanics Whites, and Hispanics of all races. For the two-group comparison we use a longer time-series distinguishing between Blacks and Whites, irrespective of Hispanic origin, since 1970.

undoubtedly play a role in shaping US mortality in the future. Estimates of life expectancy by Hispanic origin are relatively recent, starting in 2006 (Arias, 2010). In spite of their lower socioeconomic status, on average, Hispanics consistently exhibit higher life expectancy and lower lifespan variability than non-Hispanic Whites (Lariscy et al., 2015, 2016). Debates about the Hispanic mortality advantage are ongoing, but it appears to reflect a combination of selection mechanisms, both inward and outward, as well as favourable health behaviours such as lower smoking prevalence (Fenelon, 2013; Hummer et al., 2007; Palloni & Arias, 2004).[6]

Both race and ethnicity are important determinants of mortality in the United States and will continue to be so in the coming years. We illustrate our method, and what new insights can be gained from it, using these two case studies. First, using US decennial (1970–2000) and annual (2001–2017) life tables (NCHS, 2021), we decompose Black–White mortality inequalities by gender, and show how they have evolved over time (Figures 3, 4 and 5). Second, to extend our method to three groups we add to our analysis Hispanic origin. From 2006 to 2018, we explore the probabilities that non-Hispanic Blacks, non-Hispanic Whites, and Hispanic Americans outlive each other (Figure 6). While all analyses are based on official US life tables, there have been important changes in racial and ethnic categorization over the years. Since 2000, the US Census permits mixed race identification whereas the National Vital Statistics System does not. In order to obtain consistent estimates of population shares throughout the study period, we relied on the NCHS bridged-population estimates from 2000 onward (US DHHS, 2020). Life tables have also changed slightly over the years. The US decennial life tables are given in single years of age all the way through 110, whereas the annual life tables from 2001 and onward end at age 100. In order to make them comparable we extended the latter to age 109+ using P-splines, a common method for smoothing mortality rates (Camarda, 2012).

Assessing longevity inequalities in the United States

Figure 3 shows the age-at-death distribution for Blacks and Whites by gender in the United States in 1970 and 2017. The vertical lines indicate the corresponding life expectancies at birth. In all cases, life expectancies are higher for Whites than for Blacks, but the gap has shrunk over time, a result that coheres with Harper et al. (2014), and Firebaugh et al. (2014a),. Beyond means, a visual inspection of the shape of the age-at-death distributions suggests that: (a) lifespan inequality has decreased over time; (b) lifespan inequality among Whites is lower than that of Blacks; and (c) lifespan inequality among men is higher than among women.

These observations are confirmed in Figure 4, which shows the trends of the Theil and the Gini indices of lifespan inequality since 1970 for the US Black and White population, and for women and men separately. Most importantly for the purposes of this paper, Figure 4 also depicts the decomposition of these indices in their basic constituents.

The Theil index is representative of the current approaches to assess the influence of population partitions on overall inequality levels (e.g., Gakidou & King, 2002; Permanyer et al., 2018; Permanyer & Scholl, 2019; Seaman et al., 2019; van Raalte et al., 2012). The values of the overall Theil index for US females decline from 0.041 in 1970 to 0.024 in 2017. For US males, it goes down from 0.056 in 1970 to 0.033 in 2017. In both cases, the overall Theil index has been declining from 1970 until 2000 and remained relatively stable from 2000 onwards. What about the additive decomposition of the Theil index in its within-group and between-group components ($T = T_w + T_b$; see Appendix A)? As shown in the graphs, the between-group component is extremely small, and only explains around 1% of total inequality in lifespans among the US Black and White population.

The right panels of Figure 4 show the results corresponding to the absolute Gini coefficient and its different decompositions. The values of the overall Gini index (i.e., the Gini index applied to the whole US Black and White population) follow a very similar path when compared to those of the Theil index. While the levels and scale are different, the trends roughly go in the same direction: we also observe a decline in inequalities from 1970 to 2000, followed by a period of relative

[6] The healthy immigrant effect is a well-established phenomenon, in the Unite States and elsewhere, in which immigrants are healthier on average than the receiving population because they are selected based on their health status, socioeconomic status, human capital, or character (Kennedy et al., 2015). By contrast, 'salmon bias' is a form of outward selection in which less healthy immigrants return to their country of origin as their health deteriorates, eventually dying abroad (Abraido-Lanza et al., 1999). US mortality data are census-unlinked, meaning that that numerator (death counts) and denominator (person-years) are estimated from different sources. Thus, returning migrants might be included in the at-risk population but their deaths might be omitted from the numerator if they had died outside the country, thus resulting in downward bias of mortality rates for that group.

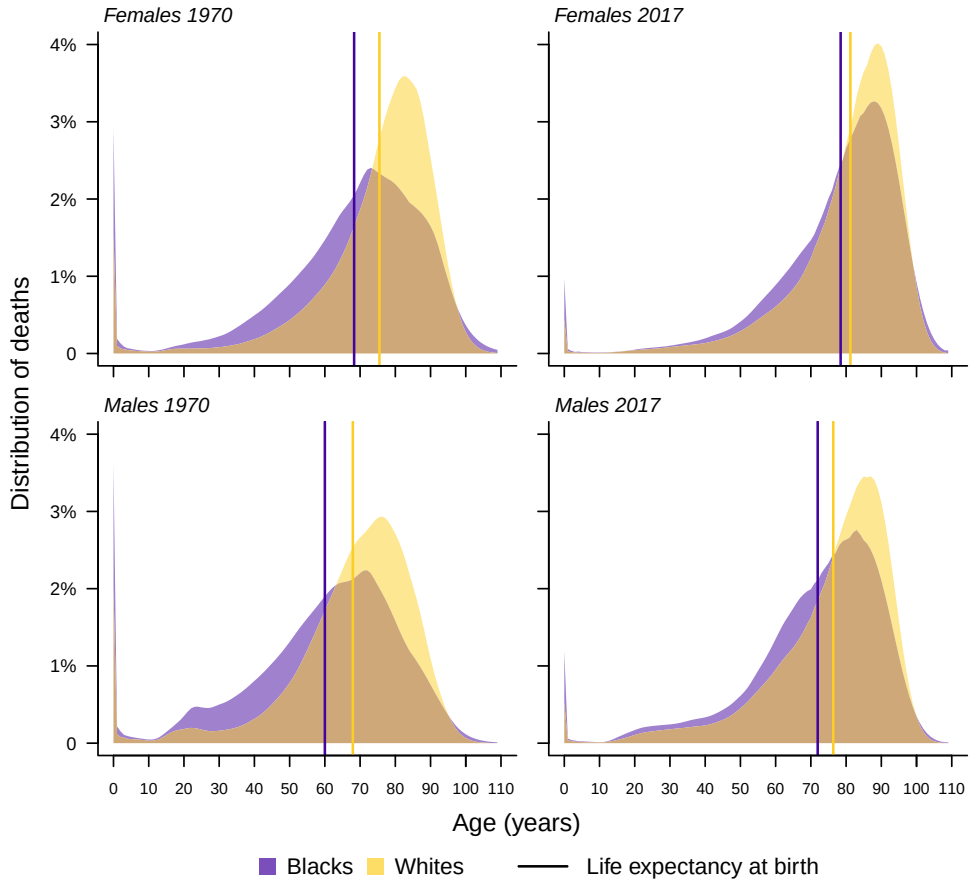

Fig. 3. Age-at-death distribution for Black and White females and males in the United States, 1970 (left panels) and 2017 (right panels). Source: NCHS (2021).

stability until 2017. For females, the overall Gini coefficient goes down from 9.148 in 1970 to 7.965 in 2017, while for males it goes down from 10.050 to 9.087. Interestingly, these levels of overall lifespan inequality can be broken down into clearly interpretable components. Take the value of 10.050 observed for males in 1970. That year, the population share for Whites was $p_1 = 0.891$, whereas Blacks represented $p_2 = 0.109$. Following (12), one has that

$$\begin{aligned}10.050 &= \boldsymbol{p}' \times \tilde{\mathcal{M}}_D \times \boldsymbol{p} = \boldsymbol{p}' \times (\tilde{\mathcal{M}}_A + \tilde{\mathcal{M}}'_A) \times \boldsymbol{p} \\ &= \begin{pmatrix}0.891 & 0.109\end{pmatrix}\begin{pmatrix}9.685 & 11.416 \\ 11.416 & 12.081\end{pmatrix}\begin{pmatrix}0.891 \\ 0.109\end{pmatrix} = \\ &= \begin{pmatrix}0.891 & 0.109\end{pmatrix}\left(\begin{pmatrix}\frac{9.685}{2} & 7.695 \\ 3.721 & \frac{12.081}{2}\end{pmatrix} + \begin{pmatrix}\frac{9.685}{2} & 3.721 \\ 7.695 & \frac{12.081}{2}\end{pmatrix}\right)\begin{pmatrix}0.891 \\ 0.109\end{pmatrix}.\end{aligned}$$

As shown in these identities, the overall Gini coefficient is a combination of inequalities among Blacks and among Whites, together with other components indicating not only the extent to which Blacks and Whites differ from each other, but also whether Blacks outperform Whites and vice versa. More specifically, we can see that the Gini coefficient for Blacks is notably larger than that of Whites (12.081 vs. 9.685). In addition, the average longevity difference between Blacks and Whites (the term I_B^{12} in (5)) is larger than the average longevity difference within Whites (11.416 vs. 9.685). In other words: the

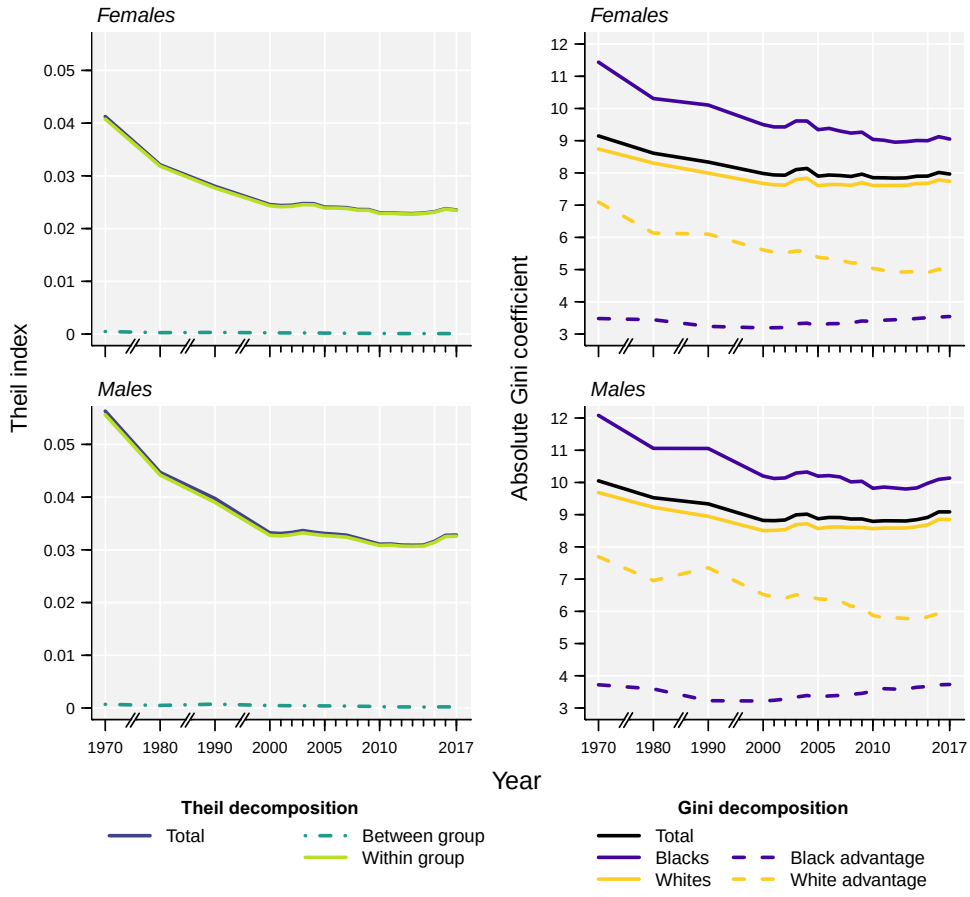

Fig. 4. Evolution of the Theil and Gini indices and their decompositions over time between 1970 and 2017 for the Black–White population in the United States. Source: Authors' elaboration based on data from NCHS (2021) and US DHHS (2020).

difference in longevity between a randomly chosen Black and a randomly chosen White is larger than the difference in longevity among two randomly chosen Whites. Breaking down the $I_B^{12} = 11.416$ term in its group-specific advantage components, we can also see that the average longevity advantage of Blacks over Whites is 3.721, while that of Whites over Blacks is 7.695 (in such a way that $11.416 = 3.721 + 7.695$). Thus, the average longevity advantage of Whites outliving Blacks ($A_{12} = 7.695$) is twice as large as the average longevity advantage of Blacks outliving Whites ($A_{21} = 3.721$). All these numbers are plotted in the bottom right panel of Figure 4 for the year 1970.

Performing now the same decomposition exercise for the overall Gini index for males in 2017, we get

$$9.087 = \begin{pmatrix} 0.851 & 0.149 \end{pmatrix} \begin{pmatrix} 8.850 & 9.673 \\ 9.673 & 10.136 \end{pmatrix} \begin{pmatrix} 0.851 \\ 0.149 \end{pmatrix} =$$

$$= \begin{pmatrix} 0.851 & 0.149 \end{pmatrix} \left(\begin{pmatrix} \frac{8.850}{2} & 5.942 \\ 3.731 & \frac{10.136}{2} \end{pmatrix} + \begin{pmatrix} \frac{8.850}{2} & 3.731 \\ 5.942 & \frac{10.136}{2} \end{pmatrix} \right) \begin{pmatrix} 0.851 \\ 0.149 \end{pmatrix}$$

As can be seen, lifespan inequality has declined both among Blacks (10.136) and among Whites (8.850), but especially among the former. We observe that the differences between Blacks and Whites have diminished as well (now $I_B^{12} = 9.673$). The decomposition of the between-group component I_B^{12} in its group-specific advantage components has also changed. In 2017, the average longevity advantage of male Whites outliving male Blacks has gone down to $A_{12} = 5.942$, while the average

longevity advantage of male Blacks outliving male Whites has remained at a similar level of $A_{21} = 3.731$. These results indicate that the age-at-death distributions of Blacks and Whites are becoming increasingly similar, with the advantage of the later over the former gradually declining over time. Very similar patterns can be identified for women as well (see upper right panel in Figure 4). Yet, this convergence in age-at-death distributions has not been a steady process throughout the study period. Periods of great progress—i.e., equalization of Black and White lifespan distributions—occurred in the 1970s and 1990s, whereas stagnation and even reversal occurred in the 1980s and early 2000s. We address some of the reasons underlying the uneven progress in the *Discussion* section.

Summing up, the three approaches discussed in this paper offer complementary information about longevity inequality trends in the United States. First, according to the group-based means approach, the life expectancy gap between Blacks and Whites declined between 1970 and 2017. Second, the decomposition of the Theil index shows that lifespan inequality declined in the United States between 1970 and 2000 and stagnated from 2000 onwards, and that differences in life expectancy between Blacks and Whites only explain 1% of the total inequality in lifespans. Finally, the new decomposition approach suggested here provides more information and permits a more thorough interpretation of the inequalities between and within the two groups: (a) lifespan inequality has been declining both among Blacks and Whites between 1970 and 2017; (b) the average longevity advantage of Whites outliving Blacks has declined over time; and (c) the average longevity advantage of Blacks outliving Whites has remained fairly constant during that period—thus resulting in a decreasing White advantage in the age-at-death distribution.

Outliving probabilities

We conclude this empirical section by analysing the time trends in the performance probabilities defined in (13) and (14). Because of data availability, we first show the results for Blacks and Whites (irrespective of Hispanic origin) from 1970 to 2017 (Figure 5). Next, we examine these performance probabilities between non-Hispanic Whites, non-Hispanic Blacks and Hispanics between 2006 and 2018 (Figure 6). As it can be seen in Figure 5, the probability that a randomly chosen White outlives a randomly chosen Black is higher than the opposite probability, but it tends to decline over time. For females, such probability was 0.604 in 1970, but declined to 0.536 in 2017. For males, such probability dropped from 0.610 to 0.567, but following a more erratic

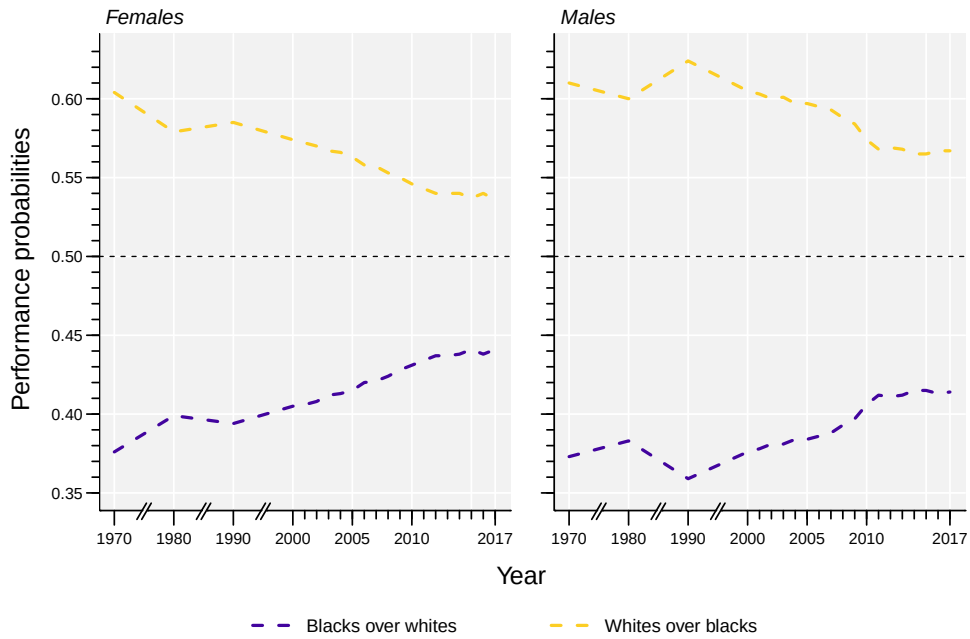

Fig. 5. Over- and under-performance probabilities for Blacks and Whites between 1970 and 2017. Source: Authors' elaboration based on data from NCHS (2021).

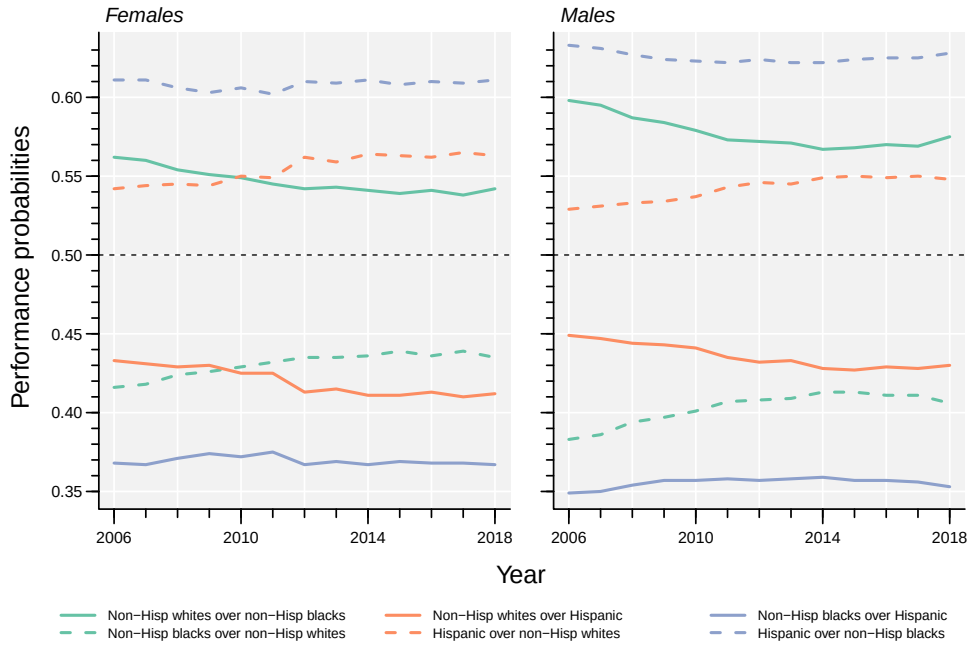

Fig. 6. Over- and under-performance probabilities for non-Hispanic Blacks, non-Hispanic Whites, and Hispanics between 2006 and 2018. Source: Authors' elaboration based on data from NCHS (2021).

trajectory over time. Symmetrically, the probability that a randomly chosen Black outlives a randomly chosen White has tended to increase over time. In 1970, such probabilities were 0.376 and 0.373 for females and males, respectively, while in 2017 they were 0.441 and 0.414.

Figure 6 depicts the performance probabilities over time for non-Hispanic Blacks, non-Hispanic Whites, and Hispanics between 2006 and 2018. We observe that the probability that Hispanics outlive non-Hispanic Whites has tended to increase over time, from 0.542 in 2006 to 0.563 in 2018 for females, and from 0.529 to 0.548 for males. In addition, the probability that Hispanics outlive non-Hispanic Blacks is considerably high, with values around 0.6 for both sexes that have remained relatively stable over time. In this three-group comparison, Hispanics perform better than non-Hispanic Blacks and non-Hispanic Whites, with the latter two becoming increasingly similar over time.

Discussion

Analysts are often interested in measuring inter-individual health inequalities in settings where the populations under study are partitioned across socially relevant groups (for instance, defined across socioeconomic status, ethnic, religious, or racial lines). A simple and very popular way of measuring health inequalities in such settings is to assess groups' performances by comparing the means of the corresponding health outcome—a group-based approach that ignores intra-group heterogeneity. Another common approach is to determine how much of the inter-individual variability can be explained by the variable that partitions the population into mutually exclusive groups. For that purpose, analysts often choose the so-called 'additively decomposable inequality measures' (like the Theil index, or the Variance) that break down total inter-individual variability in two components: (a) the between-group component (which measures the amount of inequality we would observe if we counterfactually removed intra-group variability), and (b) the within-group component (which is a weighted sum of the inequalities within the different groups)—see (1). Unfortunately, the second approach to inequality measurement does not provide any information about the groups' relative performance. In this paper, we propose a third approach that allows determining not only the health variability that might exist across individuals within and between groups, but also the relative performance of the different groups vis-a-vis each other. The decomposition method suggested here has

the advantage of integrating comparisons across individuals *and* social groups into a coherent whole and responds to the need of having a summary measure of health inequality 'which gives an overall picture of health inequalities in the population while maintaining pertinent information on [group-based] health inequalities' (Asada, 2010:3).

It is important to clarify that in this paper we focus our attention on measures of inter-individual inequality (like the Theil index, the variance or the Gini coefficient) admitting mathematically *exact* decompositions in those settings where the populations under study are partitioned across social groups. Thus, we are *not* delving into other very popular approaches to explore health inequalities that either (a) do not have the individuals as the basic units of analysis or (b) are not aiming at primarily obtaining exact inequality decompositions. Among the former, we include those approaches where the health outcome variable is only meaningful at an aggregate level (e.g., hazard ratios, relative risks, odd ratios, or mortality ratios applying to different population subgroups defined across geographical, social, cultural, or economic lines). Among the latter, we include a long and variegated list of models that aim at *predicting* individuals' health outcomes on the basis of some independent variables (e.g., income level, educational attainment, employment status, place of residence, household type, age, sex, and so on and so forth). In those settings, total inter-individual variability of the outcome variable is partly (but not fully) explained by the variability of the independent variables, and the remaining part is a residual component that the model fails to explain. Inter alia, such methods allow incorporating multiple factors to predict the outcome variable and estimate their relative role in explaining total inter-individual variability.[7] The exact inequality decomposition methods discussed in this paper offer alternative analytical perspectives that can complement the aforementioned approaches when assessing health inequalities.

Interpreting the new decompositions

The decomposition approach suggested in this paper applies to inequality measures that are based on making all pairwise comparisons across individual outcomes. Prominent members of this class include the AID index (or absolute Gini coefficient), or the (relative) Gini coefficient. While both the standard 'additive decomposition approach' of the Theil index or the variance and the new approach proposed here have a 'within-group' and a 'between-group' component (see Appendix A), the meaning and interpretation of the latter differs completely when moving from one approach to the other.[8] For standard additively decomposable measures, the between-group component has a *single* term that measures the inequality that would be observed in a hypothetical distribution suppressing variability within each of the G groups in which the population is partitioned. For the 'pairwise comparisons'-based approach proposed in this paper and described in (9) and (12), the between-group component contains the $G(G-1)$ terms off the diagonal in the advantage matrix $\mathcal{M}_A$, indicating not only how 'close' or 'far away' any two group-specific health distributions are, but also the extent to which one is (dis)advantaged with respect to the other. In this regard, the new approach is much more informative than the currently existing ones as for the groups' relative performance—something that has been clearly illustrated in the empirical section of the paper.

To further clarify the different kind of information that both approaches convey, let us consider two hypothetical scenarios. In the first one, suppose the G groups analyzed have exactly the same age-at-death distribution. In that scenario, additively decomposable measures would conclude that, since there are no differences in the group-specific life expectancies, the between-group component is zero. In contrast, all the terms off the diagonal in the distance matrix $\mathcal{M}_D$ would be a constant number $x > 0$ equal to half the expected age-at-death difference between two individuals randomly chosen from the age-at-death distribution shared across all groups. Likewise, the terms off the diagonal in the advantage matrix $\mathcal{M}_A$ would be all constant and equal to $x/2$. Suppose a second hypothetical scenario in which there is no within-group variability and all individuals

[7] In the 'exact decompositions' setting discussed in this paper, this goal can be achieved inspecting the size of the between-group components ensuing from the corresponding population partitions. That is, each of the factors (e.g., race, sex, education, or income) generates a partition of the population, and the larger the size of the corresponding between-group component, the more relevant that factor is to explain total inter-individual variability.

[8] Indeed, neither the absolute nor the relative Gini coefficients are additively decomposable in the sense described in (1). Attempts at breaking down the Gini coefficient along these lines results in a three-term decomposition, $G = G_w + G_b + R$, where G_w and G_b are the standard within-group and between-group components, and R is a residual component which is only zero in case the group-specific distributions do not overlap with each other (a circumstance that is very unlikely to occur in practice; see Lambert and Aronson 1993).

from a given group die at the same age—which naturally coincides with the group-specific life expectancy at birth and can vary across groups. In that setting, additively decomposable measures collapse into a single term, measuring the extent of inequality among the different life expectancies. Alternatively, in the pairwise comparison approach advocated in this paper, the terms off the diagonal in $\mathcal{M}_D$ would measure the distance between life expectancy pairs $|e_0^g - e_0^h|$, and the terms off the diagonal in $\mathcal{M}_A$ would be equal to either $|e_0^g - e_0^h|$ or 0, depending on whether $e_0^g > e_0^h$ or $e_0^g < e_0^h$ (e_0^g, e_0^h being the life expectancy of groups g and h, respectively).

Given the fundamentally different ways in which the two types of decompositions are defined, it is not surprising that the results they generate vary considerably. The findings reported in the empirical section of the paper are very illustrative in this regard. While the between-group component of the Theil index is almost non-existent in the US Black–White comparison (it barely explains 1% of the total variation) and tells us nothing about the relative performance of the two groups vis-a-vis each other, the terms included in the between-group component of the Gini coefficient uncover new and valuable information about the longevity advantages of Whites over Blacks, and vice versa. Inter alia, they show that while the average longevity advantage of Blacks outliving Whites has remained fairly constant, the average longevity advantage of Whites outliving Blacks has declined between 1970 and 2017. The small explanatory power of the between-group component when using additively decomposable measures like the Theil index or the variance has already been identified in previous studies (e.g., Edwards, 2011; Permanyer et al., 2018; Permanyer & Scholl, 2019; Seaman et al., 2019; Smits & Monden, 2009, van Raalte et al., 2012). This suggests that the variables that are commonly used to partition populations into groups (e.g., educational attainment, country of residence, ZIP code) are relatively weak predictors of individuals' longevity and that mortality is a highly stochastic process (Caswell, 2009). Yet, as our findings clearly illustrate, the low predictive power of these groupings on individuals' longevity does not mean that differences across groups are unimportant or non-existent; we can obtain very different insights adopting alternative measurement strategies like the one proposed here. The different approaches to inequality measurement discussed in this paper are extremely useful and can offer complementary insights to better understand past, contemporary, and prospective health dynamics.

The evolution of racial and ethnic inequalities in length of life in the United States

The US population is racially and ethnically diverse, with each group characterized by its unique mortality regime. Black Americans have long been at a disadvantage compared with the majority White population (Hummer & Chinn, 2011). Our study period, beginning in 1970, captures only the past five decades of this long history of racial inequality. Overall, this period was characterized by progress, though uneven at times, toward greater equality in lifespan distributions between Blacks and Whites in the United States. We find that among White men and women, lifespan inequality declined steadily from 1970 to 2000 but stagnated since the turn of the 21st century and even increased in recent years. Among Blacks, lifespan inequality diminished substantially in the 1970s, stagnated in the 1980s, and declined steadily during the 1990s and 2000s. Around 2010, however, this trend came to a halt and since 2014 an upsurge in lifespan inequality is observed among Blacks.

Applying the new decomposition approach proposed in this paper we can see how the composition of lifespan inequality across and within racial groups has shifted dramatically over time. In 2017, the longevity advantage of Whites over Blacks plays a much less prominent role than it did in the 1970s, while the longevity advantage of Blacks over Whites has remained relatively stable over time. Overall, these trends have resulted in a decreasing White advantage in the age-at-death distribution. These findings are consistent with prior research: the disproportionate impact of HIV/AIDS on Black communities in the 1980s leading to increased mortality in that group (Kochanek et al., 1994); the greater reductions in Black mortality from homicide, HIV, and heart disease (first among women, followed by men) during the 1990s and 2000s (Firebaugh et al., 2014a; Harper et al., 2007); and the effect of the opioid epidemic on White Americans and its lagged effect on the Black population (Alexander et al., 2018; Sasson & Hayward, 2019). The White advantage relative to Blacks, however, is far from disappearing entirely and might have worsened since 2018, the last year in our time series. COVID-19 deaths in particular have taken a far greater toll on Blacks (as well as Hispanics) in the United States, potentially reversing years of progress toward narrowing the Black–White mortality gap (Andrasfay & Goldman, 2021).

Unlike US Blacks, Hispanic hold a survival advantage relative to the non-Hispanic White population, in spite of their lower socioeconomic status (Boen & Hummer, 2019; Markides & Eschbach, 2005). We replicate this finding but show in addition that from 2006 to 2018 the probability that a randomly selected Hispanic would outlive his or her non-Hispanic White counterpart has increased. The Hispanic advantage over non-Hispanic Blacks was even greater but did not change substantially during those years. Explanations for the Hispanic survival advantage range from health behaviours to selection effects—both of healthy immigrants and unhealthy emigrants—but little attention has been given to changes in this advantage over time (Borrell & Lancet, 2012). As we have shown here, the Hispanic survival advantage is dynamic and warrants additional research to understand why it has changed over time.

Final remarks

While the decomposition approach suggested in this paper has many advantages, it is important to be aware of its limitations. Indeed, what is unarguably one of its most important strengths (the fact that it generates a detailed decomposition informing about many aspects of the health distribution) might turn out to be a disadvantage when the number of groups (G) is relatively large. Interpreting the levels and trends of the G within-group and $G(G-1)$ between-group terms shown in (9) and (12) can be daunting if, for instance, one is interested in measuring health inequalities around the world—where the population is partitioned in around $G \approx 150$ or 200 countries (see, for instance, Edwards, 2011; Permanyer & Scholl, 2019, Smits & Monden, 2009). In this line, one might further argue that when the number of groups becomes very large, the between-group component explains almost all the observed variability. While true, it is important to highlight two important issues: (a) by definition, this will happen to *any* other inequality measure or decomposition method;[9] and (b) the number of groups in which populations are partitioned is usually fixed and tends to remain unchanged across time or space. The decomposition approach proposed in this paper is both practical and very informative when the number of groups is relatively low, for example, when populations are partitioned by educational attainment, religion, ethnic groups, race, or income/wealth quartiles/quintiles.

Addressing health inequalities is becoming a top priority for public health planners all over the world. The implementation of efficient and fair policies to curb health inequalities can greatly benefit from indicators that inform planners not only about the extent of inter-individual variability, but also about the relative performance of the groups into which populations might be partitioned. This is the case of the measures proposed in this paper, which can be very useful to identify the main drivers of inequality change over time. While we have illustrated their usefulness using age-at-death distributions, the same approach can be implemented with any other health outcome measured in a cardinal scale (e.g., length of life lived in different health states (e.g., 'in good health', 'free from specific diseases or conditions'), height, frailty, grip strength, and so on), so the scope for practical applications is extraordinarily large.

Funding

The research leading to this publication was partially funded by the European Research Council under the European Union's Horizon 2020 research and innovation programme (grants 2019-CoG-864616 and 2019-AdG-884328), and the Spanish Ministry of Science and Innovation R+D LONGHEALTH project (grant PID2021-128892OB-I00). I.S. acknowledges funding from the Israel Science Foundation (grant 2532/22). F.V. acknowledges funding from the Spanish State Research Agency under the Juan de la Cierva (grant IJC2019-039144-I) and Ramón y Cajal (grant RYC2021-033979-I) programmes. The authors are grateful to Anna Turu for her support transcribing the US life tables.

Conflicts of interest: None declared.

[9] In the limit, when the number of groups approaches the number of individuals, the within-group component vanishes and the between-group one explains all the variation.

Data availability

The data and code to replicate all the results and figures presented here are publicly available for research purposes on the GitHub repository https://github.com/panchoVG/GiniDecomp.

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.

Aburto J.M., Basellini U., Baudisch A., & Villavicencio F. (2022) Drewnowski's index to measure lifespan variation: Revisiting the Gini coefficient of the life table. *Theoretical Population Biology*, *148*, 1–10.

Aburto J. M., Villavicencio F., Basellini U., Kjærgaard S., & Vaupel J. W. (2020). Dynamics of life expectancy and life span equality. *Proceedings of the National Academy of Sciences of USA*, *117*(10), 5250–5259.

Alexander M. J., Kiang M. V., & Barbieri M. (2018). Trends in Black and White opioid mortality in the United States, 1979-2015. *Epidemiology*, *29*(5), 707–715.

Andrasfay T., & Goldman N. (2021). Reductions in 2020 US life expectancy due to COVID-19 and the disproportionate impact on the Black and Latino populations. *Proceedings of the National Academy of Sciences USA*, *118*(5).

Arias E. (2010). United States life tables, 2006. *National Vital Statistics Reports*, *58*(21), 1–40.

Arias E., & Xu J. (2019). United States life tables, 2017. *National Vital Statistics Reports*, *68*(7), 1–76.

Asada Y. (2010). A summary measure of health inequalities for a pay-for-population health performance system. *Preventing Chronic Disease*, *7*(4), A72.

Atkinson A. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263.

Boen C. E., & Hummer R. A. (2019). Longer—but harder—lives? : The Hispanic health paradox and the social determinants of racial, ethnic, and immigrant–native health disparities from midlife through late life. *Journal of Health and Social Behavior*, *60*(4), 434–452.

Borrell L. N., & Lancet E. A. (2012). Race/ethnicity and all-cause mortality in US adults: Revisiting the Hispanic paradox. *American Journal of Public Health*, *102*(5), 836–843.

Braveman P., Krieger N., & Lynch J. (2000). Health inequalities and social inequalities in health. *Bulletin of the World Health Organization*, *78*(2), 232–235.

Braveman P., Starfield B., & Geiger H. J. (2001). World Health Report 2000: how it removes equity from the agenda for public health monitoring and policy. *BMJ*, *323*(7314), 678–681.

Camarda C. G. (2012). MortalitySmooth: An R package for smoothing Poisson counts with P-splines. *Journal of Statistical Software*, *50*(1), 1–24.

Caswell H. (2009). Stage, age and individual stochasticity in demography. *Oikos*, *118*(12), 1763–1782.

Colchero F., Rau R., Jones O. R., Barthold J. A., Conde D. A., Lenart A., Nemeth L., Scheuerlein A., Schoeley J., Torres C., Zarulli V., Altmann J., Brockman D. K., Bronikowski A. M., Fedigan L. M., Pusey A. E., Stoinski T. S., Strier K. B., Baudisch A., Alberts S. C., & Vaupel J. W. (2016). The emergence of longevous populations. *Proceedings of the National Academy of Sciences USA*, *113*(48), E7681–E7690.

Edwards R. (2011). Changes in world inequality in length of life: 1970-2000. *Population and Development Review*, *37*(3), 499–528.

Edwards R. D., & Tuljapurkar S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review*, *31*(4), 645–674.

Engelman M., Canudas-Romo V., & Agree E. (2010). The implications of increased survivorship for mortality variation in aging populations. *Population and Development Review*, *36*(3), 511–539.

Fenelon A. (2013). Revisiting the Hispanic mortality advantage in the United States: The role of smoking. *Social Science & Medicine*, *82*, 1–9.

Firebaugh G., Acciai F., Noah A. J., Prather C., & Nau C. (2014a). Why the racial bap in life expectancy is declining in the United States. *Demographic Research*, *31*, 975–1006.

Firebaugh G., Acciai F., Noah A. J., Prather C., & Nau C. (2014b). Why lifespans are more variable among Blacks than among Whites in the United States. *Demography*, *51*(6), 2025–2045.

Gakidou E., & King G. (2002). Measuring total health inequality: adding individual variation to group-level differences. *International Journal of Equity in Health*, *1*(1), 3.

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

Gillespie D. O. S., Trotter M. V., & Tuljapurkar S. D. (2014). Divergence in age patterns of mortality change drives international divergence in lifespan inequality. *Demography*, *51*(3), 1003–1017.

Harper S., Lynch J., Burris S., & Smith G. D. (2007). Trends in the Black-White life expectancy gap in the United States, 1983-2003. *JAMA*, *297*(11), 1224–1232.

Harper S., MacLehose R. F., & Kaufman J. S. (2014). Trends in the Black-White life expectancy gap among US states, 1990–2009. *Health Affairs*, *33*(8), 1375–1382.

Harper S., Rushani D., & Kaufman J. S. (2012). Trends in the Black-White life expectancy gap, 2003-2008. *JAMA*, *307*(21), 2257–2259.
Hummer R. A., & Chinn J. J. (2011). Race/ethnicity and U.S. adult mortality. *Du Bois Review*, *8*(1), 5–24.
Hummer R. A., Powers D. A., Pullum S. G., Gossman G. L., & Frisbie W. P. (2007). Paradox found (again): Infant mortality among the Mexican-origin population in the United States. *Demography*, *44*(3), 441–457.
Kennedy S., Kidd M. P., McDonald J. T., & Biddle N. (2015). The healthy immigrant effect: Patterns and evidence from four countries. *Journal of International Migration and Integration*, *16*(2), 317–332.
Kochanek K. D., Maurer J. D., & Rosenberg H. M. (1994). Why did Black life expectancy decline from 1984 through 1989 in the United States? *American Journal of Public Health*, *84*(6), 938–944.
Lambert P., & Aronson R. (1993). Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal*, *103*(420), 1221–1227.
Lariscy J. T., Hummer R. A., & Hayward M. D. (2015). Hispanic older adult mortality in the United States: New estimates and an assessment of factors shaping the Hispanic paradox. *Demography*, *52*(1), 1–14.
Lariscy J. T., Nau C., Firebaugh G., & Hummer R. A. (2016). Hispanic-White differences in lifespan variability in the United States. *Demography*, *53*(1), 215–239.
Markides K. S., & Eschbach K. (2005). Aging, migration, and mortality: Current status of research on the Hispanic paradox. *The Journals of Gerontology: Series B*, *60*(Special_Issue_2), S68–S75.
Murray C. J. L. (2001). Commentary: comprehensive approaches are needed for full understanding. *BMJ*, *323*-(7314), 680–681.
Nau C., & Firebaugh G. (2012). A new method for determining why length of life is more unequal in some populations than in others. *Demography*, *49*(4), 1207–1230.
NCHS – National Center of Health Statistics. (2021, February 27). *Life tables. Centers for disease control and prevention (CDC)*. https://www.cdc.gov/nchs/products/life_tables.htm
Palloni A., & Arias E. (2004). Paradox lost: Explaining the Hispanic adult mortality advantage. *Demography*, *41*(3), 385–415.
Permanyer I., & Scholl N. (2019). Global trends in lifespan inequality: 1950-2015. *PLoS ONE*, *14*(5), e0215742.
Permanyer I., Spijker J., Blanes A., & Renteria E. (2018). Longevity and lifespan variation by educational attainment in Spain: 1960–2015. *Demography*, *55*(6), 2045–2070.
Preston S. H., Heuveline P., & Guillot M. (2001). *Demography: Measuring and modeling population processes*. Blackwell.
R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.
Sasson I. (2016). Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010. *Demography*, *53*(2), 269–293.
Sasson I., & Hayward M. D. (2019). Association between educational attainment and causes of death among White and Black US adults, 2010-2017. *JAMA*, *322*(8), 756–763.
Seaman R., Riffe T., & Caswell H. (2019). Changing contribution of area-level deprivation to total variance in age at death: a population-based decomposition analysis. *BMJ Open*, *9*(3), e024952.
Seligman B., Greenberg G., & Tuljapurkar S. (2016). Equity and length of lifespan are not the same. *Proceedings of the National Academy of Sciences USA*, *113*(30), 8420–8423.
Shkolnikov V. M., Andreev M. E., & Begun A. (2003). Gini coefficient as a life table function: Computation from discrete data, decomposition of differences and empirical examples. *Demographic Research*, *8*, 305–358.
Shorrocks A. (1980). The class of additively decomposable inequality measures. *Econometrica*, *48*(3), 613–625.
Silber J. (1999). *Handbook on income inequality measurement (edited by Jacques Silber)*. Kluwer Academic Publishers.
Smits J., & Monden C. (2009). Length of life inequality around the globe. *Social Science & Medicine*, *68*(6), 1114–1123.
US DHHS – United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), and National Center for Health Statistics (NCHS) (2020) *Bridged-Race Population Estimates, United States July 1st resident population by state, county, age, sex, bridged-race, and Hispanic origin. Compiled from 1990-1999 bridged-race intercensal population estimates (released by NCHS on 7/26/2004); revised bridged-race 2000-2009 intercensal population estimates (released by NCHS on 10/26/2012); and bridged-race Vintage 2019 (2010-2019) postcensal population estimates (released by NCHS on 7/9/2020)*. http://wonder.cdc.gov/bridged-race-v2019.html
van Raalte A. A., & Caswell H. (2013). Perturbation analysis of indices of lifespan variability. *Demography*, *50*(5), 1615–1640.
van Raalte A. A., Kunst A. E., Lundberg O., Leinsalu M., Martikainen P., Artnik B., Deboosere P., Stirbu I., Wojtyniak B., & Mackenbach J. P. (2012). The contribution of educational inequalities to lifespan variation. *Population Health Metrics*, *10*(1), 3.
van Raalte A. A., Martikainen P., & Myrskylä M. (2014). Lifespan variation by occupational class: Compression or stagnation over time? *Demography*, *51*(1), 73–95.

van Raalte A. A., Sasson I., & Martikainen P. (2018). The case for monitoring life-span inequality. *Science*, *362*-(6418), 1002–1004.
Vaupel J., Bergeron-Boucher M.-P., & Kashnitsky I. (2021). Outsurvival as a measure of the inequality of lifespans between two populations. *Demographic Research*, *44*, 853–864.
Vaupel J., Zhang Z., & Van Raalte A. (2011). Life expectancy and disparity: an international comparison of life table data. *BMJ Open*, *1*(1), e000128.
Wilmoth J., & Horiuchi S. (1999). Rectangularization revisited: Variability of age at death within human populations. *Demography*, *36*(4), 475–495.
Wrigley-Field E. (2020). US racial inequality may be as deadly as COVID-19. *Proceedings of the National Academy of Sciences USA*, *117*(36), 21854–21856.
Yitzhaki S., & Schechtman E. (2013). *The Gini methodology: A primer on a statistical methodology*. Springer, New York, NY.

Appendices

Appendix A. Additively decomposable inequality indices

Here, we show the different additively decomposable inequality indices (i.e., the measures satisfying (1)) referred to in the main text and the corresponding additive decompositions. We will express them using life table notation. Thus, $\boldsymbol{d}$ represents an age-at-death distribution, ℓ_0 is the radix of the population, e_0 is the life expectancy at birth, d_x and α_x are the life table number of deaths and the average age-at-death in the age interval x to $x+1$, respectively, and ω is the maximum possible age. Assuming the population is partitioned across G groups, we denote by e_0^g the life expectancy at birth of group g.

We begin with the Theil index, which is defined as

$$T(\boldsymbol{d}) = \frac{1}{\ell_0}\sum_{x=0}^{\omega} d_x \left(\frac{\alpha_x}{e_0}\right)\log\left(\frac{\alpha_x}{e_0}\right).$$

When the population is partitioned in G groups, the Theil index can be decomposed as

$$T(\boldsymbol{d}) = T_b(\boldsymbol{d}) + T_w(\boldsymbol{d}) = \sum_{g=1}^{G}\frac{n_g}{n}\frac{e_0^g}{e_0}\log\left(\frac{e_0^g}{e_0}\right) + \sum_{g=1}^{G}\frac{n_g}{n}\frac{e_0^g}{e_0}T_g. \tag{A1}$$

The first part in (A1) is the *between-group* component, which is obtained assuming all individuals in each group die at the same age e_0^g, so there is no within-group variation. The second term is the *within-group* component, which is a weighted sum of the within-group inequalities measured by T_g, the Theil index applied to group g.

Another inequality index that is additively decomposable is the mean log deviation, which is defined as

$$L(\boldsymbol{d}) = \frac{1}{\ell_0}\sum_{x=0}^{\omega} d_x \log\left(\frac{e_0}{\alpha_x}\right).$$

When the population is partitioned in G groups, $L(\boldsymbol{d})$ can be decomposed as

$$L(\boldsymbol{d}) = L_b(\boldsymbol{d}) + L_w(\boldsymbol{d}) = \sum_{g=1}^{G}\frac{n_g}{n}\log\left(\frac{e_0}{e_0^g}\right) + \sum_{g=1}^{G}\frac{n_g}{n}L_g,$$

where L_g is the mean log deviation applied to group g. Finally, the variance is the only absolute inequality index that is additively decomposable. It is defined as

$$V(\boldsymbol{d}) = \frac{1}{\ell_0}\sum_{x=0}^{\omega} d_x(\alpha_x - e_0)^2.$$

When the population is partitioned in G groups, $V(\boldsymbol{d})$ can be decomposed as

$$V(\boldsymbol{d}) = V_b(\boldsymbol{d}) + V_w(\boldsymbol{d}) = \sum_{g=1}^{G} \frac{n_g}{n}(e_0^g - e_0)^2 + \sum_{g=1}^{G} \frac{n_g}{n} V_g,$$

where V_g is the variance applied to group g.

Appendix B. Decomposition of the relative Gini coefficient

The same approach used to decompose the absolute Gini coefficient $\Delta(\boldsymbol{d})$ can be applied to $\mathcal{G}(\boldsymbol{d})$, the relative Gini coefficient defined in (3). It is straightforward to check that (5) can be adapted to the relative context, expressed as

$$\mathcal{G}(\boldsymbol{d}) = \mathcal{G}_W(\boldsymbol{d}) + \mathcal{G}_B(\boldsymbol{d}) = \sum_{g=1}^{G} \tilde{\mathcal{S}}_g \tilde{I}_W^g + \sum_{g=2}^{G} \sum_{h=1}^{g-1} \tilde{\mathcal{S}}_{gh} \tilde{I}_B^{gh}, \tag{A2}$$

where

$$\tilde{I}_W^g := \mathcal{G}(\boldsymbol{d}_g) = \frac{I_W^g}{e_0^g}$$

$$\tilde{I}_B^{gh} := \frac{I_B^{gh}}{(e_0^g + e_0^h)/2}$$

$$\tilde{\mathcal{S}}_g := \left(\frac{n_g}{n}\right)^2 \frac{e_0^g}{e_0}$$

$$\tilde{\mathcal{S}}_{gh} := \frac{n_g n_h}{n^2} \frac{(e_0^g + e_0^h)}{e_0}$$

e_0^g, e_0^h being the life expectancy of groups g and h, respectively. The interpretation of (A2) is very similar to that of (5): The first part includes G within-group inequality terms and the second part includes $G(G-1)/2$ between-group terms. The only difference now is that the within- and between-group inequality elements are defined in relative terms (i.e., the absolute variation in the numerator is put in relation with respect to the corresponding mean/life expectancy in the denominator). Like in the previous section, the between-group terms in (A2) measuring the distance between pairs of groups can be further broken down in two components as follows:

$$\tilde{I}_B^{gh} = 2\frac{I_B^{gh}}{e_0^g + e_0^h} = 2\left(\frac{A_{gh} + A_{hg}}{e_0^g + e_0^h}\right) = \frac{2A_{gh}}{e_0^g + e_0^h} + \frac{2A_{hg}}{e_0^g + e_0^h}. \tag{A3}$$

Defining

$$\tilde{A}_{gh} := \frac{2A_{gh}}{e_0^g + e_0^h} \text{ and } \tilde{A}_{hg} := \frac{2A_{hg}}{e_0^g + e_0^h}, \tag{A4}$$

we can rewrite (A2) as

$$\mathcal{G}(\boldsymbol{d}) = \mathcal{G}_W(\boldsymbol{d}) + \mathcal{G}_B(\boldsymbol{d}) = \sum_{g=1}^{G} \tilde{\mathcal{S}}_g \tilde{I}_W^g + \sum_{g=2}^{G} \sum_{h=1}^{g-1} \tilde{\mathcal{S}}_{gh}(\tilde{A}_{gh} + \tilde{A}_{hg}), \tag{A5}$$

which is the 'relative version' of (9). In matrix notation, (A2) and (A5) can be expressed more compactly as

$$\mathcal{G}(\boldsymbol{d}) = \boldsymbol{p}' \times \tilde{\mathcal{M}}_D \times \boldsymbol{s} = \boldsymbol{p}' \times (\tilde{\mathcal{M}}_A + \tilde{\mathcal{M}}_A') \times \boldsymbol{s},$$

where $\boldsymbol{p} = (p_1, \ldots, p_G)$ is, again, the vector of population shares across groups, and $\boldsymbol{s} = (s_1, \ldots, s_G)$ is a column vector whose elements are defined as

$$s_g := \frac{n_g}{n}\frac{e_0^g}{e_0}.$$

$\tilde{\mathcal{M}}_D$ and $\tilde{\mathcal{M}}_A$ are the 'relative version' of the distance and advantage matrices $\mathcal{M}_D$ and $\mathcal{M}_A$ given in (10) and (11); that is, replacing the I_W^g, I_B^{gh} and A_{gh} terms by the corresponding $\tilde{I}_W^g$, $\tilde{I}_B^{gh}$ and $\tilde{A}_{gh}$ elements defined in (A2)–(A4).

Appendix C. The two-group decomposition of the absolute Gini coefficient

Assume a population that is partitioned in two groups, for instance females (F) and males (M), with respective population sizes n_F and n_M, such that the total population is $n = n_F + n_M$. Following (2), given the group-specific age-at-death distributions $\boldsymbol{d}_F = \{d_0^F, \cdots, d_\omega^F\}$ and $\boldsymbol{d}_M = \{d_0^M, \cdots, d_\omega^M\}$, the overall lifespan inequality measured by the absolute Gini coefficient can be broken down as

$$\begin{aligned}
\Delta(\boldsymbol{d}) &= \frac{1}{2\ell_0^2}\sum_{a=0}^{\omega}\sum_{b=0}^{\omega} d_a d_b |a-b| \\
&= \frac{1}{2\ell_0^2}\left(\sum_{b=0}^{\omega} d_0 d_b |0-b| + \sum_{b=0}^{\omega} d_1 d_b |1-b| + \cdots + \sum_{b=0}^{\omega} d_\omega d_b |\omega - b|\right) \\
&= \frac{1}{2\ell_0^2}(d_0 d_0 |0-0| + d_0 d_1 |0-1| + d_0 d_2 |0-2| + \cdots + d_0 d_\omega |0-\omega| + \cdots \\
&\quad + d_\omega d_0 |\omega - 0| + d_\omega d_1 |\omega - 1| + d_\omega d_2 |\omega - 2| + \cdots + d_\omega d_\omega |\omega - \omega|).
\end{aligned} \tag{A6}$$

Note that for any pair of ages a and b

$$\begin{aligned}
d_a d_b &= \left(\frac{n_F \cdot d_a^F + n_M \cdot d_a^M}{n}\right)\left(\frac{n_F \cdot d_b^F + n_M \cdot d_b^M}{n}\right) \\
&= \frac{n_F^2 \cdot d_a^F d_b^F + n_M^2 \cdot d_a^M d_b^M + n_F n_M \cdot d_a^F d_b^M + n_F n_M \cdot d_b^F d_a^M}{n^2}.
\end{aligned} \tag{A7}$$

In particular, when $a = b$

$$d_a d_a = \frac{n_F^2 \cdot d_a^F d_a^F + n_M^2 \cdot d_a^M d_a^M + 2 n_F n_M \cdot d_a^F d_a^M}{n^2}.$$

Using (A6) in (A7), we get

$$\begin{aligned}
\Delta(\boldsymbol{d}) = \frac{1}{2\ell_0^2}\Bigg(& \frac{n_F^2 \cdot d_0^F d_0^F + n_M^2 \cdot d_0^M d_0^M + 2 n_F n_M \cdot d_0^F d_0^M}{n^2} |0-0| \\
& + \frac{n_F^2 \cdot d_0^F d_1^F + n_M^2 \cdot d_0^M d_1^M + n_F n_M \cdot d_0^F d_1^M + n_F n_M \cdot d_1^F d_0^M}{n^2} |0-1| + \cdots \\
& + \frac{n_F^2 \cdot d_0^F d_\omega^F + n_M^2 \cdot d_0^M d_\omega^M + n_F n_M \cdot d_0^F d_\omega^M + n_F n_M \cdot d_\omega^F d_0^M}{n^2} |0-\omega| + \cdots \\
& + \frac{n_F^2 \cdot d_\omega^F d_0^F + n_M^2 \cdot d_\omega^M d_0^M + n_F n_M \cdot d_\omega^F d_0^M + n_F n_M \cdot d_0^F d_\omega^M}{n^2} |\omega - 0| \\
& + \frac{n_F^2 \cdot d_\omega^F d_1^F + n_M^2 \cdot d_\omega^M d_1^M + n_F n_M \cdot d_\omega^F d_1^M + n_F n_M \cdot d_1^F d_\omega^M}{n^2} |\omega - 1| + \cdots \\
& + \frac{n_F^2 \cdot d_\omega^F d_\omega^F + n_M^2 \cdot d_\omega^M d_\omega^M + 2 n_F n_M \cdot d_\omega^F d_\omega^M}{n^2} |\omega - \omega|\Bigg).
\end{aligned}$$

Re-arranging terms,

$$\begin{aligned}\Delta(\boldsymbol{d}) = &\frac{1}{2\ell_0^2}\Bigg(\left(\frac{n_F}{n}\right)^2\left(d_0^F d_0^F|0-0| + \ldots + d_0^F d_\omega^F|0-\omega| + \ldots + d_\omega^F d_0^F|\omega-0| + \ldots + d_\omega^F d_\omega^F|\omega-\omega|\right)\\ &+\left(\frac{n_M}{n}\right)^2\left(d_0^M d_0^M|0-0| + \ldots + d_0^M d_\omega^M|0-\omega| + \ldots + d_\omega^M d_0^M|\omega-0| + \ldots + d_\omega^M d_\omega^M|\omega-\omega|\right)\\ &+\frac{2n_F n_M}{n^2}\left(d_0^M d_0^F|0-0| + \ldots + d_0^M d_\omega^F|0-\omega| + \ldots + d_\omega^M d_0^F|\omega-0| + \ldots + d_\omega^M d_\omega^F|\omega-\omega|\right)\Bigg)\cdot\\ &= \mathcal{S}_F I_W^F + \mathcal{S}_M I_W^M + \mathcal{S}_{FM} I_B^{FM},\end{aligned}$$

where I_W^F is the within-group variation for females

$$I_W^F := \Delta(\boldsymbol{d}_F) = \frac{1}{2\ell_0^2}\sum_{a=0}^{\omega}\sum_{b=0}^{\omega} d_a^F d_b^F \cdot |a-b|,$$

I_W^M is the within-group variation for males

$$I_W^M := \Delta(\boldsymbol{d}_M) = \frac{1}{2\ell_0^2}\sum_{a=0}^{\omega}\sum_{b=0}^{\omega} d_a^M d_b^M \cdot |a-b|,$$

and I_B^{FM} the between-group variation

$$I_B^{FM} := \frac{1}{2\ell_0^2}\sum_{a=0}^{\omega}\sum_{b=0}^{\omega} d_a^F d_b^M \cdot |a-b|.$$

Each of these three terms have respective weights

$$\mathcal{S}_F := \left(\frac{n_F}{n}\right)^2,\ \mathcal{S}_F := \left(\frac{n_M}{n}\right)^2, \text{ and } \mathcal{S}_{FM} := 2\frac{n_F n_M}{n^2}.$$