Practice of Epidemiology

Measuring the Magnitude of Health Inequality Between 2 Population Subgroup Proportions

Makram Talih*, Ramal Moonesinghe, and David T. Huang

*Correspondence to Dr. Makram Talih, Instituto de Saúde Pública da Universidade do Porto, Rua das Taipas 135, 4050-600 Porto, Portugal (e-mail: mtalih@ispup.up.pt).

Initially submitted July 24, 2019; accepted for publication March 24, 2020.

In this paper, we evaluate 11 measures of inequality, \( d(p_1, p_2) \), between 2 proportions \( p_1 \) and \( p_2 \), some of which are new to the health disparities literature. These measures are selected because they are continuous, nonnegative, equal to 0 if and only if \( |p_1 - p_2| = 0 \), and maximal when \( |p_1 - p_2| = 1 \). They are also symmetrical \([d(p_1, p_2) = d(p_2, p_1)]\) and complement-invariant \([d(p_1, p_2) = d(1 - p_2, 1 - p_1)]\). To study intermeasure agreement, 5 of the 11 measures, including the absolute difference, are retained, because they remain finite and are maximal if and only if \( |p_1 - p_2| = 1 \). Even when the 2 proportions are assumed to be drawn at random from a shared distribution—interpreted as the absence of an avoidable difference—the expected value of \( d(p_1, p_2) \) depends on the shape of the distribution (and the choice of \( d \)) and can be quite large. To allow for direct comparisons among measures, we propose a standard measurement unit akin to a z score. For skewed underlying beta distributions, 4 of the 5 retained measures, once standardized, offer more conservative assessments of the magnitude of inequality than the absolute difference. We conclude that, even for measures that share the highlighted mathematical properties, magnitude comparisons are most usefully assessed relative to an elicited or estimated underlying distribution for the 2 proportions.

divergence; effect size; health inequality; inequality measurement; information theory; statistics

Healthy People is a US public health initiative that, for 4 decades, has established national goals and measurable objectives with 10-year targets to guide evidence-based policies, programs, and other actions to improve health and wellbeing (1). An overarching goal of Healthy People 2020 is to “achieve health equity, eliminate disparities, and improve the health of all groups” (1). Healthy People 2030 includes a similar goal (2). Healthy People publications and many other population health reports rely on measures of inequality to assess health disparities and monitor trends in disparities over time (3–9). Methodological work has focused on comparative studies of the inequality measures used in epidemiologic practice (10–14), identification of substantive implications of measurement choices (15–18), and development of scalable analytical tools for calculating and tracking inequalities over time (19, 20).

Much of the above-cited literature is concerned with summary health inequality indices, summarizing all possible pairwise comparisons among population subgroups (e.g., by race/ethnicity). However, even for conceptually simple comparisons between 2 subgroup proportions, assessment of the magnitude of inequality and of the direction and magnitude of change in inequality over time depend on analytical choices. For example, an analysis could focus on health attainment versus shortfall or on absolute versus relative differences (21–26). The assessment of changes in inequality may also be affected by changes in prevalence (27–29). Moreover, statistical significance alone may be insufficient to determine the public health importance of subgroup differences (30, 31).

Many measures of inequality are available and may lead to diverging conclusions about magnitude of and change in inequality over time (10–12, 15–17, 27–29). The consensus is to examine a suite of measures instead of relying on a single measure for drawing conclusions and public health recommendations (see Breen et al. (19) and Penman-Aguilar et al. (26) and references therein). For example, in 2017, the age-adjusted proportion of US adults aged 25 years or more who reported that they were in good or better health was 0.737 (73.7%) for persons with less than a high school diploma and 0.838 (83.8%) for those with a high school diploma or equivalent (32). The absolute
difference between these 2 proportions is statistically significant. However, without additional information, one cannot ascertain whether this difference of 0.101 is small or large. Relative measures can lead to divergent conclusions regarding the magnitude of inequality. The ratio of 1.14 (0.838/0.737 = 1.44) indicates that persons with a high school diploma were 14% more likely than those with less than a high school diploma to report being in good or better health. If complementary outcomes for adults in fair or poor health are compared, the ratio 1.62 (0.263/0.162 = 1.62) indicates that persons with less than a high school diploma were 62% more likely to report fair or poor health than those with a high school diploma. Because the various inequality measures may differ depending upon the prevalence of the outcome and the reference group selected, the magnitudes of these different measures cannot be directly compared. A standardized metric may facilitate comparisons across measures.

Our main objective in this paper is to present a standard measurement unit $u(p_1, p_2)$, defined as

$$u(p_1, p_2) = \frac{d(p_1, p_2) - E[d(p_1, p_2)]}{\sqrt{\text{Var}[d(p_1, p_2)]}}.$$  

(1)

This metric is constructed in the same way as a z score, by centering and scaling the inequality measure $d(p_1, p_2)$ between 2 subgroup proportions $p_1$ and $p_2$ relative to its mean $E[d(p_1, p_2)]$ and variance $\text{Var}[d(p_1, p_2)]$, under the assumption that the proportions are drawn at random from a known underlying distribution, allowing an “apples-to-apples” comparison of magnitude among measures.

Prior to constructing the standard measurement unit, we abstract mathematical properties that may impact comparability among different inequality measures. Two mathematical properties are most useful in distinguishing among inequality measures: firstly, whether they attain a “unique maximum,” namely that $d(p_1, p_2)$ is maximal if and only if $|p_1 - p_2| = 1$ (this property facilitates a simple reading of the worst-case scenario for inequality—one subgroup proportion is 0 and the other is 1); and secondly, the extent of the “penalty” that each measure assesses when small departures from equality occur (e.g., $p_1 = p_2 + \delta$ for a small $\delta > 0$), which may affect the mean value $E[d(p_1, p_2)]$ and variance $\text{Var}[d(p_1, p_2)]$ in equation 1.

**METHODS**

We focus on pairwise comparisons between proportions (typically multiplied by 100 and reported as percentages), because 70% of the nearly 1,100 measurable Healthy People 2020 objectives are tracked using percentages and because proportions are commonly used elsewhere (33, 34). “Inequality” is defined as a measurable quantity $d(p_1, p_2)$ that separates 2 subgroup proportions $p_1$ and $p_2$. Even though this operational definition is agnostic to whether the difference was avoidable, the extent to which inequality decreases reflects progress toward achieving equity (28, 35).

Drawing from the related concepts of statistical effect size and information-theoretical divergence, we formulate 11 candidate measures of inequality. Statistical effect size quantifies the magnitude of the difference between 2 proportions (36, 37). Effect-size measures include the absolute difference (Table 1, measure 1), the absolute logit difference (Table 1, measure 5), and the absolute probit difference (Table 1, measure 6), which appear in the disparities literature (27). When standardized, the absolute difference yields 2 effect-size measures that can be used to measure departure from equality: the standardized absolute difference with pooled variance and the standardized absolute difference without pooled variance (Table 1, measures 2 and 4). The rescaled absolute arcsine difference (Table 1, measure 3) also measures inequality (28, 38). Information-theoretical divergence gauges disagreement between any 2 probability distributions $P_1$ and $P_2$ (39) and quantifies the magnitude of inequality between $p_1$ and $p_2$ by specifying $P_j = (p_j, 1 - p_j)$, for $j = 1, 2$. The Theil index and its symmetrized version (Jeffreys divergence; Table 1, measure 8) are constructed from Kullback-Leibler divergence (12, 13, 40). Jensen-Shannon divergence (41, 42) is another symmetrized version of Kullback-Leibler divergence that can be adapted to measure inequality between 2 proportions (Table 1, measure 10). Chi-squared ($\chi^2$) divergence is related to Pearson’s goodness-of-fit statistic (39, 41, 42) and yields 2 symmetrized inequality measures: the symmetrized $\chi^2$ divergence (Table 1, measure 7) and the triangle discrimination measure (Table 1, measure 9). Hellinger distance (41, 42) yields another inequality measure (Table 1, measure 11).

As is discussed elsewhere (27–29), the extent of (dis)agreement among inequality measures in their assessment of magnitude depends on the location of the proportions $p_1$ and $p_2$ within the unit interval $[0, 1]$. To investigate the impact of their underlying distribution, the proportions are conceptualized as independent beta random variables. The beta family encompasses the uniform, U-shaped, and unimodal symmetrical, right-, or left-skewed densities. In this paper, we argue that the magnitude of inequality will be affected by the mean and variance of $d(p_1, p_2)$ given the underlying distribution for $p_1$ and $p_2$, and we propose the standard measurement unit shown in equation 1, allowing for direct comparisons among the selected measures.

**Abstracted mathematical properties for measures of inequality between proportions**

Drawing from an empirical investigation of properties of 2 benchmark inequality measures, the absolute difference, $|p_1 - p_2|$, and the ratio, $p_1/p_2$, we abstract some mathematical properties that may or may not be met by any given measure of inequality $d(p_1, p_2)$. All 11 measures surveyed satisfy properties 1–3 and 5–7 below. Seven of the 11 measures also satisfy property 4 (see Table 1).

**Property 1: Nonnegativity.** A nonnegative measure of inequality satisfies $d(p_1, p_2) \geq 0$ for all $p_1$ and $p_2$.

**Property 2: Absence of inequality (“egalitarian zero”).** Absence of inequality postulates that, for some $m \geq 0$, $d(p_1$, $p_2) = m$.
Table 1. Mathematical Expressions and Properties of 11 Candidate Measures of Inequality Between 2 Proportions^a

| Measure of Inequality d(p_1, p_2) Between Proportions p_1 and p_2 | Mathematical Expression | Property 4: Uniqueness of Maximal Inequality (e.g., Does d(p_1, p_2) = M if and Only If |p_1 − p_2| = 1?) | Behavior for |p_1 − p_2| Near 0 (e.g., Approximate Value of d(p_1 + δ, p_2) for δ Approaching 0 From Above) |
|---------------------------------------------------------------|------------------------|-------------------------------------------------|-----------------|-------------------------------------------------|
| 1. Absolute difference                                         | a(p_1, p_2) = |p_1 − p_2|                                      | Yes: M = 1    | δ                                              |
| 2. Standardized absolute difference, with pooled variance^b    | D_2(p_1, p_2) = \frac{|p_1 − p_2|}{\sqrt{\frac{p_1(1 − p_1) + p_2(1 − p_2)}{2}}} | Yes: M = 1              | δ × \frac{1}{\sqrt{p_1 p_2}}           |
| 3. Rescaled absolute arcsine difference^c                      | h(p_1, p_2) = 2\left\{\arcsin(\sqrt{p_1}) − \arcsin(\sqrt{p_2})\right\} | Yes: M = 1              | δ × \frac{1}{\sqrt{p_1 p_2}}           |
| 4. Standardized absolute difference^d                         | D_1(p_1, p_2) = \frac{|p_1 − p_2|}{\sqrt{\frac{p_1(1 − p_1) + p_2(1 − p_2)}{2}}} | Yes: M = +\infty        | δ × \frac{1}{\sqrt{p_1 p_2}}           |
| 5. Absolute logit difference                                   | ℓ(p_1, p_2) = |\logit(p_1) − \logit(p_2)| = |ln(OR)|                                 | No: M = +\infty can be attained for |p_1 − p_2| < 1 (e.g., if p_1 = 0). | δ × \frac{1}{\sqrt{p_1 p_2}}           |
| 6. Absolute probit difference                                   | b(p_1, p_2) = |\probit(p_1) − \probit(p_2)|                       | No: M = +\infty can be attained for |p_1 − p_2| < 1 (e.g., if p_1 = 0). | δ × \sqrt{2\pi}\exp(\frac{1}{2}\logit^2(p_2)) / 2 |
| Measures Based on Information-Theoretical Divergence          |                                                       |                                                |                                                |                                                |
| 7. Symmetrized \chi^2 divergence                               | X(p_1, p_2) = \frac{1}{2}(p_1 − p_2)^2\left[\frac{1}{p_1 + p_2} + \frac{1}{q_1 + q_2}\right] | No: M = +\infty can be attained for |p_1 − p_2| < 1 (e.g., if p_1 = 0). | δ^2 × \frac{1}{p_1 p_2}           |
| 8. Jeffreys divergence                                         | J(p_1, p_2) = \frac{1}{2}(p_1 − p_2) \times \ln(OR)                           | No: M = +\infty can be attained for |p_1 − p_2| < 1 (e.g., if p_1 = 0). | δ^2 × \frac{1}{p_1 p_2}           |
| 9. Triangle discrimination measure^e                          | \Delta(p_1, p_2) = \frac{1}{2}(p_1 − p_2)^2\left[\frac{1}{p_1 + p_2} + \frac{1}{q_1 + q_2}\right] | Yes: M = 1              | δ^2 × \frac{1}{p_1 p_2}           |
| 10. Rescaled Jensen-Shannon divergence^f                       | S_2(p_1, p_2) = \frac{1}{1 + \phi^2}\left[\frac{\ln p_1 + \ln p_2}{2} − p_1 \ln p_1 + \left(\frac{1}{q_1 + q_2} − q_1 \ln q_1\right)\right] \times | Yes: M = 1              | δ^2 × \frac{1}{\phi(1 + \phi)}          |
| 11. Hellinger distance                                         | H(p_1, p_2) = \sqrt{1 − \sqrt{p_1 p_2} − \sqrt{q_1 q_2}}                     | Yes: M = 1              | δ × \frac{1}{\sqrt{2D(p_1, p_2)}}           |

^a All 11 measures satisfy properties 1–3 and 5–7 and equal 0 and only if the 2 proportions p_1 and p_2 are equal. Given the 2 proportions p_1 and p_2, we define: q_1 = 1 − p_1, q_2 = 1 − p_2, p^0 = (p_1 + p_2)/2, and q^0 = 1 − p^0 = (q_1 + q_2)/2; OR = (p_1/p_2)(q_1/q_2); logit(OR) = p_1/p_2(1 − p_1) and probit(OR) = Φ^−1(OR), where Φ(x) is the standard normal distribution.

^b 2 × D_2 appears in the classical test of the null hypothesis that p_1 = p_2. Additionally, D_2 = \sqrt{\Delta}; see the Web Appendix (https://academic.oup.com/aje). D_2 is resolved by continuity to 0 when the proportions are both 0 or both 1.

^c Cohen’s index of effect size for the difference in proportions is \pi = h (37, 38).

^d D_1(0, 0) and D_1(1, 1) are resolved by continuity to 0.

^e \Delta(0, 0) and \Delta(1, 1) are resolved by continuity to 0.

^f Using the convention 0 × ln 0 = 0, S_2(0, 0) and S_2(1, 1) are resolved by continuity to 0.
$p_2 = m$ if and only if $p_1 = p_2$, reflecting attainment of equality between the 2 subgroup proportions.

Different values of $m$ may arise. For the absolute difference, $p_1 = p_2$ if and only if $|p_1 - p_2| = 0$; thus, $m = 0$. For the ratio, $p_1 = p_2$ if and only if $p_1/p_2 = 1$; thus, $m = 1$. Importantly, absence of inequality is met when the proportions are either both 0 or both 1. This is true for the absolute difference, but the ratio is undefined if $p_1 = p_2 = 0$.

Property 3: Minimal and maximal inequality. The property that $d(p_1, p_2) \geq m$, with $d(p_1, p_2) = m$ if and only if $|p_1 - p_2| = 0$ (minimal absolute difference), is consistent with property 2. The property that $d(p_1, p_2) \leq M$ for some (possibly infinite) $M > 0$, with $d(p_1, p_2) = M$ if $|p_1 - p_2| = 1$ (maximal absolute difference), is concerned with defining the magnitude of inequality in the worst-case scenario $|p_1 - p_2| = 1$, when one proportion is 0 and the other is 1.

While $|p_1 - p_2| = 1$ is sufficient for $d(p_1, p_2) = M$, it is not necessary. Some measures attain their maximum even if $|p_1 - p_2| < 1$. For example, the (absolute value of the) logarithm of the odds ratio can be maximal ($M = +\infty$) when just one of the proportions is 0 or 1, even if $|p_1 - p_2|$ remains small (see Table 1). This leads to the following property, which, together with the requirement that the maximum value, $M$, be finite, distinguishes 6 of the 11 measures listed in Table 1.

Property 4: Uniqueness of maximal inequality. The property that $d(p_1, p_2) \leq M$, with $d(p_1, p_2) = M$ if and only if $|p_1 - p_2| = 1$, is known as the “orthogonal maximum” property (see Mitchell [41]).

Property 5: Continuity. The measure $d(p_1, p_2)$ is a continuous function of its arguments if, for $\delta > 0$, all 4 quantities—$d(p_1, p_2 - \delta), d(p_1, p_2 + \delta), d(p_1 - \delta, p_2), d(p_1 + \delta, p_2)$—converge to $d(p_1, p_2)$ as $\delta$ approaches 0.

The absolute difference $|p_1 - p_2|$ and ratio $p_1/p_2$ are both continuous measures. A graph is usually sufficient for visual confirmation, but calculus techniques for demonstrating continuity are available [43].

There are 2 corollaries to property 5, which allow for limiting forms of properties 2 and 3.

Property 2′: For continuous measures, absence of inequality is understood to state that $d(p_1, p_2)$ will approach minimal inequality, $m$, as $|p_1 - p_2|$ approaches 0, and that it will remain near $m$ even as $p_2$ approaches 0 (or 1). The ratio $p_1/p_2$ is undefined when both proportions are 0, yet it satisfies this limiting form of property 2 (with $m = 1$).

Property 3′: As with minimal inequality, for continuous measures, the maximal inequality property is understood to indicate that $d(p_1, p_2)$ will approach $M$ as $|p_1 - p_2|$ approaches 1.

Two derived properties

We define a doubly symmetrical inequality measure as one that is both symmetrical and complement-invariant; these 2 concepts are defined below.

Property 6: Symmetry. The measure $d(p_1, p_2)$ is symmetrical (or “undirected”) if $d(p_1, p_2) = d(p_2, p_1)$.

The absolute difference $|p_1 - p_2|$ is symmetrical. The ratio $p_1/p_2$ is not symmetrical, since $p_2/p_1 \neq p_1/p_2$, and is therefore relative to the subgroup proportion used in the denominator.

As seen below and in the paper by Borrell and Talih (40), nonsymmetrical (or “directed”) inequality measures may be symmetrized—for example, by using the average $[d(p_1, p_2) + d(p_2, p_1)]/2$. Even though their magnitude may become difficult to interpret, symmetrization remains useful for priming various measures under consideration for a comparative assessment when directionality is only a secondary concern.

Property 7: Complement-invariance. The measure $d(p_1, p_2)$ is complement-invariant if $d(p_1, p_2)$ is $d(q_j, q_1)$, where $q_j = 1 - p_j$, $j = 1, 2$, are the complementary proportions.

The absolute difference $|p_1 - p_2|$ is complement-invariant. The ratio $p_1/p_2$ is not, since $q_2/q_1 \neq p_1/p_2$. The property of complement-invariance allows one to re-express inequality between proportions so that its magnitude is independent of whether the underlying health outcome is expressed as a favorable outcome or an adverse outcome, which, as discussed previously, is a major limitation for the ratio $p_1/p_2$. As with lack of symmetry, lack of complement-invariance can be corrected, albeit at the expense of interpretability—for example, using $[d(p_1, p_2) + d(q_2, q_1)]/2$ to prepare different inequality measures or health outcomes for comparison.

Measures of inequality between proportions with selected properties

The 11 measures surveyed are formulated from the related concepts of statistical effect size and information-theoretical divergence. These 2 classes of measures are elucidated in the Web Appendix (available at https://academic.oup.com/aje). All 11 measures are nonnegative (property 1) and satisfy the absence of inequality property (number 2) with $m = 0$ and the maximal inequality property (number 3) with $M = 1$ or $M = +\infty$. They are also continuous (property 5) and doubly symmetrical (properties 6 and 7) (see Table 1).

The ratio $p_1/p_2$, with $m = 1$, satisfies only a limiting form of property 2 and is maximal ($M = +\infty$) for all $p_1 > 0$ whenever $p_2 = 0$. Additionally, it is neither doubly symmetrical nor readily corrected for lack of symmetry and complement-invariance; hence, it is excluded from further comparisons. Nonetheless, as is shown in the Web Appendix, all 11 measures listed in Table 1 can be written as functions $d(R_{12}, R_{21})$ of the 2 ratios $R_{12} = p_1/p_2$ and $R_{21} = q_2/q_1$. Thus, the selected measures, including the absolute difference, are seen as relative measures, because proportions are intrinsically relative.

The absolute difference $|p_1 - p_2|$ serves as a benchmark for interpreting the minimum and maximum of those measures. Some of the measures shown (e.g., the absolute logit difference) will attain their maximum $M$ when just one of the proportions is 0 or 1, no matter how small $|p_1 - p_2|$ may be. If one wishes to avoid such an arbitrarily large “penalty” when proportions are near the boundary of the unit interval, one may rule out those measures by requiring uniqueness of maximal inequality (property 4). Table 1 identifies those measures that meet property 4 and those that do not.
Magnitude of Health Inequality Between Proportions

Figure 1. Possible values of selected inequality measures (shaded areas) as the absolute difference varies from 0 to 1. Panels A–D show the ranges of values of \( d = h, H, \sqrt{S_2}, \text{ and } \sqrt{\Delta}, \) respectively, as the absolute difference \( a \) varies between 0 and 1. The dashed line shows the 45° line of equality. The dotted lines represent the maximum vertical spread in the values of \( d \) as \( a \) varies from 0 to 1.

RESULTS

Relationships among selected measures

Measures 5–8 in Table 1 do not satisfy uniqueness of maximal inequality. Measure 4 is infinite when the absolute difference \( a = 1 \). In the rest of this paper, we focus on the remaining 6 measures, which take values between 0 and 1, where 0 indicates absence of inequality and 1 indicates maximal inequality:

- \( a \)—the absolute difference;
- \( D_2 \)—the standardized absolute difference, with pooled variance;
- \( h \)—the rescaled absolute arcsine difference;
- \( \sqrt{\Delta} \)—the square root of the triangle discrimination measure;
- \( \sqrt{S_2} \)—the square root of the rescaled Jensen-Shannon divergence; and
- \( H \)—Hellinger distance.

As is shown in the Web Appendix, \( D_2 \) is equal to \( \sqrt{\Delta} \). Agreement between the absolute difference \( a \) and each of the measures \( d = h, H, \sqrt{S_2}, \text{ and } \sqrt{\Delta} \) is illustrated in Figure 1. The shaded areas correspond to possible values of
Figure 2. Possible values of selected inequality measures (shaded areas) as the rescaled absolute arcsine difference varies from 0 to 1, with 3 thresholds for small, medium, and large Cohen’s effect sizes. The ranges of values of \( d = a, H, \sqrt{\Delta}, \) and \( \sqrt{S_2} \) for “small,” “medium,” and “large” effect sizes are represented in panels A–D, respectively, using the thick vertical line segments. Small, medium, and large effect sizes for the rescaled absolute arcsine difference \( (d = h) \) are calculated by dividing the thresholds 0.2, 0.5, and 0.8, respectively, for Cohen’s \( h \) index, by the number \( \pi \) (37, 38). The dashed line shows the 45° line of equality.

Cohen’s rule of thumb

In Figure 2, ranges of values of \( d = a, H, \sqrt{\Delta}, \) and \( \sqrt{S_2} \) for “small,” “medium,” and “large” effect sizes are shown. The latter are from the thresholds 0.2, 0.5, and 0.8, respectively, for Cohen’s \( h \) index, defined as \( \pi \times h \) (37, 38), and Figure 2A. Hellinger distance \( H \) is in near perfect agreement with the measure \( h \) (rescaled absolute arcsine difference), though \( h \leq H \) (Figure 2B and Web Appendix). The measures \( \sqrt{\Delta} \) and \( \sqrt{S_2} \) are also larger than \( h \) (Web Appendix) and, unlike the case for \( a \), there is no overlap between the ranges of values of \( \sqrt{\Delta} \) and \( \sqrt{S_2} \) corresponding to the 3 effect-size thresholds (Figures 2C and 2D).

Impact of underlying distribution

The proportions \( p_1 \) and \( p_2 \) may be conceptualized per se as random variables on the unit interval \([0, 1]\), generated a priori from the beta\((\alpha, \beta)\) distribution, for \( \alpha > 0 \) and \( \beta > 0 \) (see Web Figure 1). For \( p_1 \approx p_2 \), values of \( p_2 \) exist such that, in turn: 1) \( h \leq H \leq \sqrt{S_2} \leq a \leq \sqrt{\Delta}; \) 2) \( h \leq H \leq a \leq \sqrt{S_2} \leq \sqrt{\Delta}; \) 3) \( h \leq a \leq H \leq \sqrt{S_2} \leq \sqrt{\Delta}; \)
and 4) \(a \leq h \leq \sqrt{S_2} \leq \sqrt{\Delta} \) (see Web Appendix). Approximating the expected values \(E[d(p_1, p_2)]\) for \(d = a, h, H, \sqrt{S_2}, \) or \(\sqrt{\Delta}\) using numerical integration (values not shown here), one finds similar relationships: For the symmetrical beta densities with \(\alpha = \beta = 0.5, 1, 2, 4, 8,\) or 16, expected values are ordered as in inequality 1 in the above statement; for the skewed beta(1, 3) and beta(2, 6) densities, with mean = 0.25, their order is as in inequality 2 above; for beta(2, 13), with mean \(\approx 0.13\), it is as in inequality 3; and for beta(1, 6.5), with mean \(\approx 0.13\), and beta(1, 9) and beta (2, 18), with mean = 0.10, it is as in inequality 4.

Thus, of the inequality measures considered, only the square root of the triangle discrimination measure \(\sqrt{\Delta(p_1, p_2)}\) remains larger than the absolute difference \(a(p_1, p_2)\) for all \(p_1\) and \(p_2\) and the distributional scenarios considered.

### Magnitude comparisons among standardized health inequality measures

Figure 3 shows contours of the 5 measures \(a, \sqrt{\Delta}, \sqrt{S_2}, h,\) and \(H,\) standardized according to the formula in equation 1 using 6 selected shapes for the beta density. Each contour shows the set of pairs \((p_1, p_2)\) for which \(d(p_1, p_2)\) is at 0, 1, and 2 standard deviations from its expected value given the underlying beta distribution. For reference, the contours for \(|p_1 - p_2|\) are shown at finer granularity.

Under the U-shaped beta(0.5, 0.5) and uniform beta(1, 1) densities (Figures 3A and 3B, respectively), all 5 measures are in good agreement in that their level curves match up closely. However, average inequality tends to be larger in the U-shaped and uniform densities than in the unimodal densities beta(4, 4), beta(16, 16), beta(2, 6), and beta(2, 13), shown in Figures 3C–3F, respectively. This is seen in Figures 3C–3F, where the “0 standard deviations” level curves are closer to the line \(p_2 = p_1\).

For the 2 unimodal symmetrical densities beta(4, 4) and beta(16, 16) (Figures 3C and 3D, respectively), there is some intermeasure agreement, though the last 4 measures, once standardized, are more sensitive than the absolute difference (i.e., penalize inequality more) when one of the proportions is near the boundary of the unit interval, which is reflected in the increased curvature near 0 or 1.

For the 2 unimodal skewed densities beta(2, 6) and beta(2, 13) (Figures 3E and 3F, respectively), the last 4 measures, once standardized, are more conservative (i.e., penalize inequality less) than the absolute difference. For example, using beta(2, 6), the pair \((p_1, p_2) = (0.6, 0.2)\), with \(|p_1 - p_2| = 0.4\), registers at 2 standard deviations away from the mean for the measure \(a\), whereas the distance between \(p_1\) and \(p_2\) needs to increase further, to \(|p_1 - p_2| = 0.5\) (e.g., \((p_1, p_2) = (0.7, 0.2)\)), for the other 4 measures to register at 2 standard deviations. Using beta(2, 13), \((p_1, p_2) = (0.45, 0.2)\), with \(|p_1 - p_2| = 0.25\), registers at 2 standard deviations for \(a\), whereas \(|p_1 - p_2| = 0.35\) (e.g., \((p_1, p_2) = (0.55, 0.2)\)) is needed for the other 4 measures to register at 2 standard deviations.

Thus, the standard measurement unit \(u(p_1, p_2)\), with mean 0 and variance 1 regardless of the choice of health inequality measure \(d(p_1, p_2)\) or the underlying distribution of \(p_1\) and \(p_2\), allows assessment of inequality relative to what one would expect given prior information on the 2 proportions.

### DISCUSSION

We evaluated 11 measures of inequality between proportions \(p_1\) and \(p_2\). These measures were selected because they are continuous, nonnegative, equal to 0 if and only if \(|p_1 - p_2| = 0\), maximal when \(|p_1 - p_2| = 1\), and doubly symmetrical. To assess intermeasure agreement and develop a standard measurement unit for the magnitude of inequality, we retained the absolute difference and 4 other measures for further analysis, because, in addition to the mathematical properties they share with the remaining 6 measures, they are finite and are maximal if and only if \(|p_1 - p_2| = 1\).

For skewed underlying beta distributions, the retained measures, once standardized relative to their mean and variance, were more conservative than the absolute difference in their assessment of magnitude of inequality, demonstrating the potential impact of the underlying distribution.

In this paper, we did not address the difficult methodological issue that different measures may lead to divergent assessments of changes in inequality over time. Theoretical bounds for the range of proportions where the selected measures converge in their assessment of trends are not readily available. Instead, Web Tables 1–9 present empirical findings in relation to 4 Healthy People 2020 objectives and 5 leading causes of death. Those examples were selected to illustrate various distributional scenarios for the proportions, and, together with the relatively narrow areas of disagreement shown in Figure 2, suggest that the retained measures (the last 4 in each Web table) may agree more than they disagree in their assessment of trends.

The proposed standard measurement unit depends on the specification of an underlying distribution. We adopted a Bayesian perspective and assumed that the 2 proportions \(p_1\) and \(p_2\) were independent and identically distributed beta random variables; thus, the assumed-known mean and variance of \(d(p_1, p_2)\) were calculated directly (albeit via numerical approximation) rather than estimated from data. In practice, analysts could elicit a prior distribution for \(p_1\) and \(p_2\) from consensus or expert opinion (44). Alternatively, analysts could adopt an empirical Bayes approach, estimating the distribution from historical data (45), or proceed within the frequentist setting, modeling the 2 sample proportions \(x_j/n_j\) as being overdispersed relative to their binomial variance, leading to the beta-binomial distribution (46). Although standardized measures may be confounded by sample properties and may not be comparable to their unstandardized counterparts (47), expressing competing inequality measures as multiples of a standard inequality unit remains useful for direct comparisons among those measures.

The comparative analysis demonstrated in this paper was restricted to mathematical properties that were formulated following an empirical investigation of 2 benchmark measures, the absolute difference and the ratio. In selecting from the 11 measures surveyed, we did not consider the interpretability and clinical or public health relevance of those measures, nor did we consider ease of communication to stakeholders. The standard measurement unit may offer...
Figure 3. Level curves of selected inequality measures after standardization relative to their expected values and variances for various choices of the underlying beta distribution for the proportions. Level lines for the absolute difference ($|\omega|$) are drawn for $|p_1 - p_2|$ at 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 standard deviations from the mean $E[|p_1 - p_2|]$ under each of the selected beta distributions—beta(0.5, 0.5) (A), beta(1, 1) (B), beta(4, 4) (C), beta(16, 16) (D), beta(2, 6) (E), and beta(2, 13) (F)—and are labeled using the oblique set of labels shown above the x-axis in each graph. Level curves for the other 4 measures—$d = \sqrt{\Delta}$, $\sqrt{S_2}$, $H$, and $h$—are only drawn for $d(p_1, p_2)$ at 0.0, 1.0, and 2.0 standard deviations from the mean $E[d(p_1, p_2)]$ and are labeled using the vertical set of labels shown to the right of each graph. Recall that $\sqrt{\Delta}$ denotes the square root of the triangle discrimination measure, $\sqrt{S_2}$ the square root of the rescaled Jensen-Shannon divergence, $H$ the Hellinger distance, and $h$ the rescaled absolute arcsine difference. The dashed line shows the 45° line of equality.
an easily accessible gauge for the magnitude of inequality, useful in meta-analyses, but its dependence on a potentially elicited prior distribution may remain a barrier to interpretability.

Starting with Healthy People 2020, the Healthy People initiative has moved to considering a suite of measures to examine health disparities instead of relying on a single measure (9, 20). In a recent editorial, Duran et al. (48) urged harmonization for consistency in measurement and reporting of health disparities and promoted sentinel disparities indicators. A standard measurement unit is a step toward achieving this goal, because it allows an “apples-to-apples” comparison among inequality measures and, at least in the context of proportions, obviates the need to endorse (implicitly or explicitly) a value judgment in selecting among measures.

ACKNOWLEDGMENTS

Author affiliations: Division of Analysis and Epidemiology, National Center for Health Statistics, Hyattsville, Maryland (Makram Talih, David T. Huang); Office of Minority Health and Health Equity, Centers for Disease Control and Prevention, Atlanta, Georgia (Ramal Moonesinghe); and Institute of Public Health of the University of Porto, Porto, Portugal (Makram Talih).

The findings and conclusions presented in this paper are those of the authors and do not necessarily represent the official position of the National Center for Health Statistics or the Centers for Disease Control and Prevention.

Conflict of interest: none declared.

REFERENCES


