HHS Public Access

Author manuscript

Am J Epidemiol. Author manuscript; available in PMC 2021 September 01.

Published in final edited form as:

Am J Epidemiol. 2020 September 01; 189(9): 987–996. doi:10.1093/aje/kwaa050.

Measuring the Magnitude of Health Inequality between Two Population Subgroup Proportions

Makram Talih [Contractor],

Division of Analysis and Epidemiology, National Center for Health Statistics, Hyattsville, Maryland

University of Porto Institute of Public Health, Porto, Portugal

Ramal Moonesinghe,

Office of Minority Health and Health Equity, Centers for Disease Control and Prevention, Atlanta, Georgia

David T. Huang

Division of Analysis and Epidemiology, National Center for Health Statistics, Hyattsville, Maryland

Abstract

The paper evaluates 11 measures of inequality $d(p_1,p_2)$ between two 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 zero if and only if $|p_1-p_2|=0$, and maximal when $|p_1-p_2|=1$. They are also symmetric [$d(p_1,p_2)=d(p_2,p_1)$] and complement-invariant [$d(p_1,p_2)=d(1-p_2,1-p_1)$]. To study inter-measure agreement, five 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 two 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, a standard measurement unit akin to a z-score is proposed. For skewed underlying beta distributions, four of the five retained measures, once standardized, offer more conservative assessments of the magnitude of inequality than the absolute difference. The paper concludes 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 two proportions.

Keywords

Divergence; health inequality measurement; information theory; statistical effect size

---

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

Conflict of interest: The authors have no conflict of interest to declare.

Disclaimer: The findings and conclusions 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.

Healthy People is a U.S. public health initiative that, for four 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 well-being (1). An overarching goal of Healthy People 2020 is to “achieve health equity, eliminate disparities, and improve the health of all groups.” 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 epidemiological practice (10–14), identification of substantive implications of measurement choices (15–18), and development of scalable analytic 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 two subgroup proportions, the assessment of the magnitude of inequality as well as the direction and magnitude of the change in inequality over time depend on analytic choices. For example, an analysis could focus on health attainment versus shortfall, or absolute versus relative differences (21–26). The assessment of changes in inequality may also be impacted 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 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 (19, 26) and references therein. For example, in 2017, the age-adjusted proportion of adults aged 25 and over in good or better health was 0.737 (=73.7%) for those 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 two 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 0.838/0.737=1.14 indicates that those with a high school diploma were 14% more likely than those with less than a high school diploma to be in good or better health. If complementary outcomes for adults in fair or poor health are compared, the ratio 0.263/0.162=1.62 indicates that those 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 magnitude of these different measures cannot be directly compared. A standardized metric may facilitate comparisons across measures.

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

$$u(p_1, p_2) = \frac{d(p_1, p_2) - \mathrm{E}[d(p_1, p_2)]}{\sqrt{\mathrm{Var}[d(p_1, p_2)]}} \quad [1]$$

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 two subgroup proportions p_1 and p_2 relative to its mean $E[d(p_1,p_2)]$ and variance $Var[d(p_1,p_2)]$ 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, the paper abstracts 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 zero and the other is one. 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 impact the mean $E[d(p_1,p_2)]$ and variance $Var[d(p_1,p_2)]$ in equation 1.

METHODS

This paper focuses 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 proportions are commonly used elsewhere (33, 34). “Inequality” is defined as a measurable quantity $d(p_1,p_2)$ that separates two 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-theoretic divergence, 11 candidate inequality measures are formulated. Statistical effect size quantifies the magnitude of the difference between two proportions (36, 37). Effect size measures include the absolute difference (Table 1, measure 1), absolute logit difference (Table 1, measure 5), and absolute probit difference (Table 1, measure 6), which appear in the disparities literature (27). When standardized, the absolute difference yields two effect size measures that can be used to measure departure from equality: the standardized absolute difference with, and without, pooled variance (Table 1, measures 2 and 4). The rescaled absolute arcsine difference (Table 1, measure 3) also measures inequality (28, 38). Information-theoretic divergence gauges disagreement between any two 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 Kullback-Leibler divergence that can be adapted to measure inequality between two proportions (Table 1, measure 10). Chi-squared divergence is related to Pearson’s goodness-of-fit statistic (39, 41,

42) and yields two symmetrized inequality measures: the symmetrized chi-squared 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 discussed in (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 symmetric, right, or left-skewed densities. The paper argues that the magnitude of inequality will be impacted by the mean and variance of $d(p_1,p_2)$ given the underlying distribution for p_1 and p_2, and proposes 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 two benchmark inequality measures, the absolute difference, $|p_1-p_2|$, and the ratio, p_1/p_2, this section abstracts 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. The absolute difference, $|p_1-p_2|$, and the ratio, p_1/p_2, are nonnegative measures of inequality.

Property 2: Absence of inequality ("egalitarian zero")—Absence of inequality postulates that, for some $m\geq 0$, $d(p_1,p_2)=m$ if and only if $p_1=p_2$, reflecting attainment of equality between the two 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 zero or both one. 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 magnitude of inequality in the worst-case scenario $|p_1-p_2|=1$, when one proportion is zero and the other is one.

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 zero or one, 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 six 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 (41).

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

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

There are two 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 zero, 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

This paper defines a *doubly symmetric* inequality measure as one that is both symmetric and complement-invariant; these two concepts are defined next.

Property 6: Symmetry—The measure $d(p_1,p_2)$ is symmetric (or "undirected") if $d(p_1,p_2)=d(p_2,p_1)$.

The absolute difference $|p_1-p_2|$ is symmetric. The ratio p_1/p_2 is not symmetric, 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 (40), non-symmetric (or "directed") inequality measures may be symmetrized, e.g., 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 $d(p_1,p_2)=d(q_2,q_1)$, 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 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, e.g., 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-theoretic divergence. These two classes of measures are elucidated in Web Appendix 1. All 11 measures are nonnegative (property 1), satisfy the absence of inequality property (no. 2) with m=0, and maximal inequality property (no. 3), with M=1 or M=+∞. They are also continuous (property 5) and doubly symmetric (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= +∞) for all $p_1>0$ whenever $p_2=0$. Additionally, it is neither doubly symmetric nor readily corrected for lack thereof, hence it is excluded from further comparisons. Nonetheless, as seen in Web Appendix 1, all 11 measures listed in Table 1 can be written as functions $d(R_{12},\check{R}_{21})$ of the two ratios $R_{12}=p_1/p_2$ and $\check{R}_{21}=q_2/q_1$. 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, then 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.

Another point of reference for comparing the selected measures is the rate of change in $d(p_1,p_2)$ as $|p_1-p_2|$ decreases toward zero. Thus, measures in Table 1 are characterized by their behavior for small values of $|p_1-p_2|$ and whether/how their rate of change depends on the location of the two proportions on the unit interval (e.g., both near 0 or 1, or both near 0.5). Mathematical derivations are included in Web Appendix 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. The rest of this paper focuses on the remaining six measures, which take values between 0 and 1, where 0 indicates absence of inequality and 1 indicates maximal inequality:

a—Absolute difference;

D_2—Standardized absolute difference, with pooled variance; h—Rescaled absolute arcsine difference;

$\sqrt{\Delta}$—Square root of triangle discrimination measure;

$\sqrt{S_2}$—Square root of rescaled Jensen-Shannon divergence; and

H—Hellinger distance.

As shown in Web Appendix 1, D_2 is equal to $\sqrt{\Delta}$. Agreement between the absolute difference a and each of the measures d=h, H, $\sqrt{S_2}$, or $\sqrt{\Delta}$ is illustrated in Figure 1. The shaded areas correspond to possible values of d as a ranges from 0 to 1. For example, Figure 1D shows that $0 \leq \sqrt{\Delta} - a \leq 0.1$ for small values of a (in fact, $a \leq \sqrt{\Delta}$ everywhere; see Web Appendix 1). Figure 1A shows that $0 \leq a - h \leq 0.2$ for large values of a. Additional comparisons are presented below and in Web Appendix 1.

Cohen's rule of thumb

In Figure 2, ranges of values of d=a, H, $\sqrt{\Delta}$, or $\sqrt{S_2}$ 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 $\pi \times h$; see (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 1). The measures $\sqrt{\Delta}$ and $\sqrt{S_2}$ are also larger than h (Web Appendix 1) and, unlike for a, there is no overlap between the ranges of values of $\sqrt{\Delta}$ and $\sqrt{S_2}$ corresponding to the three 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(α,β) distribution, for α>0 and β>0; see Web Figure 1. For $p_1 \approx p_2$, values of p_2 exist such that, in turn: (i) $h \leq H \leq \sqrt{S_2} \leq a \leq \sqrt{\Delta}$; (ii) $h \leq H \leq a \leq \sqrt{S_2} \leq \sqrt{\Delta}$; (iii) $h \leq a \leq H \leq \sqrt{S_2} \leq \sqrt{\Delta}$; or (iv) $a \leq h \leq H \leq \sqrt{S_2} \leq \sqrt{\Delta}$; see Web Appendix 1. 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 symmetric beta densities with α=β=0.5, 1, 2, 4, 8, or 16, expected values are ordered as in (i); for the skewed beta(1,3) and beta(2,6) densities, with mean=0.25, their order is as in (ii); for beta(2,13), with mean≈0.13, it is as in (iii); and for beta(1,6.5), with mean≈0.13, and beta(1,9) and beta(2,18), with mean=0.10, it is as in (iv).

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 five measures a, $\sqrt{\Delta}$, $\sqrt{S_2}$, h, and H, standardized according to the formula in equation 1 using six 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 five 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 two unimodal symmetric densities beta(4,4) and beta(16,16) (Figures 3C and 3D, respectively), there is some inter-measure agreement, though the last four 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 two unimodal skewed densities beta(2,6) and beta(2,13) (Figures 3E and 3F, respectively), the last four 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 four 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 four measures to register at 2 standard deviations.

Thus, the standard measurement unit $u(p_1,p_2)$, with mean zero and variance one 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 two proportions.

DISCUSSION

The paper evaluated 11 measures of inequality between proportions p_1 and p_2. These measures were selected because they are continuous, nonnegative, equal to zero if and only if $|p_1-p_2|=0$, maximal when $|p_1-p_2|=1$, and doubly symmetric. To assess inter-measure agreement and develop a standard measurement unit for the magnitude of inequality, the absolute difference and four other measures were retained for further analysis, because, in addition to the mathematical properties they share with the remaining six measures, they are finite, and 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.

This paper 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 four Healthy People 2020 objectives and five 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 (last four 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. The paper adopted a Bayesian perspective and assumed that the two 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 two 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 in the paper was restricted to mathematical properties that were formulated following an empirical investigation of two benchmark measures, the absolute difference and the ratio. In selecting from the 11 measures surveyed, the paper did not consider the interpretability and clinical or public health relevance of those measures, nor did it consider ease of communication to stakeholders. The standard measurement unit may offer 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). A recent editorial urges harmonization for consistency in measurement and reporting of health disparities and promote sentinel disparities indicators (48). 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 judgement in selecting among measures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding: The authors have no funding to report.

References

1. US Department of Health and Human Services. About Healthy People: Healthy People 2020. Washington, DC. https://www.healthypeople.gov/2020/About-Healthy-People. Accessed July 24, 2019.

2. US Department of Health and Human Services. About Healthy People: Healthy People 2030 Draft Framework. Washington, DC. https://www.healthypeople.gov/2020/About-Healthy-People/Development-Healthy-People-2030/Draft-Framework Accessed July 24, 2019.
3. National Center for Health Statistics. Healthy People 2020 Midcourse Review. Hyattsville, MD: National Center for Health Statistics; 2016 (DHHS Pub. No. 2017–1042).
4. National Center for Health Statistics. Health, United States, 2015: With Special Feature on Racial and Ethnic Health Disparities. Hyattsville, MD: National Center for Health Statistics; 2016 (DHHS Pub. No. 2016–1232).
5. Agency for Healthcare Research and Quality. 2016 National Healthcare Quality and Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality; 2017 (AHRQ Pub. No. 17–0001).
6. US Department of Health and Human Services. Healthy People in Action: State and Territorial Healthy People Plans. Washington, DC. https://www.healthypeople.gov/2020/healthy-people-in-action/State-and-Territorial-Healthy-People-Plans Accessed July 24, 2019.
7. Wisconsin Department of Health Services. Healthiest Wisconsin 2020: Wisconsin State Health Plan. Madison, WI. https://www.dhs.wisconsin.gov/hw2020/index.htm. Accessed July 24, 2019.
8. University of Wisconsin Population Health Institute. County Health Rankings & Roadmaps. Madison, WI. http://www.countyhealthrankings.org/reports/. Accessed July 24, 2019.
9. Talih M, Huang DT. Measuring progress toward target attainment and the elimination of health disparities in Healthy People 2020 Hyattsville, MD: National Center for Health Statistics Healthy People Stat Notes. 2016;(27):1–19.
10. Keppel K, Pamuk E, Lynch J, et al. Methodological issues in measuring health disparities Hyattsville, MD: National Center for Health Statistics Vital Health Stat. 2005;2(141):1–16.
11. Harper S, Lynch J. Selected Comparisons of Measures of Health Disparities: A Review Using Databases Relevant to Healthy People 2010 Cancer-Related Objectives. Bethesda, MD: National Cancer Institute; 2007 (NIH Pub. No. 07–6281).
12. Harper S, Lynch J, Meersman SC, et al. An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race/ethnicity, 1992–2004. Am J Epidemiol. 2008;167(8):889–899. [PubMed: 18344513]
13. Talih M A reference-invariant health disparity index based on Rényi divergence. Ann Appl Stat. 2013;7(2):1217–1243. [PubMed: 26568778]
14. Harper S, Lynch J. Health Inequalities: Measurement and Decomposition In Oakes JM, Kaufman JS (eds). Methods in Social Epidemiology, 2nd Edition. San Francisco: Jossey-Bass 2017.
15. Messer LC. Invited commentary: measuring social disparities in health—what was the question again? Am J Epidemiol. 2008;167:900–904. [PubMed: 18344512]
16. Bhopal RS. Re: “An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992–2004” [letter]. Am J Epidemiol. 2008;167:1214–1216.
17. Harper S, King NB, Meersman SC, et al. Implicit value judgments in the measurement of health inequalities. Milbank Q. 2010;88(1):4–29. [PubMed: 20377756]
18. Kjellsson G, Gerdtham UG, Petrie D. Lies, damned lies, and health inequality measurements: understanding the value judgements. Epidemiol. 2015;26:673–680.
19. Breen N, Scott S, Percy-Laurry A, et al. Health disparities calculator: A methodologically rigorous tool for analyzing inequalities in population health. Am J Public Health. 2014;104(9):1589–1591. [PubMed: 25033114]
20. U.S. Department of Health and Human Services. Healthy People 2020 Health Disparities Data. Washington, DC. https://www.healthypeople.gov/2020/data-search/health-disparities-data. Accessed July 24, 2019.
21. Scanlan JP. Can we actually measure health disparities? Chance. 2006;19(2):47–51.
22. Asada Y On the choice of absolute or relative inequality measures. Milbank Q. 2010;88(4):616–622. [PubMed: 21166871]
23. King NB, Harper S, Young ME. Use of relative and absolute effect measures in reporting health inequalities: structured review. BMJ. 2012;345:e5774. [PubMed: 22945952]

24. Lambert PJ, Subramanian S. Disparities in socio-economic outcomes: some positive propositions and their normative implications. Soc Choice Welf. 2014;43(3):565–576.
25. Moonesinghe R, Beckles G. Measuring health disparities: a comparison of absolute and relative disparities. Peer J. 2015;3:e1438. [PubMed: 26623191]
26. Penman-Aguilar A, Talih M, Huang D, et al. Measurement of health disparities, health inequities, and social determinants of health to support the advancement of health equity. J Public Health Manag Pract. 2016;22(1-suppl):S33–S42. [PubMed: 26599027]
27. Scanlan JP. The mismeasure of health disparities. J Public Health Manag Pract. 2016;22(4):415–419. [PubMed: 27223313]
28. Penman-Aguilar A, Talih M, Moonesinghe R, Huang D. Response to Scanlan Concerning: Measurement of Health Disparities, Health Inequities, and Social Determinants of Health to Support the Advancement of Health Equity [letter]. J Public Health Manag Pract. 2016;22(6):614–615. [PubMed: 27682733]
29. Thomas H, Hettmansperger T. Risk ratios and Scanlan's HRX. Journal of Statistical Distributions and Applications. 2017;4:27.
30. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. Am Stat. 2016;70(2):129–133.
31. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p<0.05". Am Stat. 2019;73(sup1):1–19.
32. National Center for Health Statistics. Summary Health Statistics: National Health Interview Survey, 2017 [Table P-1a]. Hyattsville, MD. https://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS/2017_SHS_Table_P-1.pdf.
33. US Department of Health and Human Services. DATA2020: Healthy People 2020 Database. Washington, DC https://www.healthypeople.gov/2020/data-search/. Accessed July 24, 2019.
34. Parker JD, Talih M, Malec DJ, et al. National Center for Health Statistics Data Presentation Standards for Proportions Hyattsville, MD: National Center for Health Statistics Vital Health Stat. 2017;2(175):1–14.
35. Braveman P Health disparities and health equity: concepts and measurement. Annu Rev Public Health. 2006;27:167–194. [PubMed: 16533114]
36. Kelley K, Preacher KJ. On effect size. Psychol Methods. 2012;17(2):137–152. [PubMed: 22545595]
37. Sánchez-Meca J, Marín-Martínez F, Chacón-Moscoso S. Effect-size indices for dichotomized outcomes in meta-analysis. Psychol Methods. 2003;8(4):448–467. [PubMed: 14664682]
38. Cohen J Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York, NY: Psychology Press, Taylor & Francis Group; 1988.
39. Cressie N, Read TRC. Multinomial goodness-of-fit tests. J R Statist Soc B. 1984;46(3):440–464.
40. Borrell LN, Talih M. A symmetrized Theil index measure of health disparities: An example using dental caries in U.S. children and adolescents. Stat Med. 2011;30(3):277–290. [PubMed: 21213344]
41. Mitchell SA. Geometric comparison of popular mixture model distances. Journal of Modern Mathematics Frontier. 2012;1(4):1–21.
42. Taneja IJ. Refinement inequalities among symmetric divergence measures. Australian Journal of Mathematical Analysis and Applications. 2005;2(1):1–23.
43. Stewart J. Calculus. 8th ed. Boston, MA: Cengage Learning; 2016.
44. O'Hagan RA. Expert knowledge elicitation: Subjective but scientific. Am Stat. 2019;73(sup1):69–81.
45. Greenland S Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol. 2006;35(3):765–775. [PubMed: 16446352]
46. Rothman KJ, Greenland S, Lash TJ. Modern Epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2008 [p. 422].
47. Greenland S Standardized regression coefficients: A Further critique and review of some alternatives. Epidemiol. 1991;2(5):387–392.

48. Duran D, Asada Y, Millum J, Gezmu M. Harmonizing health disparities measurement. Am J Public Health. 2019;109(suppl-1):S25–S27. [PubMed: 30699026]

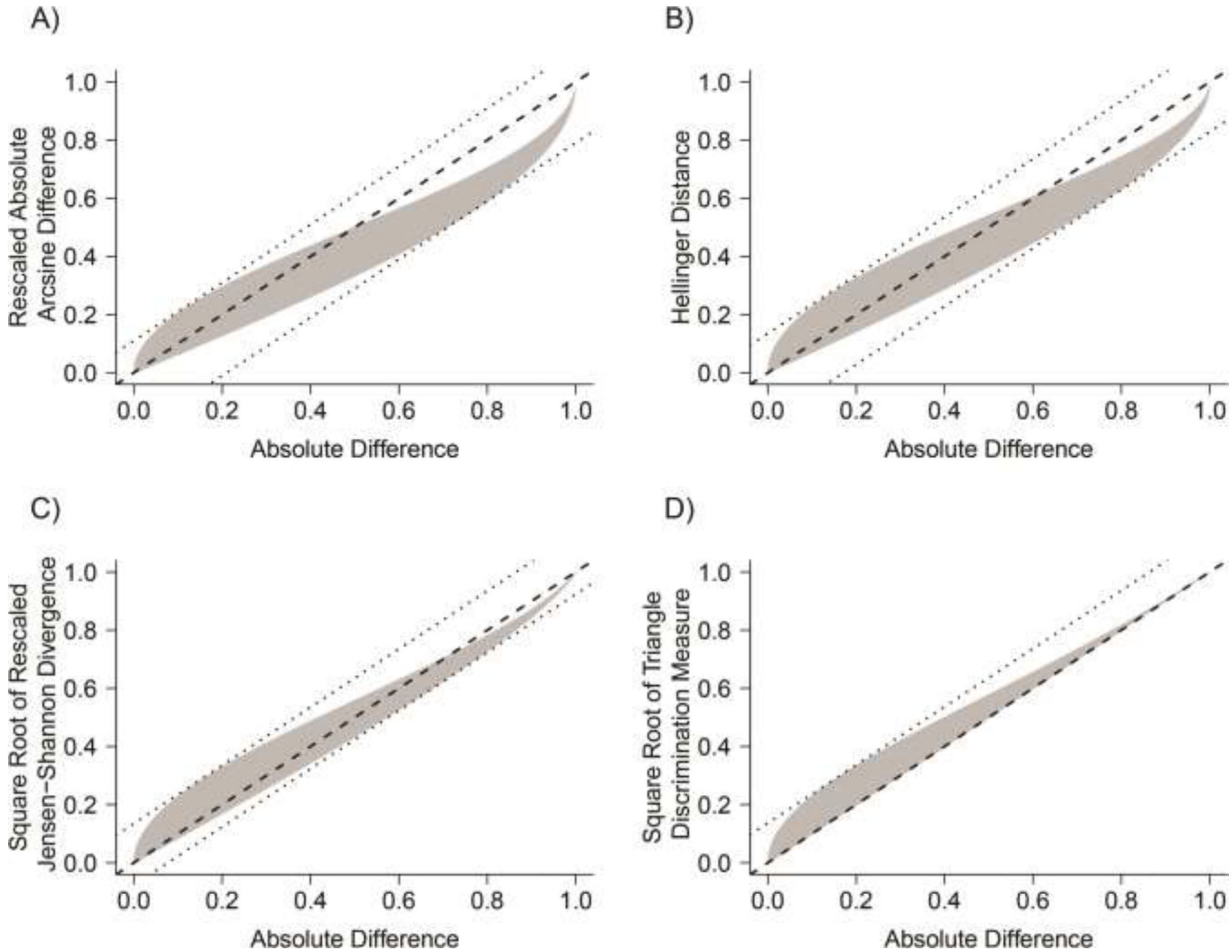

Figure 1.

Possible Values of Selected Inequality Measures (Shaded Areas) as the Absolute Difference Varies from Zero to One

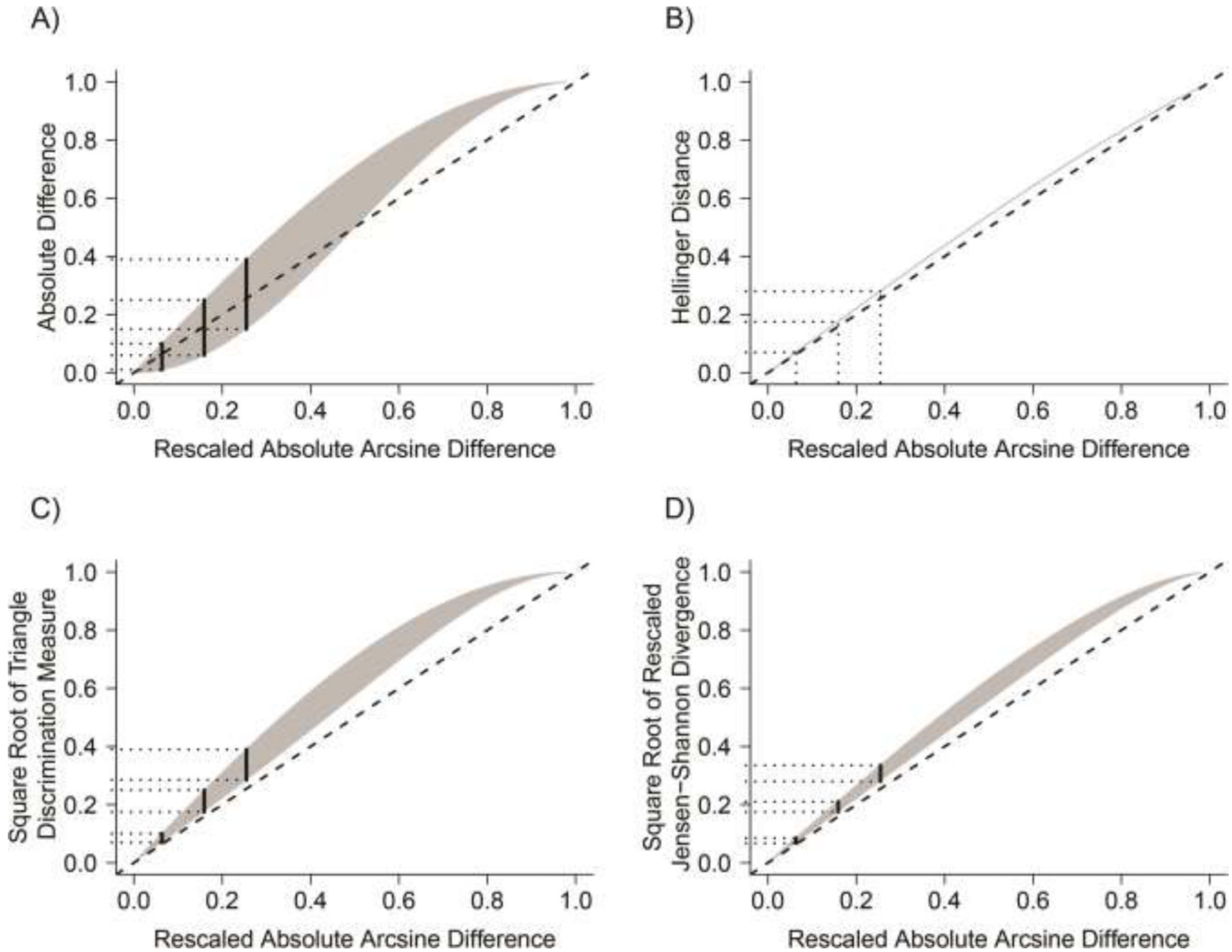

Figure 2.

Possible Values of Selected Inequality Measures (Shaded Areas) as the Rescaled Absolute Arcsine Difference Varies from Zero to One, with Three Thresholds for Small, Medium, and Large Cohen's Effect Sizes

The ranges of values of $d=a$, H, $\sqrt{\Delta}$, or $\sqrt{S_2}$ for "small", "medium", and "large" effect sizes are represented in A) through 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 π; see (37, 38).

Author Manuscript

Author Manuscript

Author Manuscript

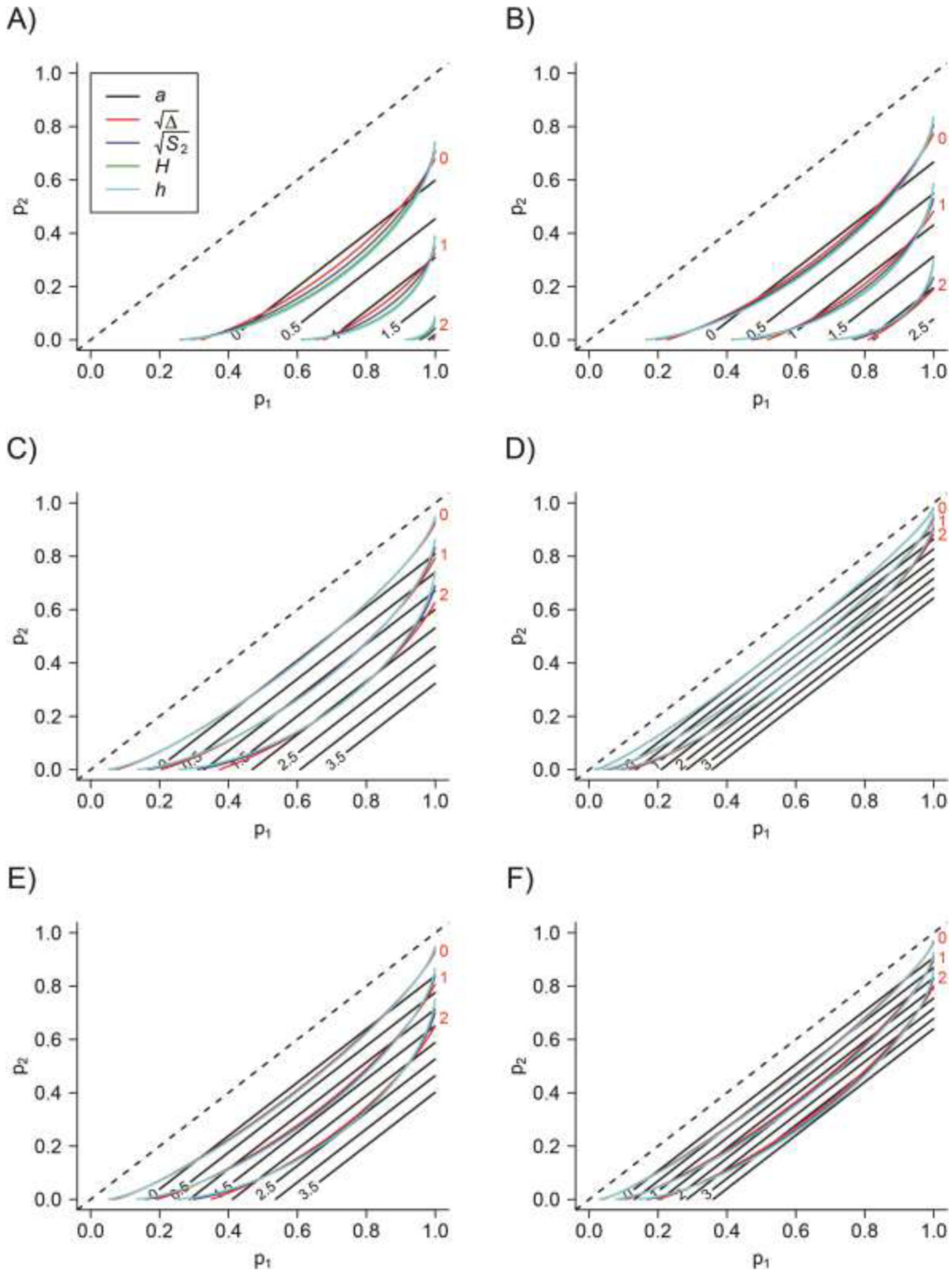

Figure 3.

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 (*a*) are drawn for $|p_1\text{-}p_2|$ at 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5 standard deviations from the mean $E[|p_1\text{-}p_2|]$ under each of the selected beta distributions—A) beta(0.5,0.5); B) beta(1,1); C) beta(4,4); D) beta(16,16); E) beta(2,6); and F) beta(2,13)—and are labeled using the oblique set of labels shown above the x-axis in each graph. Level curves for the other four 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 triangle discrimination measure, $\sqrt{S_2}$ the square root of rescaled Jensen-Shannon divergence, H the Hellinger distance, and h the rescaled absolute arcsine difference.

Table 1.

Summary of Two Mathematical Properties of 11 Measures of Inequality Between Two 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., $d(p_1,p_2)=M$ if and only if $\|p_1-p_2\|=1$?**	**Behavior for $\|p_1-p_2\|$ near 0, e.g., approximate value of $d(p_2+\delta,p_2)$ for δ approaching 0 from above.**
Measures based on statistical effect size			
1. Absolute difference	$a(p_1,p_2) = \|p_1-p_2\|$	Yes: M=1	δ
2. Standardized absolute difference, with pooled variance[b]	$D_2(p_1,p_2) = \frac{\|p_1-p_2\|}{2\sqrt{p_* q_*}}$	Yes: M=1	$\delta \times \frac{1}{2\sqrt{p_2 q_2}}$
3. Rescaled absolute arcsine difference[c]	$h(p_1,p_2) = \frac{2}{\pi}\|\arcsin(\sqrt{p_1}) - \arcsin(\sqrt{p_2})\|$	Yes: M=1	$\delta \times \frac{1}{\pi\sqrt{p_2 q_2}}$
4. Standardized absolute difference[d]	$D_1(p_1,p_2) = \frac{\|p_1-p_2\|}{\sqrt{(p_1 q_1 + p_2 q_2)/2}}$	Yes: M=+∞	$\delta \times \frac{1}{\sqrt{p_2 q_2}}$
5. Absolute logit difference	$\ell(p_1,p_2) = \|\mathrm{logit}(p_1) - \mathrm{logit}(p_2)\| = \|\ln(\mathrm{OR})\|$	No: M=+∞ can be attained for $\|p_1-p_2\|<1$, e.g., if $p_1=0$.	$\delta \times \frac{1}{p_2 q_2}$
6. Absolute probit difference	$b(p_1,p_2) = \|\mathrm{probit}(p_1) - \mathrm{probit}(p_2)\|$	No: M=+∞ can be attained for $\|p_1-p_2\|<1$, e.g., if $p_1=0$.	$\delta \times \sqrt{2\pi}\exp\left\{\left[\Phi^{-1}(p_2)\right]^2/2\right\}$
Measures based on information-theoretic divergence			
7. Symmetrized chi-squared divergence	$X(p_1,p_2) = \frac{1}{2}(p_1-p_2)^2\left[\frac{1}{p_1}+\frac{1}{p_2}+\frac{1}{q_1}+\frac{1}{q_2}\right]$	No: M=+∞ can be attained for $\|p_1-p_2\|<1$, e.g., if $p_1=0$.	$\delta^2 \times \frac{1}{p_2 q_2}$
8. Jeffreys divergence	$J(p_1,p_2) = \frac{1}{2}(p_1-p_2) \times \ln(\mathrm{OR})$	No: M=+∞ can be attained for $\|p_1-p_2\|<1$, e.g., if $p_1=0$.	$\delta^2 \times \frac{1}{2p_2 q_2}$
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	$\delta^2 \times \frac{1}{4p_2 q_2}$
10. Rescaled Jensen-Shannon divergence[f]	$S_2(p_1,p_2) = \frac{1}{\ln 2}\left\{\left[\frac{p_1\ln p_1 + p_2\ln p_2}{2} - p_*\ln p_*\right] + \left[\frac{q_1\ln q_1 + q_2\ln q_2}{2} - q_*\ln q_*\right]\right\}$	Yes: M=1	$\delta^2 \times \frac{1}{(8\ln 2)p_2 q_2}$
11. Hellinger distance	$H(p_1,p_2) = \sqrt{1-\sqrt{p_1 p_2}-\sqrt{q_1 q_2}}$	Yes: M=1	$\delta \times \frac{1}{(2\sqrt{2})\sqrt{p_2 q_2}}$

[a]All 11 measures satisfy properties 1–3 and 5–7, and equal zero if and only if the two proportions p_1 and p_2 are equal. Given the two proportions p_1 and p_2, we define: $q_1=1-p_1$, $q_2=1-p_2$, $p^*=(p_1+p_2)/2$, and $q^*=1-p^*=(q_1+q_2)/2$; $\mathrm{OR}=(p_1/p_2)/(q_1/q_2)$; $\mathrm{logit}(p)=p/(1-p)$; and $\mathrm{probit}(p)=\Phi^{-1}(p)$ where $\Phi(x)$ is the standard normal distribution.

[b] $2 \times D_2$ appears in the classical test of the null hypothesis that $p_1=p_2$. Additionally, $D_2=\sqrt{\Delta}$; see Web Appendix 1. 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 $\pi \times h$; see (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 $0 \times \ln 0 = 0$, $S_2(0,0)$ and $S_2(1,1)$ are resolved by continuity to 0.