

James P. Scanlan
Attorney at Law
1529 Wisconsin Avenue, NW
Washington, D.C. 20007
(202) 338-9224
jps@jpscanlan.com

October 8, 2015

ELECTRONICALLY TRANSMITTED

With corrections of Oct. 12, 2015
[Table of Contents](#)

David R. Morganstein, President
Jessica Utts, President-Elect
Jeri Metzger Mulrow, Vice-President
James L. Rosenberger, Vice-President
Robert L. Santos, Vice-President
Ronald L. Wasserstein, Executive Director
Stephen Pierson, Director of Science Policy
Kenneth C. Land, Chair, Committee on Law and Justice Statistics
Jerome P. Reiter, Chair,
American Statistical Association
732 North Washington Street
Alexandria, VA 22314-1943

Re: Recommendation That ASA (a) Form a Committee to Address the Unsoundness of Statistical Analyses of Demographic Differences That Fail to Consider Ways That Measures Tend to Be Systematically Affected by the Frequency of an Outcome and (b) Advise Arms of the U.S. Government That Reducing the Frequency of an Outcome Tends to Increase Relative Differences in Outcome Rates

Dear President Morganstein and other Members of the Leadership of the American Statistical Association:

This letter has two purposes. One purpose, which involves what could be a long-term undertaking by the American Statistical Association (ASA), is to urge the organization to form a committee to explore the ways analyses by statisticians and others of demographic and other differences in outcome rates are fatally undermined as a result of the failure to recognize patterns by which standard measures of differences between outcome rates tend to be systematically affected by the frequency of an outcome.

A second purpose of the letter, which involves a matter that ASA can address immediately, is to urge the organization to formally advise arms of the United States government that a statistical belief underlying important civil rights law enforcement policies is patently

incorrect. In particular, many policies are based on the belief that reducing the frequency of an adverse outcome will tend to reduce relative differences between rates at which advantaged and disadvantaged groups experience the outcome. In fact, the opposite is the case. While reducing the frequency of an outcome tends to reduce relative differences in rates of avoiding the outcome, it tends to increase relative differences in rates of experiencing the outcome.

The following are two examples of the perverse consequences of the federal government's misunderstanding of this issue. For at least two decades the government has encouraged lenders to relax mortgage lending standards in order to reduce relative racial differences in adverse borrower outcomes like rejection of mortgage loan applications. For at least several years, the government has encouraged public schools to relax discipline standards in order to reduce relative racial and other differences in adverse discipline outcomes like suspension and expulsion. But relaxing standards and thereby generally reducing adverse outcomes, while tending to reduce relative differences in the corresponding favorable outcomes, tends to increase relative differences in the adverse outcomes. Unaware that reducing the frequency of an outcome tends to increase relative difference in rates of experiencing it, the government continues to monitor the fairness of lending and discipline practices on the basis of relative differences in adverse outcomes. Thus, by complying with government encouragements to relax standards, lenders and public schools increase the chances that the government will sue them for discrimination.

This letter arises from discussions in 2012 with ASA Director of Science Policy Stephen Pierson regarding my suggestion that ASA advise Congress and government agencies of fundamental problems in the government's analyses of demographic differences. In those discussions, Dr. Pierson advised that the appropriate course to secure ASA action of the type I suggested would be to contact the relevant ASA committees. Those discussions also led to publication of my Statistician's View column in the December 2012 issues of *Amstat News* titled "[Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies](#)."¹ That item, which explained statistical patterns I had previously described (among many other places since 1987) in an article in the Fall 1994 issues of *Chance* titled "[Divining Difference](#)," and a guest editorial in the Spring 2006 issue of *Chance* titled "[Can We Actually Measure Health Disparities?](#)," gives a fairly succinct summary of the civil rights law enforcement anomalies mentioned above as of late 2012.²

¹ To facilitate consideration of issues raised in letters such as this I include links to referenced materials in electronic copies of the letters. I do not include links with every mention of an item, but attempt to include them often enough to make online items readily available without the reader's having to search for an earlier link. Electronic copies of letters such as this are available by means of the [Institutional Correspondence](#) subpage of the [Measuring Health Disparities](#) page of [jpscanlan.com](#). In this case, electronic copies of the letter are being emailed to recipients mentioned in the text of the letter. Since the online copy of this letter may be corrected, I include here a [link](#) to the most recent version (which, if containing corrections, will indicate on the first page when it was last corrected).

² Briefer treatments of the key concept addressed in the *Amstat News* column may be found in "[The Paradox of Lowering Standards](#)," *Baltimore Sun* (Aug. 5, 2013); "[Things government doesn't know about racial disparities](#)," *The Hill* (Jan. 28, 2014); and "[It's easy to misunderstand gaps and mistake good fortune for a crisis](#)," *Minneapolis StarTribune* (Feb. 8, 2014).

More recent and more extensive treatments of those anomalies and related matters may be found in my [“Race and Mortality Revisited,”](#) *Society* (July/Aug. 2014); [“The Perverse Enforcement of Fair Lending Laws,”](#) *Mortgage Banking* (May 2014); [“Measuring Health and Healthcare Disparities,”](#) Federal Committee on Statistical Methodology 2013 Research Conference (March 2014) (FCSM paper); [“The Mismeasure of Discrimination,”](#) Faculty Workshop, University of Kansas School of Law (Sept. 2013) (Kansas Law paper); and *amicus curiae* [brief](#) in *Texas Department of Housing and Community Affairs et al. v. The Inclusive Communities Project, Inc.*, Sup. Ct. No. 13-1371 (Nov. 2014) (TDHCA brief).³

Listed as addressees of the letter, in addition to Dr. Pierson and ASA Executive Director Ronald L. Wasserstein, are the President, President-Elect and three Vice-Presidents of ASA, along with the Chairs of the Committee on Law and Justice and the Scientific and Public Affairs Advisory Committee.⁴ Given that the issues raised in the letter apply to a wide range of ASA and ASA member activities, I will be emailing the letter to the leadership and members of various committees and sections. As with other recent letters of this nature, I will also post it on ASA Connect.

As a preliminary matter, I note that ASA President David R. Morganstein may wish to consider whether it would be appropriate to recuse himself from any ASA role in consideration of the issues raised in this letter. President Morganstein is a Vice President and Director of the statistical consulting firm Westat. Pursuant to U.S. Department of Education, Office Special Education Programs grants, Westat, through the IDEA Data Center, has produced guides on the

³ Recent, extended treatments of these issues may also be found in the following methods workshops at American universities: [“The Mismeasure of Discrimination,”](#) Center for Demographic and Social Analysis, University of California, Irvine (Jan. 20, 2015); [“The Mismeasure of Demographic Differences in Outcome Rates”](#) Public Sociology Association of George Mason University (Oct. 18, 2014); [“Rethinking the Measurement of Demographic Differences in Outcome Rates,”](#) Maryland Population Research Center of the University of Maryland (Oct. 10, 2014); [“The Mismeasure of Association: The Unsoundness of the Rate Ratio and Other Measures That Are Affected by the Prevalence of an Outcome,”](#) Minnesota Population Center and Division of Epidemiology and Community Health of the School of Public Health of the University of Minnesota (Sept. 5, 2014); [“The Mismeasure of Group Differences in the Law and the Social and Medical Sciences,”](#) Institute for Quantitative Social Science at Harvard University (Oct. 17, 2012); [“The Mismeasure of Group Differences in the Law and the Social and Medical Sciences,”](#) Department of Mathematics and Statistics of American University (Sept. 25, 2012).

⁴ This letter may be deemed a follow-up to earlier communications with certain of the recipients, including a September 29, 2014 email to members of the ASA Committee on Law and Justice, an October 2, 2014 email to the ASA Scientific and Public Affairs Advisory Committee, and a September 30, 2014 email to Dr. David Morganstein (prior to his serving as ASA President and solely regarding his role at Westat). In addition to advising the recipients of the issues now being addressed in this letter (and, in the cases of the two ASA committees, that I would likely be addressing the issues with ASA), the letters invited the recipients to attend (or, in the case of Dr. Morganstein, to have Westat staff attend) an October 10, 2014 methods [workshop](#) I was giving on these issues at the Maryland Population Center of the University of Maryland. A September 16, 2014 letter to President-Elect Jessica Utts inviting her to attend my January 20, 2015 [workshop](#) on these issues at the Center for Demographic and Social Analysis of the University of California, Irvine also advised of my intention to encourage ASA to take a role in advising Congress and government agencies on these issues. More information on these and the other workshops mentioned in the following note may be found in note 3 *supra*.

measurement of “significant disproportionality” and “significant discrepancies” regarding certain special education matters. I specifically discuss these guides in materials referenced below, including the [IDEA Data Center Disproportionality Guide](#) and [Disabilities – Public Law 104-446](#) subpages of the [Discipline Disparities](#) page of [jpscanlan.com](#), as reflecting the same types of failures of understanding that are the subject of this letter (or as involving certain additional interpretive issues given brief attention in Section A.3 of the letter).⁵ Further, as indicated immediately below, in August 2014 I sent the IDEA Data Center a letter similar to this letter discussing problems in one of its measurement guides. I also followed up with several emails to the organization. The points I make in this letter call into question the utility of the IDEA Data Center measurement guides and potentially other products in which Westat has an interest.⁶

This letter follows on a number of similar letters advising various types of governmental and nongovernmental entities of the ways analyses of demographic or other differences in outcome rates that the entities conduct, fund, or provide guidance on – or that in some manner pertain to the entities’ activities – are undermined by the failure to recognize patterns by which measures of differences between outcome rates tend to be systematically affected by the frequency of an outcome. Recipients of such letters since 2009 include [Robert Wood Johnson Foundation](#) (Apr. 8, 2009), [National Quality Forum](#) (Oct. 22, 2009), [Institute of Medicine](#) (June 1, 2010), [The Commonwealth Fund](#) (June 1, 2010), [United States Department of Education](#) (Apr. 18, 2012), [United States Department of Justice](#) (Apr. 23, 2012), [Board of Governors of the Federal Reserve System](#) (Mar. 4, 2013), [Harvard University](#) (Oct. 9, 2012), [Harvard Medical School, Massachusetts General Hospital, et al.](#) (Oct. 26, 2012), [Senate Committee on Health, Education, Labor and Pensions](#) (Apr. 1, 2013), [Mailman School of Public Health of Columbia University](#) (May 24, 2013), [Investigations and Oversight Subcommittee of House Finance Committee](#) (Dec. 4, 2013), [Education Trust](#) (April 30, 2014), [Annie E. Casey Foundation](#) (May 13, 2014), [Institute of Medicine II](#) (May 28, 2014), [IDEA Data Center](#) (Aug. 11, 2014), [Education Law Center](#) (Aug. 14, 2014), [Financial Markets and Community Investment Program, Government Accountability Office](#) (Sept. 9, 2014), [Wisconsin Council on Families and Children’s Race to Equity Project](#) (Dec. 23, 2014), [Portland, Oregon Board of Education](#) (Feb.

⁵ I have given the IDEA Data Center technical assistance guide “[Methods for Assessing Racial/Ethnic Disproportionality in Special Education](#)” particular attention in a number of items discussed or referenced in this letter, including the TDHCA [brief](#) (at 23-27), the September 8, 2015 [letter](#) to the Chief Data Scientist of the White House Office of Science and Technology Policy (at 9, n.13), the August 31, 2015 [letter](#) to the McKinney, Texas Independent School District (at 8), the August 24, 2015 [letter](#) to the Department of Health and Human Services and the Department of Education (at 3, 8-11), the April 7, 2015 [letter](#) to Texas Appleseed (at 3-4), the January 20, 2015 University of California, Irvine [workshop](#) (slides 55-59), the October 18, 2014 George Mason University [workshop](#) (slides 98-108), the October 10, 2015 University of Maryland [workshop \(slides 98-108\)](#), and the September 5, 2015 University of Minnesota [workshop](#) (slides 62-63). I have also brought certain of these materials to the attention of Department of Education personnel overseeing the pertinent IDEA Data Center grant by emails of August 25, 2015, and August 12, 2014.

⁶ Points made in this letter call into question most standard analyses of demographic differences in outcome rates. Thus, there may be a basis for recusal of other members of ASA leadership should they have interests in products employing such analyses. But I have no knowledge of such interests other than that noted with respect to President Morganstein.

25, 2015), [Vermont Senate Committee on Education](#) (Feb. 26, 2015), [United States Department of Justice and City of Ferguson, Missouri](#) (Mar. 9, 2015), [Senate Committee on Health, Education, Labor and Pensions II](#) (Mar. 20, 2015), [Texas Appleseed](#) (Apr. 7, 2015), [City of Minneapolis, Minnesota](#) (June 8, 2015), [Agency for Healthcare Research and Quality](#) (July 1, 2015), [Department of Health and Human Services and Department of Education](#) (Aug. 24, 2015), [McKinney, Texas Independent School District](#) (Aug. 31, 2015), and [Chief Data Scientist of the Office of Science and Technology Policy](#) (Sept. 8, 2015). [Comments](#) I posted on May 16, 2014, regarding Federal Drug Administration proposed subgroup regulations could be deemed a similar communication to that agency.⁷

These letters collectively reveal how widespread, even among institutions and organizations that are universally esteemed for their statistical and scientific expertise, is the failure to understand patterns by which measures of differences between outcome rates tend to be affected by the frequency of an outcome and how fatal that failure is to crucial activities of those institutions and organizations. I have posted most recent letters on ASA Connect, which has commonly led to large numbers of reviews of the letters by ASA members. So there should currently be an understanding of these issues, both in general and as they bear on the situations of various recipient entities, among many ASA members.

A number of those letters are of particular pertinence to the subjects of this letter. Particularly pertinent to both subjects of this letter is the September 8, 2015 [letter](#) to DJ Patil, the Chief Data Scientist of the White House Office of Science and Technology Policy. Like this letter, the letter to Dr. Patil, while addressing the need to give attention to broader problems in statistical analyses arising from the failure to understand patterns by which measures tend to be affected by the frequency of an outcome, urges Dr. Patil to cause the federal government to immediately cease leading the public and entities covered by civil rights law erroneously to believe that reducing the frequency of an outcome will tend to reduce relative differences in rates of experiencing the outcome. The letter also stresses that the continued failure of the

⁷ In recent letters such as this, I usually list only formal letters to other entities since 2009. I note, however, that for more than two decades I have communicated by letter or email with many individuals or organizations regarding issues of the type addressed in this letter. Some of these communications (and responses thereto) are pertinent to the recognition by the National Center for Health Statistics (NCHS), discussed *infra*, of the pattern whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it. Among these are (a) a November 28, 1988 [letter](#) to Surgeon General David Satcher and Department of Health and Human Services Assistant Secretary for Planning and Evaluation Margaret A. Hamburg; (b) a January 20, 1999 [letter](#) from NCHS Director Edward J. Sondick; and (c) a January 25, 1999 [letter](#) to Dr. Sondick. The exchange with Dr. Sondick is discussed or alluded to in my "[Race and Mortality](#)," *Society* (Jan./Feb. 2000) (at 6), and "[Race and Mortality Revisited](#)," *Society* (July/Aug. 2014) (at 332)). Email exchanges of special pertinence to the subject of this letter include a number of exchanges with NCHS statistician Dr. Kenneth, B. Keppel, beginning in 2001, that led directly to the formal recognition by NCHS that determinations of whether health and healthcare disparities were increasing or decreasing would commonly turn on whether one examines relative differences in favorable outcomes or relative differences in the corresponding adverse outcomes. I have also had many other email exchanges with officials or staff of other federal agencies involved with health and healthcare disparities research or oversight of clinical trials.

government to understand the simple fact that reducing the frequency of an outcome tends to increase, not decrease, relative differences in rates of experiencing it may undermine public confidence in the government's expertise regarding more complex matters.

Particularly pertinent to the broader subject of this letter is the July 1, 2015 [letter](#) to the Agency for Healthcare Research and Quality (AHRQ). The letter explains that the health and healthcare disparities research the agency conducts or funds has been universally undermined by the failure, on the parts of both AHRQ and recipients of AHRQ funding, to recognize the ways measures employed in such research tend to be affected by the frequency of an outcome. The letter also explains that, apparently as a result of the agency's failure to understand certain measurement issues, AHRQ has failed to measure health and healthcare disparities in the manner it believed it was doing, with the resultant anomaly that the National Healthcare Disparities Report would find substantial decreases in disparities between point A and point B, even though it would also find those disparities to be substantially larger at point B than point A. See discussion in Section A.3 *infra*.

Particularly pertinent to the narrower subject of this letter are three recent letters involving actions or proposed actions by federal government entities based on beliefs about the consequences of those actions that are the opposite of reality. The March 9, 2015 [letter](#) to the Department of Justice (DOJ) and the City of Ferguson, Missouri addresses the fact that findings in the DOJ's March 4, 2015 report titled "[Investigation of the Ferguson Police Department](#)" that police and court procedures of Ferguson, Missouri had a disparate impact on the city's African American residents is based on the mistaken premise that reducing the frequency of adverse interactions between the police/courts and the city's residents would tend to reduce the proportion African Americans make up of persons subject to those interactions. The report adds another area of federal law enforcement to the list of situations where an entity's compliance with implied or explicit guidance of the government will tend to increase the chances that the government will accuse the entity of discrimination.⁸

The March 20, 2015 [letter](#) to the Senate Committee on Health, Education, Labor and Pensions discusses that the Keep Kids in School Act that was introduced in the Senate in March 2015 is based on the mistaken premise that reducing public school suspensions will tend to reduce relative racial/ethnic differences in suspension rates. That bill creates potential for the enactment of a law requiring actions that, although specifically aimed at reducing relative differences in an adverse outcome, will tend to increase those differences. See "[Race and Mortality Revisited](#)" (at 343) regarding a provision in existing law, the Individuals with

⁸ Both the DOE/Ferguson letter and the letter to the Department of Health and Human Services and Department of Education discussed *infra* add a complicating element to this subject by mainly casting demographic disparity issues in terms of comparisons between the proportion a group makes up of persons potentially experiencing an outcome and the proportion the group makes up of persons actually experiencing the outcome. I explain *infra*, as I have previously illustrated in Table 1 of the 1994 *Chance* [article](#) and Table 1 of the 2006 *Chance* [editorial](#), that reducing the frequency of an outcome tends to increase the proportions groups most susceptible to the outcome make up of persons experiencing the outcome (as well as of persons failing to experience the outcome).

Disabilities Education Act, that since 2004 has mandated responses to large relative differences in certain adverse outcomes in schools receiving federal funds that are of a type that tends to increase such differences. See also the [Disabilities – Public Law 104-446](#) subpage of the [Discipline Disparities](#) page.

The August 24, 2015 [letter](#) to the Department of Health and Human Services (HHS) and Department of Education (DOE) discusses that the [Policy Statement on Expulsion and Suspension Policies in Early Childhood Settings](#) that HHS and DOE jointly issued in December 2014 is based on the mistaken premise that generally reducing preschool suspensions and expulsions will tend to reduce the proportions disadvantaged groups make up of persons experiencing those outcomes. The Policy Statement adds HHS to the list of federal agencies leading the public and entities covered by federal civil rights laws to believe things that are the opposite of reality and the opposite of what the National Center for Health Statistics, an arm of HHS, recognized to be the case more than a decade ago. In addition to calling for action that will tend to increase the large relative differences that are prompting the agencies' action, the Policy Statement reflects the incongruous, but not uncommon, situation in which the government gives special attention to a matter based on large relative demographic differences in an adverse outcome, while failing to recognize that the relative differences are so large precisely because the outcome is so rare.⁹

Section A below discusses the near universal unsoundness of analyses of demographic and other differences between outcome rates as a result of the failure to recognize patterns by which measures of those differences tend to be systematically affected by the frequency of the outcome and recommends that ASA form a committee to address the matter. Section B addresses the narrower subject of the government's affirmatively leading the public and entities covered by civil rights laws erroneously to believe that reducing the frequency of an outcome will tend to reduce relative differences in rates of experiencing the outcome or the proportion groups most susceptible to the outcome make up of persons experiencing it. While the sections seek different actions from ASA, they are closely related. One important connection is that the facts (a) that federal law enforcement policy based on a statistical belief that was not simply incorrect, but the exact opposite of reality, could persist for decades, and (b) that the belief could go entirely unquestioned by the statistical and social science communities, compellingly illustrate the need for ASA to educate its own members, other members of the scientific and research communities, the public, and the federal government regarding certain fundamental statistical concepts.

⁹ See pages 1 and 3-5 of the HHS/DOE letter regarding the agencies' unsupported statement that the outcome rates at issue are "high."

A. The General Unsoundness of Analyses of Differences Between Outcome Rates of Advantaged and Disadvantaged Groups and the Need for the American Statistical Association to Form a Committee to Address the Situation

1. Patterns by Which Measures of Differences Between Outcome Rates Tend to Be Systematically Affected by the Frequency of an Outcome

There are four principal measures by which statisticians and others appraise differences in rates at which advantaged and disadvantaged groups experience favorable or adverse outcomes: (1) relative (percentage) differences between rates of experiencing the outcome; (2) relative differences between rates of avoiding the outcome (*i.e.*, experiencing the opposite outcome); (3) absolute (percentage point) differences between the outcome rates; and (4) odds ratios. None of these measures provides a sound basis for quantifying differences in the circumstances of advantaged and disadvantaged groups reflected by their outcome rates (or, otherwise put, the strength of the forces causing the groups' outcome rates to differ) because, for reasons inherent in the underlying risk distributions, each measure tends to be systematically affected by the frequency of an outcome.¹⁰

a. The Two Relative Differences

Relative differences in rates of experiencing and avoiding an outcome tend to be affected by the frequency of an outcome in the following manner. The rarer an outcome the greater tends to be the relative difference between rates at which advantaged and disadvantaged groups experience the outcome and the smaller tends to be the relative difference between rates at which such groups avoid the outcome.¹¹ Thus, for example, as mortality and poverty decline, relative differences in experiencing those outcomes tend to increase while relative differences in avoiding them tend to decrease. As rates of appropriate healthcare increase, relative differences in failing to receive such care tend to increase while relative differences in receipt of such care tend to decrease. Relaxing mortgage lending criteria or public school discipline standards tends to increase relative differences in adverse borrower and discipline outcomes while reducing relative differences in the corresponding favorable outcomes. Interventions to reduce foreclosure rates will tend to increase relative differences in foreclosure rates while reducing relative differences in rates of avoiding foreclosure. Generally improving academic performance (or lowering standards) will tend to increase relative differences in (a) failing to achieve proficiency, (b) retention in grade, and (c) failing to graduate, while reducing relative differences in (a') achieving proficiency, (b') avoiding retention in grade, and (c') graduating.

¹⁰ I use the terms "frequency" and "prevalence," sometimes modified by "overall," essentially interchangeably. Lately I have principally used "frequency." See Section A.8 of the [Scanlan's Rule](#) page of [jpscanlan.com](#) regarding precision issues pertaining to "prevalence." The points made there would also apply to "frequency."

¹¹ A more precise description of the pattern would state, rather than "the rarer an outcome," "the more the outcome is restricted toward either tail of the overall distribution." But I have characterized the pattern in the manner done in the text above for some time and those discussing it have not been confused by the usage. Thus, I am not at this time inclined to depart from the usage in the text.

Similarly, in populations or settings where adverse outcomes are comparatively rare (*e.g.*, persons with high education or high income, the young, British civil servants,¹² highly-qualified job applicants, the countries of Norway and Sweden or the states of Minnesota and Massachusetts) relative demographic differences in adverse outcomes tend to be larger, while relative differences in the corresponding favorable outcomes tend to be smaller, than in populations or settings where adverse outcomes are comparatively common. To put the point in the context of the subject of some of the recent letters to government entities, relative demographic differences in adverse public school discipline outcomes tend to be larger, though relative differences in avoiding those outcomes tend to be smaller, in (a) suburban schools, (b) preschool, and (c) schools without zero tolerance policies, where those outcomes are less common than in (a') urban schools, (b') K-12, and (c') schools with zero tolerance policies, than they are in the latter settings.

One manifestation of the above-described pattern of relative differences that is of special pertinence to the subjects of the above-discussed DOJ/Ferguson and HHS/DOE letters – and something I sometimes term a corollary to the pattern of relative differences described above – is that reducing the frequency of an outcome tends to increase the proportions groups most susceptible to the outcome make up of (a) persons experiencing the outcome and (b) persons failing to experience the outcome.¹³

Another manifestation of the pattern whereby the two relative differences change as the frequency of an outcome changes – and another that I sometimes term a corollary to the pattern of relative differences – is that a factor that affects outcome rates for two groups with different baseline rates for the outcome will tend to cause a larger proportionate change in the rate for the

¹² Readers unfamiliar with the Whitehall Studies or their findings may be puzzled by mention of British civil servants. Thus, I note that the Whitehall Studies of health disparities among British civil servants play an important role in the history of health disparities research, and interpretations of implications of the studies' findings that relative differences in adverse outcomes are greater among British civil servants than in the United Kingdom population at large exemplify the impossibility of soundly interpreting data on demographic differences without understanding the patterns described in this letter. See "[Race and Mortality Revisited](#)" (at 340), my 2006 British Society for Populations Studies paper "[The Misinterpretation of Health Inequalities in the United Kingdom](#)," and the [Whitehall Studies](#) subpage of the [Measuring Health Disparities](#) page of [jpscanlan.com](#).

¹³ The pattern whereby reducing the frequency of an adverse outcome tends to increase the proportions disadvantaged groups make up of persons experiencing it and failing to experience it could just as well be termed the cause of the pattern of relative differences. For it is the fact that the disadvantaged group will make up a larger proportion of persons below and above each increasingly lower cut point on a continuum that underlies the described pattern of relative differences. See Table 1 of the 2006 *Chance* [editorial](#). My earliest treatments of this subject principally focused on perceptions about disproportionate representations of certain groups among persons experiencing an adverse outcome. See "[The 'Feminization of Poverty' is Misunderstood](#)," *Plain Dealer* (Nov 11, 1987); "[An Issue of Numbers](#)," *National Law Journal* (Mar. 5, 1990); and "[The Perils of Provocative Statistics](#)," *Public Interest* (Winter 1991).

group with the lower baseline rate while causing a larger proportionate change in the opposite outcome rate for the other group.¹⁴

These patterns can be easily illustrated with normally distributed test score data. Table 1 below, which is a version of Table 1 of "[Race and Mortality Revisited](#)" (at 329) as adapted for Table 2 of the Chief Data Scientist [letter](#), and which reflects the same hypothetical described in the *Amstat News* column, is based on a situation where the means of normal test score distributions of an advantaged group (AG) and a disadvantaged group (DG) differ by half a standard deviation and both distributions have the same standard deviation. In addition to showing the pass and fail rates of each group, the table shows the ratio of AG's pass rate to DG's pass rate and the ratio of DG's fail rate to AG's fail rate at each cutoff (the first pair of shaded columns, with tan shading in the electronic copy of the letter).¹⁵ Based on a situation where AG and DG each make up half of the test takers, the final columns (shaded red in the electronic copy of the letter) show the proportion DG makes up of those who pass and those who fail at each cutoff.

¹⁴ The pattern whereby reducing the frequency of an outcome (a) tends to increase relative differences between rates of experiencing the outcome at the same time that it (b) tends to reduce relative differences between rates of avoiding the outcome may initially seem counterintuitive. In fact, however, (b) is implied in (a), if in fact (b) is not exactly the same thing as (a). For if reducing the frequency of an outcome tends to increase relative differences in rates of experiencing the outcome, it necessarily follows that increasing the frequency of an outcome tends to reduce relative differences in rates of experiencing the outcome. And if one outcome declines in frequency (hence, tending to increase relative differences as to that outcome), it necessarily follows that the opposite outcome increases in frequency (hence, tending to reduce relative differences as to that outcome).

The same point can be made more simply with regard to the contrasting pattern of proportionate changes. For if a factor affecting an outcome tends to cause a larger proportionate change for the group with the lower baseline rate, it necessarily follows that the factor will tend to cause a large proportionate change in the opposite outcome for the other group, since the other group has the lower baseline rate for the opposite outcome.

¹⁵ While I commonly refer to patterns of relative differences in this letter, the table actually presents rate ratios. The relative difference is the rate ratio minus 1 where the rate ratio is above 1 and 1 minus the rate ratio where the rate ratio is below one. One should be careful not to mistakenly refer to the rate ratio as the relative difference. But the distinction between the two terms is not pertinent to the discussion here of patterns by which relative differences tend to be affected by the frequency of an outcome. In recent years I commonly present the rate ratios for both outcomes with the larger figure in the numerator, in which case, as to both outcomes, the larger the rate ratio, the larger the relative difference. In the 1994 *Chance* [article](#) and the 2006 *Chance* [editorial](#) I used the disadvantaged group's rate as the numerator in both ratios (which is the approach of the "four-fifths" or "80 percent" rule for identifying disparate impact under the Uniform Guideline for Employee Selection Procedures, see note 18 *infra*) yielding a rate ratio for the favorable outcome that was below 1. Choice of numerator in the ratio, however, has no bearing on the patterns by which the two relative differences tend to be affected by the frequency of the outcome.

Table 1. Illustration of effects on relative differences in pass and fail rates of lowering a cutoff from a point where 80% of AG passes to a point where 95% of AG passes, with proportions DG comprises of persons who pass and of persons who fail (when mean scores differ by approximately half a standard deviation and DG comprises 50% of test takers)

Cutoff	AG Pass	DG Pass	AG Fail	DG Fail	AG/DG Pass Ratio	DG/AG Fail Ratio	DG Prop of Pass	DG Prop of Fail
High	80%	63%	20%	37%	1.27	1.85	44%	65%
Low	95%	87%	5%	13%	1.09	2.60	48%	72%

According to the specifications underlying the table, at the cutoff where 80% of AG passes the test, approximately 63% of DG would pass the test (with corresponding failure rates of 20% for AG and 37% for DG). The ratio of AG's pass rate to DG's pass rate would be 1.27 while the ratio of DG's fail rate to AG's fail rate would be 1.85.

When the cutoff is lowered to the point where the pass rate for AG is 95%, the pass rate for DG would be approximately 87% (with corresponding failure rates of 5% for AG and 13% for DG). The ratio of AG's pass rate to DG's pass rate would thus decrease to 1.09 (from 1.27), while the ratio of DG's fail rate to AG's fail rate would increase to 2.60 (from 1.85). That is, the relative difference in the outcome that was reduced in frequency (test failure) increases, while the relative difference in the opposite outcome (test passage, which increased in frequency) declines.

It warrants note at this point that the pattern whereby lowering a test cutoff tends to reduce relative differences in pass rates is well known. It underlies the universal belief that lowering test cutoffs tends to reduce the disparate impact of employment and other tests where some groups outperform others as well as the requirement that employers justify how high they have set a test cutoff.¹⁶ But even though the fact that lowering a cutoff will tend to increase relative differences in failure rates is implied in the widely known fact that lowering the cutoff tends to reduce relative differences in pass rates (see note 14 *supra*), at least so far as the published record reveals, the fact that lowering a cutoff tends to increase relative differences in failure rates is virtually unknown. That seems to be the case even among agencies like the Departments of Justice and Education that have been dealing with issues concerning demographic differences in testing outcomes for decades.

As shown in the final two columns of Table 1, lowering the cutoff and reducing the frequency of test failure caused an increase in the proportion DG makes up of those who pass the test (from 48% to 52%) and the proportion DG makes up of persons who fail the test (from 65% to 72%). Because the proportion DG makes up of persons taking the test is unaffected by the

¹⁶ Whether relaxing a standard in fact reduces the standard's impact, properly measured, is a complex subject. See Section E (at 27-32) of the Kansas Law [paper](#) mentioned in the introduction.

cutoff, lowering the cutoff would increase both the relative difference and the absolute difference between the proportion DG makes up of test takers and the proportion it makes up of persons who fail (while reducing both the relative difference and the absolute difference between the proportion DG makes up of test takers and the proportion it makes up of persons who pass).

One can also divine from the table that lowering the cutoff caused a larger proportionate decline in failure rates for AG (the group with the lower baseline failure rate) while causing a larger proportionate increase in the pass rate for DG (the group with the lower baseline pass rate). That is, lowering the cutoff caused failure rates to decrease by 75% for AG but only 65% for DG, while causing pass rates to increase by 38% for DG but only 19% for AG. Raising the cutoff back to the original point would similarly show a larger proportionate increase in the failure rate for AG but a larger proportionate decrease in the pass rate for DG, thus causing the relative difference in failure rates to decrease and the relative difference in pass rates to increase.

Table 1 can also illustrate the pattern whereby relative differences in the adverse outcome tend to be larger, while relative differences in the corresponding favorable outcome tend to be smaller, in advantaged populations/settings than in disadvantaged populations/setting. The reader need simply regard the second row as reflecting the situation in the advantaged population/setting (where the adverse outcome is less common) and the first row as reflecting the situation in the disadvantaged population/setting (where the adverse outcome is more common). In terms of the point of the preceding paragraph, this means that being in the advantaged population/setting tends to reduce adverse outcome rates proportionately more for advantaged groups than disadvantaged groups while increasing favorable outcome rates proportionately more for disadvantaged groups than advantaged groups. To make the last point more concrete, having high income or high education tends to reduce adverse outcome rates proportionately more for whites than disadvantaged racial minorities, while increasing favorable outcome rates proportionately more for disadvantaged racial minorities than whites.

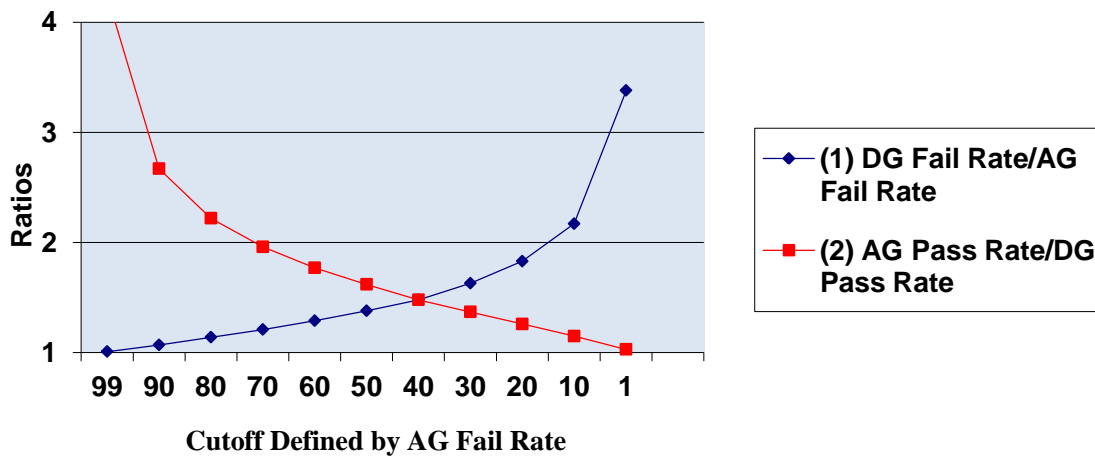
It warrants note at this juncture that, irrespective of the pattern of proportionate changes in outcome rates just described, the widespread notion that in the normal course (*i.e.*, absent what would be termed a subgroup effect, effect heterogeneity, or interaction) one would expect a factor that affects an outcome rate to cause equal proportionate changes for groups with different baseline outcome rates is illogical as well as unsound. For if a factor causes two groups to experience equal proportionate changes in different baseline rates for one outcome, it necessarily causes the groups to experience unequal proportionate changes in their rates for the opposite outcome. Since there is no more reason to expect equal proportionate changes in one outcome than there is to expect equal proportionate changes in the opposite outcome, it is illogical to expect equal proportionate changes in either outcome.

For similar reasons, it is illogical to regard the rate ratio as a sound measure of association. For anytime the rate ratio is the same for different pairs of outcome rates (*e.g.*, a rate ratio of 2.0 for underlying rates of 20% and 10% and for underlying rates of 10% and 5%), the rate ratios will necessarily be different for the opposite outcomes rates (*i.e.*, a rate ratio of 1.125

for rates of 80% and 90% and a rate ratio of 1.06 for rates of 90% and 95%).¹⁷ See "Race and Mortality Revisited" (at 339-341) and the [Illogical Premises](#), [Illogical Premises II](#), [Subgroup Effects](#), [Subgroup Effects – Nonclinical](#), and [Inevitability of Interaction](#) subpages of the [Scanlan's Rule](#) page.¹⁸

The pattern of relative differences in pass and fail rates shown in Table 1 exists across the full range of test scores. Figure 1, which employs the same specifications as Table 1, shows the effects on the two relative differences of lowering a cutoff from a point where almost everyone fails to a point where almost every passes. The relative difference in the decreasing outcome (test failure) consistently increases, while the relative difference in the increasing outcome (test passage) consistently decreases.

Figure 1. Ratios of (1) DG Fail Rate to AG Fail Rate, (2) AG Pass Rate to DG Pass Rate, (3) DG Failure Odds to AG Failure Odds; and (4) Absolute Difference Between Rates



Similar patterns can be found in virtually any data that show the proportions of advantaged and disadvantaged groups falling above and below various points on a continuum of factors associated with experiencing some favorable or adverse outcome. For example, income and credit score data show that the lower the income or credit score, the greater tends to be the relative racial difference in falling below it and the smaller tends to be the relative racial difference in falling above it. National Health and Nutrition Survey Data show that generally

¹⁷ See note 15 *supra* regarding numerator issues.

¹⁸ See also the [Four-Fifths Rule](#) subpage of the [Disparate Impact](#) page of [jpscanlan.com](#) regarding the different strengths of association (reasonably measured), and different rate ratios for the adverse outcome, in situations where the favorable outcome rate of the disadvantaged group is 80% of the favorable outcome rate of the advantaged group (the benchmark for identifying disparate impact under the Uniform Guideline for Employee Selection Procedures), at different frequencies for the favorable outcome.

lowering blood pressure will tend to increase relative racial differences in hypertension while reducing relative racial differences in rates of avoiding hypertension or that generally improving folate levels will tend to increase relative racial difference in low folate while reducing relative racial differences in rates of adequate folate.¹⁹

The patterns can also be observed in data on rates at which advantaged and disadvantaged groups experience favorable and adverse outcomes within advantaged and disadvantaged populations or settings.²⁰ For example, life tables show that relative demographic differences in mortality tends to greater among the young than the old, while relative differences in survival tend to be greater among the old than the young.²¹ Home Mortgage Disclosure Act data show that relative differences in mortgage rejection rates tend to be greater, while relative differences in mortgage approval rates tend to be smaller, among higher-income than lower-income mortgage loan applicants.²² Income data show that relative differences in poverty rates of single-parent and married-couple families tend to be greater, while such differences in rates of avoiding poverty tend to smaller, among whites than among blacks, or that, correspondingly, relative differences between black and white poverty rates tend to be larger, while such differences in rates of avoiding poverty tend to be smaller, among married-couple families than among single-parent families.²³

Tables 2 through 4 below, which are based on Tables 4 (at 334), 8 (at 342), and 7 (at 341) of "Race and Mortality Revisited," present illustrations of the patterns described above along with some information on how observers, unaware of or ignoring those patterns, would interpret the data. For simplicity, the tables present only the outcome rates commonly reported (from which opposite outcome rates can be inferred)

Table 2 is based on data from a 2008 study in *Pediatrics* that examined of a school-entry Hepatitis B vaccination requirement on racial and ethnic differences in vaccination rates. The table presents vaccination rates for white and black fifth graders in the year before and the year in which the requirement was imposed, along with rate ratios for receipt and non-receipt of vaccination.

¹⁹ See [Collected Illustrations](#) subpage of Scanlan's Rule page of [jpscanlan.com](#)

²⁰ Examples abound even though the strength of the forces causing the outcome rates of advantaged and disadvantaged groups to differ may vary greatly from population/setting to population/setting. See Section A.2 *infra*.

²¹ See [Life Tables Illustrations](#) subpage of the [Scanlan's Rule](#) page of [jpscanlan.com](#). See also Table 1 of the [Mortality and Survival](#) page of [jpscanlan.com](#) (regarding racial differences in cancer mortality and survival at different ages) and [Table 6](#) (slide 14) of my 16th Nordic Demographic Symposium (2008) presentation "[Measures of Health Inequalities That are Unaffected by the Prevalence of an Outcome](#)" (regarding relative differences in mortality and survival of renters and homeowners at different ages).

²² See the [Disparities – High Income](#) subpage of the [Lending Disparities](#) page of [jpscanlan.com](#).

²³ See data cited in my "[The Perils of Provocative Statistics](#)," *Public Interest* (Winter 1991).

Table 2. Hepatitis B vaccination rates for white and black fifth graders before and after implementation of school-entry vaccination requirement, with ratio of white to black vaccination rates and ratio of black to white rates of no vaccination

Period	White Vac Rate	Black Vac Rate	White/Black Vac Ratio	Black/White No Vac Ratio
Pre Impl Year	8%	3%	2.67	1.05
Impl Year	46%	33%	1.39	1.24

Consistent with the patterns described above, the general increases in vaccination rates presumably caused by the requirement reduced relative differences in vaccination rates while increasing relative differences in rates of failing to be vaccinated. The authors, relying on the former relative difference, found that the requirement dramatically reduced disparities. The National Center for Health Statistics, which, as discussed in Section A.3 *infra*, would have relied on relative differences in non-receipt of vaccination to measure disparities, would have found a substantial increase in disparities.

The underlying table in "[Race and Mortality Revisited](#)" (Table 4, at 334) also shows that to the extent that we can effectively measure it, the strength of the forces causing outcome rates to differ declined (something that it is reasonable to expect following imposition of a mandatory requirement). In order to simplify this letter, I will leave discussion of that method to "Race and Mortality Revisited" (at 336-337) and other materials discussed in Section A.3. But I note that the reduction in the strength of the forces causing the outcome rates to differ was insufficient to cause an increase in the relative difference in the adverse outcome not to be observed. The same holds for the increase in the absolute difference (from 5 to 13 percentage points), which is the subject of Section A.1.b *infra*.²⁴

Table 3 is based on data on racial differences in preschool suspensions released by the DOE in March 2014 concerning a situation that would then be regarded as involving enormous disparities (and that would later be cited in the HHS/DOE [Policy Statement](#) (at 4) discussed in the recent HHS/DOE letter). The table presents white and black rates of multiple suspensions in preschool and in K-12, along with rate ratios for experiencing multiple suspensions and avoiding multiple suspensions. The table shows the common pattern whereby in the setting where the adverse outcomes is less common (preschool), the relative difference in the adverse outcome is larger, while the relative difference in the corresponding favorable outcome is smaller, than in the setting where the adverse outcome is more common (K-12).²⁵

²⁴ More complete information on the results of this study may be found in Table 4 (at 334) of "Race and Mortality Revisited" and Table 7 (at 20) of the FCSM [paper](#) and the [tables](#) accompanying my February 23, 2010 [Second Comment on Morita Pediatrics 2008](#).

²⁵ I used the multiple suspension figures in the table in "Race and Mortality Revisited" because figures on multiple suspensions in the DOE's March 2014 [Data Snapshot: Early Childhood Education](#) were commonly cited in reportage of the perceived enormity of the disparities (though cast in terms of a comparison of the proportion

Table 3. White and black rates of multiple suspensions in preschool and K-12, with ratio of black to white multiple suspension rates and ratio of white to black rates of avoiding multiple suspensions

Level	White Rate Multi Suspension	Black Rate Mult Suspensions	Black/White Ratio Mult Suspensions	White/Black Ratio No Mult Suspension
Preschool	0.15%	0.67%	4.41	1.01
K12	2.23%	6.72%	3.01	1.05

The underlying table in "Race and Mortality Revisited" (Table 8, at 342) also shows that to the extent that we can measure the strength of the forces causing black and white multiple suspension rates to differ, whatever those forces may be, it is approximately the same in preschool as in K-12. But, as noted above, I leave discussion of the pertinent method to "Race and Mortality Revisited" and other materials.

Table 4 is based on a study of the effects of having a criminal record on the callback rates of tester pairs of job applicants comprised of two black or two white job applicants with fabricated resumes that were comparable in all respects but for the indication of a criminal conviction. In this case, the advantaged and disadvantaged groups are applicants without a conviction and applicants with a conviction, while white and black applicants are the advantaged and disadvantaged populations. The outcome rates shown are the favorable outcome rates (rates of receiving a callback), and the favorable and adverse outcome ratios are the (a) ratios of receiving a callback of the applicant without a conviction to that of the applicant with a

African Americans made up of preschoolers with the proportion they made up of persons with multiple suspensions.) As in the HHS/DOE Policy Statement, figures on multiple suspensions were presumably cited because they reflect a greater perceived disparity than found in analyses of any (*i.e.*, one or more) suspensions. The latter, involving a more common outcome than multiple suspensions, will typically show a smaller relative difference in experiencing the outcome, but a larger relative difference in avoiding the outcome, than analyses of multiple suspensions. See discussion in Section A.3 *infra* regarding misperceptions based on the comparative size of relative differences in rates of experiencing increasingly more severe outcomes in the criminal justice system.

See also the [Intermediate Outcomes](#) subpage the Scanlan's Rule page of jpscanlan.com regarding the impossibility, even with appropriate tools (such as that discussed in "Race and Mortality Revisited" (at 336-337) or like measure that is not affected by the frequency of an outcome), of analyzing data on rates of experiencing single suspensions, as distinguished from one or more suspensions. That is, one can soundly analyze data on rates at which advantaged and disadvantaged groups fall above and below a particular point (*e.g.*, in a school setting, rates of falling below grades A, B, C, or D, or the corresponding opposite outcomes); but one cannot soundly analyze rates at which such groups receive Bs, Cs, or Ds. For the latter rates would be functions of the proportions of the advantaged and disadvantaged groups falling into both more favorable categories and less favorable categories. Analyses of rates of falling into intermediate categories are not widespread. But one finds them in discussions of school discipline issues (as in the DOE preschool data and the suspension rates data discussed in the Intermediate Outcomes subpage) and criminal justice issues, and occasionally in discussion of things involving mortality relative to time from event (*e.g.*, postneonatal mortality and the maternal mortality categories discussed on the subpage), as well in discussions of moderately low birthweight and certain types of coronary care.

conviction, and (b) the ratio of not receiving a callback of the applicant with a conviction to that of the applicant without a conviction.

Table 4. White and black rates of receiving callbacks for testers with and without convictions indicated on application, with ratio of no conviction to conviction callback rates and ratio of conviction to no conviction rates of no callback

Race	Conviction Callback Rate	No Conviction Callback Rate	No Conv/Conv Callback Ratio	Conv/No Conv No Callback Ratio
White	17%	34%	2.00	1.26
Black	5%	14%	2.80	1.10

The table shows the common pattern described above where the factor that affects an outcome rate causes a larger proportionate change for the group with the lower baseline rate (blacks) while causing a larger proportionate change in the opposite outcome for the other group.²⁶ Focusing on the relative effect on receipt of a callback (the favorable outcome), the author found that the effect of having a criminal record was greater for blacks than for whites and considered possible explanations for the differing effect. But had the author instead examined the effect of a criminal record on the adverse outcome, she would have found a greater relative effect for whites than blacks.

The table in "Race and Mortality Revisited" (Table 7, at 341) shows that to the extent we can measure the strength of the effect of having a criminal conviction on treatment of the applicants by the employers, the effect was essentially equal for blacks and whites. The study may be contrasted with a later study by the same author that is discussed on the [Criminal Record Effects](#) subpage of the Scanlan's Rule page. In the case of the later study, a sound measure would find a substantially larger effect of a criminal record on the treatment of blacks than whites. But it requires a sound measure, or at least an understanding of the patterns described here, to effectively distinguish between the results of the two studies.

Two things about tester studies warrant mention here. There is a tendency for observers to rely on the larger of the two relative differences, with or without giving thought to the other relative difference, but typically without recognizing even the possibility, much less the likelihood, that the other relative difference would support an opposite conclusions.²⁷ Had the

²⁶ Consistent with my preference for always using the higher rate in the numerator of the rate ratio (see note 15 *supra*) I show the effect on the favorable outcome in terms of the ratio of the callback rate of the advantaged group (applicants without a conviction) to the callback rate of the disadvantaged group (applicants with a conviction). So the table shows that having no conviction increases the favorable outcome proportionately more for blacks than whites, which is the same thing as that having a conviction reduces the favorable outcome proportionately more for blacks than whites.

²⁷ The tendency to rely on the larger of the two relative differences sometimes causes researchers to rely both on relative differences in favorable outcomes and on relative differences in adverse outcomes in the same study. See the [Immune Disparities](#) page of [jpscanlan.com](#) regarding a study that examined immunization disparities in terms of relative differences in the adverse outcome for receipt/non-receipt of any immunization and relative

fabricated credentials of the tester pairs been much greater, or had the labor market been much tighter, the callback rates could have been in ranges where the relative effects on the adverse outcome could have been larger than the relative effects on the favorable outcome. The author might then have relied on the relative effect on the adverse outcome and reached opposite results.

Further, in the case of any sort of tester study, the creators of the study, by deciding whether the fabricated backgrounds of the testers should make a particular outcome frequent or infrequent, will substantially dictate the pattern of relative differences that a particular underlying effect, or lack of comparability of members within a tester pair, will yield.²⁸ See my [Comment on Schulman NEJM 1999](#) (at 4) regarding a 1999 *New England Journal of Medicine* tester study of racial and gender differences in recommendation of cardiac catheterization in which, by fabricating symptom scenarios where the overwhelming majority of physicians would recommend cardiac catheterization, the authors largely dictated that any modest physician bias, or modest lack of comparability of actors describing the symptom scenarios, would result in small relative differences in cardiac catheterization recommendations and large relative differences in rates of failing to recommend cardiac catheterization.²⁹

These points are not indictments of the good faith of tester studies.³⁰ But they do illustrate some of the anomalies that can arise from reliance on a measure that is fundamentally

differences in the favorable outcome for receipt/non-receipt of full immunization and the [McKinsey Achievement Gap Study](#) subpage of the [Educational Disparities](#) page regarding a study where the authors relied on the relative difference in the adverse outcome for reaching/failing to reach the basic proficiency level but on the relative difference in the favorable outcome for reaching/failing to reach the advanced proficiency level. In both cases the authors relied on the larger of the two relative differences with respect to each of the subjects examined and without recognizing that general increases in favorable outcome would tend to increase relative differences in the adverse outcomes while reducing relative differences in the favorable outcome.

²⁸ While not pertinent to any to any important point in this letter, I note that the tester study discussed in the text, because members of a tester pair were of the same race, was able to obviate matching issues respecting the two members of each pair by alternating which of the two members of the pair would be associated with the resume showing a conviction.

²⁹ For reasons discussed in the next section, one could similarly predict that the results of that particular study would show small absolute differences between rates and large differences measured by odds ratios (which is the measure on which the authors relied). However, in general, the relationship between the degree to which fabricated backgrounds call for a particular outcome and the size of absolute differences and odds ratios is more complicated than the relationship between such factor and the two relative differences.

³⁰ There are many difficult problems in tester studies. Apart from matching issues, it often will be difficult or impossible to divine an appropriate numerator and denominator for tester outcomes rates (see my "[Measuring Hiring Discrimination](#)," *Labor Law Journal* (July 1993)) making it difficult or impossible to interpret results even with a sound measure. In the NEJM study, some of the fabricated scenarios might have elicited automatic catheterization recommendations by all examining physicians 100 percent of the time and other scenarios might have elicited automatic catheterization recommendations by a high proportion the examining physicians 100 percent of the time. Catheterization recommendation rates for each group within the situations where serious judgment was actually exercised by the examining physicians would likely be very different from those based on the entire universe of physician recommendations. But there exist many like issues in non-tester contexts.

unsound. There are many other situations not involving testers where researcher decisions pertaining to the frequency of the outcome examined or degree of changes in the frequency of an outcome will similarly largely dictate the findings that will be yielded by the measure on which the researchers have chosen to rely.

Numerous other examples of the above-described patterns of relative differences in publicly available data may be found in the materials referenced in Section A.3 *infra*. But, notwithstanding the many situations where these patterns are evident, with the rare exceptions discussed on the [Consensus](#) subpage of the Scanlan's Rule page, observers relying on some relative measure to quantify a demographic difference, or otherwise to appraise a situation reflected by a pair of outcome rates, have done so without recognizing that the patterns described above exist or of the implications of those patterns. The same holds for the above-described related approaches to appraising demographic differences, *i.e.*, (a) comparisons of the proportion a group makes up of persons potentially experiencing an outcome with the proportion it makes up of persons actually experiencing the outcome and (b) comparisons of proportionate changes in outcome rates undergone by different groups over time or as the result of some intervention or some general change in the frequency of an outcome.

Indeed, as discussed in the *Amstat New* column and the above-discussed Chief Data Scientist, DOE/Ferguson, Senate Committee on Health, Education, Labor and Pensions, and HHS/DOE letters (and dozens or scores of other places, including all of the earlier letters to federal government entities), the government believes that reducing the frequency of an outcome will tend to reduce, not increase, relative differences in experiencing it. That is, the government does have a notion about the relationship between the frequency of an outcome and relative differences in rates of experiencing it. But, in the case of adverse outcomes, the notion is the exact opposite of reality.

And, as reflected by the scores or hundreds of scholarly and non-scholarly articles observing that “despite” a general decline in mortality or some other adverse outcome, relative demographic differences either “remain” or “have increased,” the mistaken notion that reducing the frequency of an outcome will or should tend to reduce relative differences in rates of experiencing it is shared by a substantial proportion of trained and untrained persons analyzing and discussing demographic differences in the law or the social and medical sciences.³¹ The pervasiveness of this mistaken notion is also reflected in the fact that the government's

³¹ The failure to understand that reducing the frequency of an adverse outcome tends to increase relative differences in rates of experiencing it is somehow pervasive even among the great numbers of persons with statistical training (perhaps the overwhelming majority of such persons) who readily understand that as a favorable outcome is increasingly restricted toward the right tail of the overall distribution, relative differences in rates of experiencing the outcome increase. For reasons mentioned above (see note 14 *supra*) that understanding should compel an understanding that reducing the frequency of an adverse outcome will tend to increase relative differences in experiencing it. It may be, however, that the correct understanding as to favorable outcomes contributes to the misunderstanding as to adverse outcomes, just as the understanding that lowering a test cutoff tends to reduce relative differences in pass rates may contribute to the mistaken assumption that relaxing lending and discipline standards will tend to reduce relative differences in adverse borrower and discipline outcomes.

longstanding efforts to lead the public and entities covered by federal civil rights laws to believe that reducing the frequency of an outcome will tend to reduce relative differences in rates of experiencing it has yet to be questioned either by the research community or by the entities whom the government accuses of discrimination on the basis of large relative differences in adverse outcomes.

b. Absolute Differences and the Odds Ratios

Appraisals of differences in the circumstances of advantaged and disadvantaged groups reflected by their outcome rates in terms of absolute (percentage point) differences between rates and differences measured by odds ratios are unaffected by which outcome one examines. But in order for a measure to effectively quantify the strength of the forces causing outcome rates to differ, it must remain unchanged as there occurs a general change in the frequency of an outcome akin to that effected by the lowering of a test cutoff. And, like the two relative differences, absolute differences and odds ratios also tend to be affected by the frequency of an outcome, though in a more complicated way than the two relative differences.

Roughly, as uncommon outcomes (below 50% for both groups) become more common, absolute differences between rates tend to increase; as common outcomes (above 50% for both groups) become even more common, absolute differences tend to decrease. The frequency-driven direction of change of the absolute difference is harder to predict when the outcome is neither common nor uncommon or changes from being uncommon to common (or vice-versa) during a period examined.

The absolute difference and both relative differences may all change in the same direction as the frequency of an outcome changes (in which case the difference measured by the odds ratio will also change in the same direction as the other measures). But in the common situation where all measures do not change in the same direction as the frequency of an outcome changes, the absolute difference will tend to change in the same direction as the smaller relative difference. Since observers who rely on relative differences to appraise the difference in the circumstances of two groups reflected by their differing outcome rates tend usually to rely on the larger of the two relative differences (as discussed above), there exists a systematic tendency for observers relying on the absolute difference to reach opposite conclusions about directions of changes in the size of demographic differences, or about the comparative size of the differences in different settings, from observers relying on a relative difference.

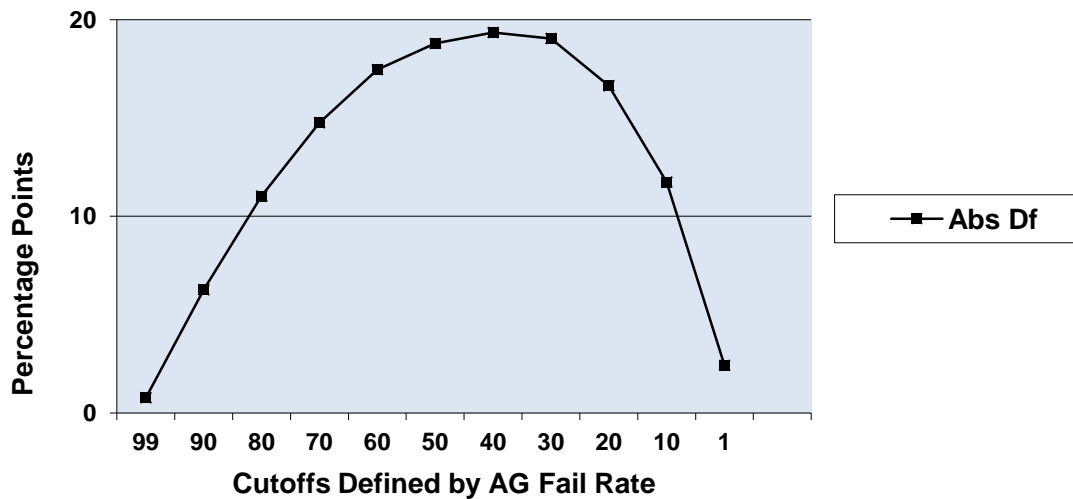
Further, anytime an observer notes that the absolute difference has changed in a different direction from the relative difference the observer happens to be examining, the unmentioned relative difference will necessarily have changed in the opposite direction of the mentioned relative difference and the same direction as the absolute difference. Thus, observers maintaining that one must make a value judgment in choosing between the relative difference and the absolute difference with respect to the appraisal of a particular change in some demographic disparity over time have already made a choice (usually without thought) to

rely on the relative difference that yields an opposite conclusion from the absolute difference rather than the relative difference that yields the same conclusion as the absolute difference.

As the frequency of an outcome changes, and all measures do not change in the same direction, the difference measured by the odds ratio tends to change in the opposite direction of the absolute difference and in the same direction as the larger relative difference.

These patterns are illustrated in Figures 2 and 3, which are based on the same specifications as Figure 1, and similarly show the implications of lowering a cutoff from a point where almost everyone fails to a point where almost everyone passes. Figure 2 presents the pattern for the absolute difference. Figure 3 presents the pattern for the ratio of the disadvantaged group's odds of failing the test to the advantaged group's odds of failing the test.³² But in order to illustrate the relationship with the two relative differences, a line for the odds ratio is simply added to the lines for the two relative differences previously shown in Figure 1.

Figure 2. Absolute differences between rates of AG and DG pass (or fail) rates at various cutoff points defined by AG fail rate



³² There are four odds ratios depending on which outcome is examined and which group's odds is used as the numerator. Two yield one value and two yield a value that is the reciprocal of the first value. The ratio of DG's failure odds to AG's failure odds is the same as the ratio of AG's pass odds to DG's pass odds. Thus, the ratios shown are also the ratios of AG's odds of passing the test to DG's odds of passing the test.

Figure 3. Ratios of (1) DG fail rate to AG fail rate, (2) AG pass rate to DG pass rate, (3) DG failure odds to AG failure odds

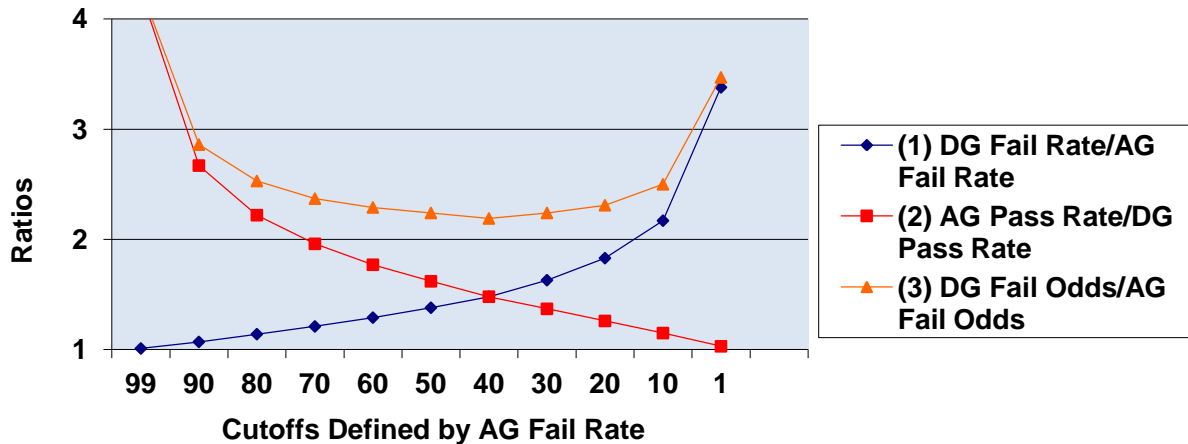


Table 5 below, which is based on the same specifications as Table 1 and Figures 1 to 3, presents favorable outcome rates at four frequency levels, along with rate ratios for the favorable and adverse outcomes as well as absolute differences and odds ratios. The parenthetical numbers indicate the ways one would rank the size of the disparity between the situation of the advantaged and disadvantaged groups according to each of the four measures. Those numbers show that rankings according to relative differences in favorable outcomes are the opposite of rankings according to relative differences in adverse outcomes. They also show that rankings according to absolute differences and odds ratios, while different from the rankings according to either relative difference, are the opposite of each other.

Table 5. Favorable outcome rates of advantaged group (AG) and disadvantage group (DG) at four settings with different favorable outcome frequencies, with measures of difference

Setting	AG Fav Rate	DG Fav Rate	AG/DG Fav Ratio	DG/AG Adv Ratio	Abs Df (Perc Pnts)	Odds Ratio
A	20.0%	9.0%	2.22 (1)	1.14 (4)	11.0 (4)	2.53 (1)
B	40.0%	22.6%	1.77 (2)	1.29 (3)	17.4 (2)	2.28 (3)
C	70.0%	51.0%	1.37 (3)	1.63 (2)	19.0 (1)	2.24 (4)
D	80.0%	63.4%	1.26 (4)	1.83 (1)	16.6 (3)	2.31 (2)

A version of this table appears as Table 5 (at 335) of "Race and Mortality Revisited," where, as in many other places, it is used to refute the notion that choice of a measure in analyses of demographic differences in health or healthcare outcomes involves a value judgment. I will rest here on the treatment of that issue in "Race and Mortality Revisited" (at 335-336) and other places (save for some points made below in Section A.2). I include the table here principally to illustrate certain points about the absolute difference in contexts where observers commonly rely on that measure. The premise of the table, as reflected by its specifications, is that there is no

rational basis to maintain that the strength of the forces causing the favorable (or adverse) outcome rates of the advantaged and disadvantaged groups to differ varies among the rows.

Areas of research or commentary where observers commonly or increasingly rely on absolute differences between rates include appraisals of demographic differences in academic proficiency and healthcare. In the case of academic proficiency, irrespective of any change in the forces causing outcome rates of advantaged and disadvantaged groups to differ, general improvements in proficiency scores will tend to increase absolute differences between rates at which such groups reach the advanced level (where favorable outcome rates tend to be well below 50%) as reflected by movement from Row A to Row B in Table 5. On the other hand, general improvements in scores will tend to reduce such differences between rates of achieving the basic level (where favorable outcome rates are often well above 50%), as illustrated by movement from Row C to Row D in Table 5. General deterioration in test performance will tend to have the opposite effect. Like patterns will tend to be observed when proficiency standards are altered or a proficiency test is replaced with one that is easier or harder.

Similarly, increases in rates of appropriate healthcare will tend to increase absolute differences between rates of advantaged and disadvantaged groups for uncommon procedures/outcome, while reducing absolute differences for common procedures/outcomes. Such patterns are reflected by movement from Row A to Row B for the former procedures/outcomes and from Row C to Row D for the latter procedures/outcomes. When healthcare outcomes increase from being fairly uncommon to being very common (as in the case of certain vaccines and screening practices) the absolute differences will tend to increase for a time and then decrease.

Further, higher-performing hospitals (which tend to have generally higher appropriate care rates than lower-performing hospitals) will tend to show larger absolute differences between appropriate care rates than lower-performing hospitals for procedures/outcomes where rates are generally low, while showing smaller absolute differences for procedures/outcomes where rates are generally high. This pattern, too, is illustrated in Table 5 if Rows A and B are regarded as the lower- and higher-performing hospitals with respect to procedures/outcomes with generally low rates and Rows C and D are regarded as such hospitals with respect to procedures/outcomes with generally high rates.

Odds ratios would tend to show patterns that are the opposite of those just described for absolute differences. In situations where observers draw initial conclusions about such things as changes in the size of healthcare disparities over time on the basis of absolute differences between rates, and then employ logistic regression to adjust for the role of possible confounders, the odds ratio yielded by the adjustment will tend to show a pattern that is the opposite of that shown by the absolute difference.

Observers relying on absolute differences, however, have yet to recognize the ways absolute differences tend to be affected by the frequency of an outcome or that other measures would tend to systematically yield opposite results (*i.e.*, the larger relative difference and the

odds ratio) or consistent results (*i.e.*, the smaller relative difference). Commonly observers relying on absolute differences, like those relying on one of the two relative differences, show no awareness that choice of measure is of any consequence.

An extreme example of the confusion in this area may be found in the following situation. A 2004 *American Journal of Public Health* (AJPH) study examined changes in racial disparities in certain fairly uncommon procedures among Medicare beneficiaries between 1986 and 1997, a period in which the procedures were generally increasing. As commonly occurs in the circumstances, the results showed that, in the main, relative difference in receipt of procedures decreased while absolute differences (and relative differences in non-receipt) increased. Relying on relative differences between rates of receiving such procedures to measure disparities, a common approach at the time, the authors found that disparities generally decreased. A 2005 study in the *New England Journal of Medicine* (NEJM) examined changes in racial disparities in similarly uncommon procedures among Medicare beneficiaries between 1992 and 2001, also a period when rates were generally increasing. The results were much like those of the 2004 AJPH study, *i.e.*, usually decreasing relative differences in receipt of procedure and increasing absolute differences (and relative differences in non-receipt of the procedures). In this case, however, because the authors relied on absolute differences between rates to measure disparities, they found disparities usually to be increasing.³³ A 2008 article in *Medical Care Research and Review* then discussed the contrasting findings of directions of changes in the two studies, and without consideration of the role of choice of measure in the reportage of results, or showing any awareness of the measures used in the two studies, then opined that differing conclusions in the studies may have had to do with the absence of complete overlap of the time periods studied. See the [Spurious Contradictions](#) subpage of the [Measuring Health Disparities](#). See also the discussion in "Race and Mortality Revisited" (at 333) of an AHRQ-funded review of the effects of improvement in healthcare on healthcare disparities that gave no consideration to the measures employed in the 4,258 studies it examined.

The government and most researchers examining racial differences in public school discipline usually measure discipline disparities in terms of relative differences in adverse

³³ I discuss the 2005 NEJM study in "Race and Mortality Revisited" (at 338), and quite a few other places, as one of two studies in the same issue of the journal that relied on absolute differences to measure racial disparities in healthcare. The other study, however, examined comparatively common outcomes and found absolute differences usually to be decreasing for those outcomes. Neither study, nor a commentary discussing their differing results, recognized the role of the frequency of the outcomes examined on absolute differences or the way that relative differences in favorable outcomes (a common approach to healthcare disparities measurement at the time) or relative differences in adverse outcomes (an approach adopted by National Center for Health Statistics in 2004) would have yielded findings consistent or inconsistent with those in the studies. See the above-mentioned October 9, 2012 [letter](#) to Harvard University (at 34-35). Two studies on healthcare disparities in a recent issue of the NEJM are by groups of authors that include co-authors of one of the 2005 studies ((a)Trivedi AN, Nsa W, Hausmann LRM, et al. Quality and equity of care in U.S. hospitals. *N Engl J Med* 2014;371:2298-308. (b) Ayanian JZ, Landon BE, Newhouse JP, Zaslavsky AM. Racial and ethnic disparities among enrollees in Medicare Advantage plans. *N Engl J Med* 2014;371:2288-97.) These studies also rely on absolute differences, again without consideration of the effects of frequency of the outcomes or mention that other measures might yield different conclusions.

outcomes or differences between the proportion a group makes up of the student population and the proportion it makes up of persons experiencing the adverse discipline outcome. But some researchers analyze discipline disparities in terms of absolute differences. Researchers doing so commonly (though not invariably, see the [NEPC National Study](#) subpage of the Discipline Disparities page) reach opposite conclusions about directions of changes over time or the comparative size of discipline disparities in different settings from researchers relying on relative differences in discipline rates, typically without recognition that the relative measure would yield a different conclusions. One recent study that relied on absolute differences did offer a justification for that approach (see Losen et al., “[Are We Closing the Discipline Gap](#)” (2015) (at 48)) though still without reflecting an awareness of the way any measure tends to be systematically affected by the frequency of an outcome or discussing that most observers, including the government, would be relying on the relative difference in discipline rates (or the difference between the proportion a group makes up of students and the proportion it makes up of persons disciplined) and commonly reaching opposite conclusions from those based on absolute differences.

It used to be the case that demographic differences in poverty and unemployment were almost always appraised in relative terms – with observers’ examining the less common outcome of the favorable/adverse dichotomy (which would involve the larger relative difference). Thus, decreases in poverty and unemployment would tend to be associated with increasing demographic disparities and increases in poverty and unemployment would tend to be associated with reduced disparities (or the corresponding larger proportionate changes in outcome rates for the groups with lower baseline rates). See my “[The Perils of Provocative Statistics](#),” *Public Interest* (Winter 1991).

Recently, however, observers had been examining changes in unemployment and poverty disparities in terms of absolute differences between rates or the comparative size of absolute changes in each group’s rate. Such approach will tend to yield opposite conclusions from those based on relative differences in adverse outcomes or the size of various groups’ relative changes in adverse outcomes. See the discussion of the “suburbanization of poverty” on the [Feminization of Poverty](#) page of [jpscanlan.com](#)³⁴ and the illustration pertaining to the

³⁴ The “suburbanization of poverty” theme and the “feminization of poverty” theme both involve the misguided examination of a phenomenon that is a function of two quite different matters: (a) changes in the proportion of a group falling into some category with (b) changes in the outcome rates of groups in the category and groups not in the category. But pertinent to the subject of this letter is that in discussions of the feminization of poverty, one fact that was deemed to reflect a worsening of the situation of single-parent families compared with other persons was that, during a period of general declines in poverty, the relative difference between the poverty rates of male-headed and female-headed families increased. Reliance on the absolute difference, as is now being done in discussions of the suburbanization of poverty, would commonly have shown decreases in the poverty rate disparity between male-headed and female-headed families. See also discussion *infra* regarding (1) the consistency between (a) patterns of changes in the relative differences in rates of experiencing an outcome and (b) patterns of changes in the proportion disadvantaged groups make up or persons experiencing the outcome, and (2) the inconsistency between (a) patterns of changes in absolute differences in rates of experiencing an uncommon adverse outcome and (b) the proportion disadvantaged groups make up of persons experiencing the outcome.

2011 Center for American Progress study in Table 30 (slide 118) of the University of Maryland [workshop](#) referenced in note 3 *supra*.

The connection between patterns of changes in relative differences in rates of experiencing an outcome and the size of differences between the proportions the group more susceptible to the outcome makes up of persons potentially experiencing the outcome and actually experiencing the outcome is explained above and illustrated in Table 1. Because of that connection, at least where there are no substantial population shifts, and regardless of the rates ranges at issue for outcomes examined, observers relying on relative differences between rates of experiencing any outcome, will tend to reach the same conclusion about patterns of changes as observers relying on differences between the proportions disadvantaged groups make up of persons potentially experiencing the outcome (the pool) and persons actually experiencing the outcome. And that holds whether observers measure the difference between those two proportions in relative or absolute terms.³⁵ But in the case of the rate ranges at issue for things like poverty, unemployment, or school discipline, observers relying on absolute differences between rates will commonly reach opposite conclusions from observers relying on comparisons between the proportion a group makes up of the pool and the proportion it makes up of persons experiencing the outcome. See slides 103-104 of the Maryland workshop.

2. The Impossibility of Soundly Analyzing Differences Between Outcome Rates Without Consideration of the Ways the Measures Used Tend to Be Affected by the Frequency of the Outcome

The extent of scholarly consensus with my descriptions of patterns by which measures of differences between outcome rates tend to be systematically affected by the frequency of an outcome is summarized on the [Consensus Subpage](#) of the [Scanlan's Rule](#) page. Treatments of these issues by others have been largely limited to discussions of the pattern by which the two relative differences tend to be affected by the frequency of an outcome.³⁶ In general, such

³⁵ This point will generally hold in analyses of changes within districts over time, since periods examined usually are not long enough for major demographic shifts. But it often may not hold for comparisons across school districts. Districts can vary greatly in the proportion a group makes up of the pool, and that proportion will affect both relative and absolute differences between the proportion a group makes up of the pool and the proportion it makes up of persons experiencing the outcome in ways that are unrelated to the underlying outcome rates of the advantaged and disadvantaged groups. Moreover, the proportion a group comprises of the pool will tend to affect the relative difference between the two proportions and the absolute difference between the two proportions in different ways. See Table 24 (slide 108) of the University of Maryland [workshop](#) and the Addendum to the [IDEA Data Center Disproportionality Guide](#) subpage of the Discipline Disparities page.

³⁶ One exception is an article by Houweling TAJ, Kunst AE, Huisman M, Mackenbach JP ([Using relative and absolute measures for monitoring health inequalities: experiences from cross-national analyses on maternal and child health](#). *International Journal for Equity in Health* 2007;6:15). In responding to my "[Race and Mortality](#)," *Society* (Jan./Feb. 2000), while being unaware of the 2006 *Chance* [editorial](#), the authors independently identified the same pattern of correlation between the absolute difference and the frequency of an outcome that I had described in the *Chance* editorial.

treatments recognize that the two relative differences tend to behave in the manner I have described, while not necessarily describing the basis for the pattern in the same way I do.³⁷ But some treatments, including an internal National Center for Health Statistics document, suggest that the pattern may be ignored because it is not seen in all cases.

One will of course find many situations where observed patterns do not accord with the patterns described above. Observed patterns are functions of (a) the described frequency-related patterns and (b) the strength of the forces causing the outcome rates to differ in the settings being compared. The strength of those forces can vary greatly from setting to setting and can change substantially over time (especially longer periods of time) particularly during periods of dramatic changes in the frequency of an outcome. That can cause the frequency-related patterns not to be observed even though they are having a substantial effect.³⁸

The crucial consideration is that so long as measures tend to be in any manner affected by the frequency of an outcome, one cannot rely on the measures to appraise the strength of the forces causing rates to differ without consideration of the role of the frequency of the outcome. That various measures tend to exhibit the described patterns of changes, including that whereby the relative differences tend to change in opposite directions as the frequency of an outcome changes (and similarly contrasting patterns for the absolute difference and odds ratios) highlights the flaws in standard measures. But even if there existed only one measure, the fact that it is in some manner affected by the frequency of an outcome would undermine it as a means of quantifying the forces causing the outcome rates of two groups to differ.

As discussed above, "Race and Mortality Revisited" addresses the argument that two measures yielding different conclusions about such things as directions of changes over time can

³⁷ Since I am led to understand that some statisticians may best understand the basis for the described pattern of relative differences in terms of the density function, I note that two recent treatments of the pattern of relative differences with reference to the density function may be found in Lambert PJ, Subramanian S ([Disparities in Socio-Economic outcomes: Some positive propositions and their normative implications](#), Soc Choice Welf 2014;43:565-576), and Lambert PJ, Subramanian S ([Group inequalities and "Scanlan's Rule": Two apparent conundrums and how we might address them](#), Working Paper 84/2014, Madras School of Economics (2014)).

³⁸ Even when the comparative roles of the two factors are such that all standard measures change in the same direction, the frequency-related forces typically will have a role (driving certain measures in one direction and others in the opposite direction). Where the frequencies are dramatically different in the settings being compared, one will almost invariably find that the relative difference in experiencing the outcome is larger, while the relative difference in failing to experience the outcome is smaller, in the setting with lower rates for the outcome than in the setting with higher rates for the outcome, even when the comparative strength of the forces causing the rates to differ in each setting has a substantial countervailing effect. The matter is different for absolute differences and odds ratios, since as an outcome goes from being very rare to being very common, the directions of change of the measures reverse course. Thus, in the case of a very large increase in the frequency of an outcome, absolute differences and odds ratios may appear to have changed not at all when in fact they have changed a great deal in one direction and then changed the same amount in the opposite direction. See Table 1 of the 1994 *Chance* [article](#), which (unfortunately and perhaps misleadingly) involves a hypothetical change where the absolute difference is the same at both cutoff points.

both be in some manner valid and that a value judgment is involved in choosing between them. Inasmuch as those proposing the argument apparently did so while unaware of the patterns whereby measures tend to change as the frequency of an outcome changes, they ought to rethink the argument in light of those patterns (as well as knowledge of the simple fact that anytime a relative difference and absolute difference show different directions of changes over time, the other relative difference will necessarily have changed in the same direction as the absolute difference). At any rate, any reasonable discussion of the factors favoring one measure over another must be informed by an understanding of these patterns. And persons wed to a particular measure, whether rationally or not, ought to recognize that society has an interest in knowing whether the measure changed more than or less than (or in a different direction from) the pattern that would be observed if the change is entirely a function of a change in the frequency of the outcome. One can only make sound judgments about such issues, however, with an understanding of the patterns by which the measure tends to change as the frequency of the outcome changes.³⁹

Further, everyone ought to recognize that when there occurs a change in the various measures that is observably a consequence of changes in the frequency of the outcome, there can be no justification for devoting resources to studying the underlying causes of the observed pattern of change with respect to what may be the observer's preferred measure or any other measure. See discussion of Table 2 in "[Race and Mortality Revisited](#)" at 329-330, 343.

Similarly, even if there were a plausible argument that a value judgment is involved in choosing among measures to determine whether inequality in a particular situation has decreased or increased, there would be no conceivable argument that a value judgment is involved in drawing inferences about underlying processes, as in the case of subjects of the type addressed in the section of "Race and Mortality Revisited" (at 339-341) headed "Illogical Expectations and

³⁹ One recent exception to the failure of observers addressing value judgment issues to recognize patterns by which measures tend to change as the frequency of an outcome changes may be found in a recent *European Journal of Public Health* (EJPH) commentary by Johan Mackenbach (Should we aim to reduce relative or absolute inequalities in mortality? *Eur J Pub Health* 2015/3/28, <http://dx.doi.org/10.1093/eurpub/cku21709>) to which I respond in a comment titled "[Time to stop thinking in terms of relative inequalities and absolute inequalities.](#)" In the EJPH commentary, Mackenbach, who is coauthor of the article by Houweling *et al.* cited in note 16 *supra* and whose work I allude to in "Race and Mortality Revisited" (at 334), discusses the way the patterns by which measures change as the frequency of an outcome changes affects the value judgment argument. But he does so in a manner to suggest that the frequency-related patterns contribute to (or counter) the underlying inequality itself rather than that they indicate the unsoundness of the measures. Mackenbach does not seem to take a particular position respecting value judgments. But while focusing on the declines in the types of adverse health outcomes like mortality, where declines in the outcome typically reduce absolute differences but increase relative differences in those outcomes, Mackenbach concludes by observing that, while society should endeavor to reduce both relative and absolute differences, society should "count [its] blessings when only absolute inequalities go down." But society should not be pleased at the situation when absolute differences decline but do so to a smaller degree than they would tend to do solely on the basis of the decline in the outcome. For, in such circumstances, inequality should be deemed to have increased. Similarly, in the case of the increases in uncommon healthcare outcomes that commonly are attended by increasing absolute differences between rates, society should be pleased if the absolute differences increase to a smaller degree than they would tend to do solely on the basis of the general increase in the outcome.

Unfounded Inferences.” The same holds for efforts to apply results observed in a clinical trial to determine the number-needed-to-treat in situations involving different baseline rates from that in the trial. See [Subgroup Effects](#) subpage of the [Scanlan’s Rule](#) page.

3. Summary of Treatments of the Pertinent Pattern

“[Race and Mortality Revisited](#)” provides an extensive treatment of the above-described patterns and the broader implications of the failure of researchers and law enforcement officials to understand them. It explains that, to date, no one examining demographic differences has sought to determine whether observed patterns of changes in some measure of differences between outcome rates reflects something other than changes in the frequency of the outcome, and rarely, if ever, has anyone drawn a sound inference about underlying processes based on an examination of the size of some measure of difference in outcome rates in one setting compared with another. The article also suggests a method for appraising the strength of the forces causing a pair of outcome rates to differ that is theoretically unaffected by the frequency of the outcomes at issue. The article, while merely scratching the surface of many subjects, should alone provide a sufficient summary of the pertinent issues to persuade ASA leadership of the need for ASA to form a committee charged with reforming the analysis of demographic and other differences in outcome rates.

Nevertheless, in order that ASA leadership be as informed as possible about the pertinent issues in considering the recommendation in this letter, and that any committee formed pursuant to that recommendation also be as informed as possible about those issues, I summarize below certain additional materials.

Recent, extended discussions of the pertinent measurement issues with a particular focus on health and healthcare disparities research may be found in my Federal Committee on Statistical Methodology 2013 Research Conference paper “[Measuring Health and Healthcare Disparities](#),” as well as in the above-mentioned August 24, 2015 [letter](#) to the Agency for Healthcare Research and Quality and October 9, 2012 [letter](#) to Harvard University (each of which items contains further development of the ideas presented in the 2006 *Chance* editorial). Each item gives attention (as does “Race and Mortality Revisited” (at 331-335)) to the recognition, beginning in 2004, by the National Center for Health Statistics (NCHS) that determinations of directions of changes in health and healthcare disparities would commonly turn on whether one examines relative differences in the favorable outcome or relative differences in the corresponding adverse outcome. In doing so, each addresses the failure of understanding on the part of NCHS reflected by its manner of dealing with that recognition. Specifically, rather than regarding the pattern by which the two relative differences tend to change in opposite directions as the frequency of an outcome changes as calling into question the value of either relative difference for appraising health and healthcare disparities, the agency merely recommended that all disparities be measured in terms of relative differences in adverse outcomes. That action changed (a) a situation where, with healthcare disparities commonly measured in terms of relative difference in favorable outcomes (receipt of care), improvements in healthcare tended to be associated with decreasing healthcare disparities to (b) a situation where

improvements in healthcare tend to be associated with increasing disparities. But it left the crucial measurement issues unaddressed.

As explained in the HHS/DOE [letter](#) (at 12), NCHS is now proposing to measure healthcare disparities in terms of relative differences in favorable outcomes. That would return the situation to that existing prior to 2004. But it would still leave unaddressed the fact that neither relative difference provides useful information on whether the forces causing healthcare (or health) outcome rates of advantaged and disadvantaged groups to differ have increased or decreased over time and what factors contribute to, or mitigate, those forces.

Meanwhile, few persons at other federal agencies involved in health and healthcare research are aware of the conclusions NCHS reached about patterns of relative differences over a decade ago. So far as the published record reveals, no one at those agencies even recognizes that it is possible for the two relative differences to yield different conclusions as to the directions changes in health and healthcare disparities over time.⁴⁰

Each of the three items mentioned three paragraphs above also gives attention (as does "Race and Mortality Revisited" (at 337-339)) to the following anomaly in the Massachusetts Medicaid pay-for-performance (P4P) program. A perception that incentive programs would tend to increase healthcare disparities, which was based on a study finding that a general increase in a very uncommon procedure was associated with increased absolute differences between rates at which blacks and whites received the procedure, led to a call to include a healthcare disparities element in P4P programs. Neither the authors of the study, nor those responding to it by calling for disparities elements in P4P programs, gave attention to the fact that the observed increases in absolute differences were what commonly occurs when uncommon outcomes generally increase or to the fact that relative differences in receipt of care had decreased. Massachusetts responded to the call by including a disparities element in its Medicaid P4P program. The element appraises healthcare disparities through a measure that is a function of absolute differences between rates. But because the outcomes examined all involved very high overall rates (a situation where, as described with regard to Rows 3 and 4 of Table 5, higher rates tend to be associated with smaller absolute differences between rates), the program tends to systematically

⁴⁰ Section B of the AHRQ letter (at 15-23) clarifies the point in note 6 of "Race and Mortality Revisited" (at 333) and illustrated in Table 5 ([slide 21](#)) of the FCSM [presentation](#) regarding the apparent failure of the National Healthcare Disparities Report (NHDR) to measure healthcare disparities in the manner AHRQ believed the report was doing. The letter explains that, apparently without recognizing that it was doing anything of consequence, beginning in the 2010 NHDR, AHRQ changed the methodology for appraising changes in disparities over time from (a) one of quantifying changes in disparities in terms of percentage point changes in relative differences in adverse outcomes to (b) one of quantifying changes in disparities in terms of the comparative size of percentage point changes in the rates of the groups being compared. Because the NHDR continues otherwise to measure disparities in terms of relative differences in adverse outcomes, the change in method created situations where the NHDR would find substantial decreases in disparities between point A and point B, even though it would find substantially larger disparities at point B than at point A. Pages 23-24 of the letter explain that AHRQ may have recently reverted (seemingly inadvertently) to a method that would yield the same direction of changes as the pre-2010 methodology, though still quantifying the size of the changes differently from the earlier method.

favor higher-performing hospitals for reasons unrelated to the comparative racial equity from hospital to hospital. Since minorities tend to make up smaller proportions of patients at higher-performing hospitals than lower-performing hospitals, inclusion of the disparities element in the Massachusetts P4P program – by diverting resources away from providers with high proportions of minority patients – is more likely to increase healthcare disparities than to reduce them.⁴¹

The FCSM paper also gives attention (at 21-22) to the following anomaly, only touched upon in "Race and Mortality Revisited" (at 334) but treated at length in the [Mortality and Survival](#) page of [jpscanlan.com](#). Particularly in treatments of racial differences in cancer outcomes, studies commonly refer to relative differences in survival and mortality interchangeably, often stating that they are examining relative differences in survival while in fact examining relative differences in mortality. Invariably, they do so without recognizing that as survival generally increases, relative differences in survival tend to decrease while relative differences in mortality tend to increase, or that more survivable cancers tend to show smaller relative differences in survival but larger relative differences in mortality than less survivable cancers. In fact, in most cases where observers state something about demographic differences in survival in any context, they have in fact analyzed relative differences in mortality, while often reporting results that are the opposite of those one would find for survival.

Recent, extended discussions of measurement issues concerning analyses of disparate impact or other discrimination issues, including those addressed in the *Amstat News* column, may be found in my TDHCA [brief](#), my article "[The Perverse Enforcement of Fair Lending Laws](#)," *Mortgage Banking* (May 2014), and my September 20, 2013 University of Kansas School faculty workshop paper "[The Mismeasure of Discrimination](#)." See also my January 20, 2015 methods workshop at the University of California, Irvine's Center for Demographic and Social Analysis, which is also titled "[The Mismeasure of Discrimination](#)." These materials show that the failure to understand patterns by which measures between outcome rates tend to be affected by the frequency of an outcome has undermined virtually all analyses of discrimination issues involving outcome rates.

The TDHCA brief (at 23-27) and the Kansas Law paper (at 23-26) also address fatal flaws in analyses of discrimination issues in terms of the proportion that a group makes up of the population potentially experiencing an outcome (the pool) and the proportion it make up of persons experiencing the outcome – the approach of the Ferguson, Missouri disparities study discussed in the recent DOJ/Ferguson [letter](#) and of the HHS/DOE Policy Statement on preschool discipline discussed in the recent HHS/DOE [letter](#). In summary, with appropriate tools, one may be able to soundly appraise the strength of the forces causing a pair of outcome rates to differ.⁴²

⁴¹ There are other problems with the disparities measure employed in the Massachusetts Medicaid P4P program, and, at least in certain circumstances, these may cause the disparities element of the program to operate further to the disadvantage of hospitals where minorities make up high proportions of patients. See [Between Group Variance](#) subpage of the [Measuring Health Disparities](#) page of [jpscanlan.com](#).

⁴² Even with a theoretically sound measure, there are circumstances where the analysis of discrimination issues is extremely complex. See the Kansas Law paper at 21-22, as well as the [Solutions](#), [Irreducible Minimums](#), and

But one must be able to examine those rates. One can never soundly appraise the strength of such forces on the basis of comparisons of the proportion a group makes up of the pool and the proportion it makes up of persons experiencing the outcome. See also the [IDEA Data Center Disproportionality Guide](#) subpage of the Discipline Disparities page and [slides 97 to 108](#) of the University of Maryland methods [workshop](#).⁴³

The TDHCA brief (at 27-30), Kansas Law paper (at 32-35), and *Mortgage Banking* article (at 93) all address the fundamental problems in analyses of discrimination issues that examine data solely on persons who accepted some outcome or situation. Such analyses are fatally flawed because they do not examine data on the entire universe of persons subject to the processes at issue (a matter I addressed long ago in “[Illusions of Job Segregation](#)” *Public Interest* (Fall 1988), but which remains largely unknown to persons dealing with discrimination issues).⁴⁴ Further, even if the universe analyzed were an appropriate one, there would be no sound measure for appraising the strength of the forces causing outcome rates of advantaged and disadvantaged groups to differ. See TDHCA brief at 28. Nevertheless, all discrimination cases of which I am aware that have involved recoveries in excess of \$100 million, including the two lending cases discussed in the *Amstat New* column that together involved a recovery of more than half a billion dollars, have been based on analyses that examined data solely on persons who accepted some outcome or situation.⁴⁵

As discussed at the outset of this letter with regard to the DOJ/Ferguson [letter](#), the DOJ report on the disparate impact of police and court procedures in Ferguson, Missouri adds the

[Cohort Considerations](#) subpages of the [Measuring Health Disparities](#) page, the [Truncation Issues](#) subpage of [Scanlan’s Rule](#) page, and the [Ferguson Arrest Disparities](#) subpage of the [Discipline Disparities](#) page. See also note 30 *supra*.

⁴³ The tables in these references illustrate some quite complicated aspects to the problematic nature of comparisons between the proportion a group makes up of persons potentially experiencing the outcome (the pool) and the proportion it makes up of persons actually experiencing the outcome. That general reductions in the frequency of an outcome will tend to increase the proportion that more susceptible groups make up of persons experiencing and failing to experience the outcome – hence, increasing both relative and absolute differences between the proportions such groups make up of the pool and the proportions they make up of persons experiencing the outcome and reducing both relative and absolute differences between the proportions the groups make up of the pool and the proportions they make up of persons failing to experience the outcome – is straightforward enough. So, too, is the opposite pattern when the outcome increases. But there are some more complicated issues related to the effects of the proportion a group comprises of the pool on those relative and absolute differences for any given pair of outcome rates for advantaged and disadvantaged groups. See note 35 *supra*.

⁴⁴ That is, to greatly oversimplify, when 50 of 100 white hires or loan recipients receive the favored job or loan product compared with 25 of 100 minority hires or loan recipients, the pertinent favorable outcome rates are not 50% and 25%. The pertinent rates are 50 divided by (a) the number of white job or loan applicants seeking the favored job or loan product and 25 divided by (b) the number of minority job or loan applicants seeking the favored job or loan product. Commonly (a) and (b) are not even examined.

⁴⁵ See articles collected in Section A of the [Employment Discrimination](#) page of [jpscanlan.com](#) and my “[Fair Lending Studies Paint Incomplete Picture](#),” *American Banker* (April 24, 2013).

appraisal of racial differences in criminal justice outcomes to the areas where federal law enforcement is premised on a manifestly incorrect belief about the connection between the frequency of an outcome and relative differences in rates of experiencing it (or the proportion groups most susceptible to the outcome make up of persons experiencing it). Even apart from the DOJ report, there has lately been increasing media or scholarly attention to racial differences in criminal justice outcomes. That attention commonly reflects the mistaken perception that reducing adverse criminal justice outcomes will tend to reduce relative differences in rates of experiencing those outcomes (or the proportion racial minorities make up of persons experiencing those outcomes).

But discussions of racial differences in criminal justice outcomes also often involve the drawing of unsound inference about processes on the basis of increasing relative differences/disproportionality at each increasingly adverse (sometime termed “deeper”) level of the criminal justice system. Thus, although the point is implicit in the above discussion of patterns of relative differences (and disproportionality), it warrants mention that it is to be expected that the deeper the level in the criminal justice system, the greater will tend to be the relative difference in experiencing it and the smaller will tend to be the relative difference in avoiding it. Thus, as in the case of the mistaken inferences discussed in “Illogical Expectations and Unfounded Inferences” section of “Race and Mortality Revisited” (at 339-341), none of the inferences observers commonly draw on the basis of the size of relative differences in rates of experiencing varying levels of adverse outcomes in the criminal justice system has a sound statistical basis (as would also be the case were inferences drawn on the basis of the comparative size of relative differences in avoiding such outcome). See discussion of these issue in my “[Mired in Numbers](#),” *Legal Times* (Oct. 12, 1996), which also addresses such things as that making a three-strikes law a four-strikes law, or reducing the rigidity of application of such rules for imposition of a life sentence, would tend to increase relative racial differences in experiencing that outcome while reducing such differences in avoiding it.⁴⁶ See also the [Offense Type Issues](#) subpage of the Discipline Disparities page regarding mistaken inferences based on the comparative size of relative differences in discipline rates for objectively-defined and subjectively-defined offenses.⁴⁷

⁴⁶ A useful example of the failure of understanding in this area may be found in the extensive report the National Academy of Sciences issued in 2014 titled “[The Growth of Incarceration in the United States: Exploring Causes and Consequences](#).” The report gives substantial attention to racial disparities in crime rates and in rates of experiencing various adverse outcomes in the criminal justice system. It analyzes disparities in terms of relative differences in adverse outcomes, absolute differences in adverse outcomes, and the proportion racial minorities make up of persons experiencing adverse outcome compared to their representation in the population. But at no point does the discussion reveal an appreciation of the ways the measures discussed tend to systematically affected by the frequency of an outcome. As a result, as with virtually all health and healthcare disparities research, the report’s conclusions about criminal justice disparities are invariably questionable, and, even when broadly correct, misleadingly imply that the measures employed effectively quantify the forces causing outcome rates to differ.

⁴⁷ Appraisals the rates at which different races experience various types of adverse outcome are often also problematic because the category analyzed reflects some intermediate situation, as discussed in the second paragraph of note 25 *supra* and the [Intermediate Outcomes](#) subpage the Scanlan’s Rule page.

Moderately extended discussions of measurement issues focused on the interpretation of subgroup effects in clinical trials, and on methods for employing an observed effect in a clinical trial to calculate the number-needed-to-treat in situations involving different baseline rates from that observed in the trial, may be found in the [Subgroup Effects](#) subpage of the [Scanlan's Rule](#) page and my May 16, 2014 [Comments](#) on Federal Drug Administration subgroup effects regulations. Moderately extended discussions of measurement issue focused on the interpretation of data on demographic differences in achieving or failing to achieve certain academic proficiency levels may be found in the above-mentioned April 30, 2015 [letter](#) to the Education Trust and May 13, 2014 [letter](#) to the Annie E. Casey Foundation and on the [Educational Disparities](#) page of [jpscanlan.com](#) and its subpages.

Many other discussions of the above-described patterns, and of implications of failure to understand them in particular contexts, may be found in several dozen articles collected on the [Bibliography](#) subpage of the [Scanlan's Rule](#) page of [jpscanlan.com](#). More important older articles not mentioned, or given only passing mention, above include "[Race and Mortality](#)," *Society* (Jan./Feb. 2000); "[When Statistics Lie](#)," *Legal Times* (Jan. 1 1996); "[Getting it Straight When Statistics Can Lie](#)," *Legal Times* (June 23, 1993); "[The Perils of Provocative Statistics](#)," *Public Interest* (Winter 1991); "[Comment on McLanahan, Sorensen, and Watson's 'Sex Differences in Poverty, 1950-1980'](#)," *Signs* (Winter 1991); "[An Issue of Numbers](#)," *National Law Journal* (Mar. 5, 1990); "[The 'Feminization of Poverty' is Misunderstood](#)," *Plain Dealer* (Nov 11, 1987). Each item involves an essentially universal misunderstanding of an important demographic difference issue, often involving a perception about the meaning of large relative differences in adverse outcomes (or the high proportion a disadvantaged group makes up of persons experiencing such an outcome) that is the opposite of reality. Each of those misunderstandings is as pervasive today as it was when the article was published.

Scores of other examples of particular failures to understand the relationship between the frequency of an outcome and some measure of difference between outcome rates may be found on the 140-plus online comments on the [Journal Comments](#) subpage of the [Measuring Health Disparities](#) page (MHD) and the 30-plus conference presentations⁴⁸ or methods workshop on the [Conference Presentations](#) subpage of MHD.

The pages and subpages of [jpscanlan.com](#) devoted to measurement issues also discuss many examples of the instances in which measures have in fact behaved in the manner described above, and the widespread or universal failure to understand implications of the frequency of the outcome, in particular contexts. Of special pertinence to the federal government's misperceptions about the effects of reducing adverse discipline outcomes on relative racial/ethnic differences in disciplines rates are the subpages of the Discipline Disparities page discussing data from across the country showing that recent general reductions in discipline rates have been

⁴⁸ Presentations at events sponsored by the American Statistical Association or the Royal Statistical Society include those given at Joint Statistical Meetings of 2006, [2007](#), [2008](#), and [2009](#), International Conferences for Health Policy Statistics of [2001](#), [2008](#), and [2011](#), and the [2009](#) Royal Statistical Society Conference.

accompanied by increasing relative racial/ethnic differences in discipline rates. See the following subpages pertaining to patterns in the jurisdictions identified in the names of the subpages: [Los Angeles SWPBS](#), [Denver Disparities](#), [Florida Disparities](#), [Maryland Disparities](#), [California Disparities](#), [Connecticut Disparities](#), [Maryland Disparities](#), [Minnesota Disparities](#), [Rhode Island Disparities](#), [St. Paul Disparities](#), [Minneapolis Disparities](#), [Beaverton \(OR\) Disparities](#), [Portland \(OR\) Disparities](#), [Montgomery County \(MD\) Disparities](#), and [Henrico County \(VA\) Disparities](#).

See also the [DOE Equity Report](#) subpage of the Discipline Disparities page.⁴⁹ That subpage discusses data in the DOE's November 2012 report titled "[Helping to Ensure Equal Access to Education: Report to the President and Secretary](#)" showing that, notwithstanding claims of the Department of Education and Justice that zero tolerance policies are responsible for large relative racial differences in adverse discipline outcomes, relative racial differences in expulsion rates are smaller in districts with zero tolerance policies than in districts without such policies.

Other examples of failures of understanding regarding the appraisal of demographic differences in outcome rates, with a particular focus on activities of the recipient entities, may be found in the letters listed at pages 4-5 *supra*. The letters should be examined with recognition that the criticisms of methods of recipient institutions and organizations would apply to virtually every institution and organization in the world that analyzes demographic or other differences in outcome rates or provides guidance on methods to conduct analyses of such differences. The same may be said of the journals that publish studies involving analyses of such differences.

4. Recommendation

As discussed above and reflected in referenced materials, regardless of the measure employed, very little in the analysis of demographic or other differences in outcome rates has been sound. In fact, virtually never has anyone analyzed group differences on the basis of some standard measure of differences between outcome rates while recognizing the implications of the frequency of an outcome. Nor has anyone attempted to distinguish, in cases where some measure of difference between outcome rates changed over time, or where some measure of difference between outcome rates was otherwise larger in one setting than another, between the extent to which the observed patterns are functions of different frequencies of an outcome and the extent to which the patterns actually reflect something about underlying processes. Thus, even when conclusions about observed patterns are broadly correct in the sense of being to the same general effect of those one would reach while take the described frequency-related patterns into account, or by means of a measure unaffected by the frequency of an outcome, presentations of those conclusions in terms of standard measures have been invariably misleading by

⁴⁹ The principal measurement pages are: [Measuring Health Disparities](#), [Scanlan's Rule](#), [Mortality and Survival](#), [Statistical Reasoning](#), [Immunization Disparities](#), [Educational Disparities](#), [Disparate Impact](#), [Discipline Disparities](#), [Lending Disparities](#), [Employment Discrimination](#), [Feminization of Poverty](#), and [Vignettes](#). The pages have close to a hundred subpages.

suggesting that the measures employed effectively quantify the difference in the circumstances of advantaged and disadvantaged groups reflected by the differing rates of experiencing some outcome.

Thus, addressing the failure of those analyzing group differences to consider the issues discussed here is a matter urgently warranting attention from the nation's principal statistical organization. Formation of a committee to address the matter would seem the reasonable means of giving the matter the necessary attention. As suggested above, adequately addressing this subject may be a long-term undertaking. It may take some time for an ASA committee to reach a consensus on how to measure demographic differences in outcome rates. Such a consensus may not even be possible. Further, the patterns described above may have implications well beyond those I have considered and any such implications should not go unaddressed.

But, given the scope of the resources continually devoted to unsound analyses of health and healthcare disparities issues and many other matters, every effort should be made to address this subject as soon as possible. In addition, there is no reason why ASA cannot, in the very near future, bring to the attention of its members, the public, and pertinent government entities that there exists serious problems in analyses of group differences that fail to consider the ways standard measures of differences between outcome rates tend to be systematically affected by the frequency of an outcome. Thus, researchers who may not yet be aware of these issues may start to consider them immediately, even if it may take some time before ASA can provide formal guidance on how to address them.

B. Reasons That the American Statistical Association Should Immediately Advise Congress and Federal Agencies That Reducing the Frequency of an Outcome Tends to Increase, Not Decrease, Relative Differences in Rates of Experiencing the Outcome

Other measurement issues aside, I suggest that there is no plausible basis for questioning the existence of a pattern whereby the rarer an outcome the greater tends to be the relative difference in experiencing it, even if one may find many situations where the pattern is not observed. But, even if there were some basis for questioning whether reducing the frequency of an outcome will tend to increase relative differences in rates of experiencing it, there still would be no conceivable basis for maintaining that reducing the frequency of an outcome will tend to reduce relative differences between rates of experiencing it.

Yet, as discussed at the end of Section A.1.a, the belief that reducing the frequency of an outcome should reduce relative differences in experiencing the outcome is pervasive among governmental entities enforcing civil rights laws, and widespread even within the social and medical science research communities that commonly, or universally, understand that the more one restricts a favorable outcome toward the tail of a distribution, the greater will tend to be the relative differences in rates of experiencing that outcome.

The entire research community needs to be educated on this matter. But I limit the specific suggestion of this letter to one that ASA formally advise Congress, the Comptroller General, the Office of Science and Technology, and other Executive Branch agencies that reducing the frequency of an outcome tends, to increase, not decrease (a) relative differences in experiencing the outcome and (b) the proportion groups most susceptible to the outcome make up of persons experiencing it. By advising the government on this matter, ASA will do much to educate the research community as well.

The anomalies discussed in the *Amstat News* column alone provide compelling reason for action by any organization that understands the points of the column and that recognizes an obligation to promote sound government action concerning matters relevant to the organizations' mission. But irrespective of those anomalies, the simple facts that the government believes something of a statistical nature that the ASA knows to be false, and that the government also attempts to lead the public and entities covered by civil rights laws to accept that belief, obligates an organization like ASA to correct the government on the matter.

It nevertheless warrants mention that the *Amstat News* column addressed only a few of the situations where the government, either affirmatively or passively, contributes to a failure of understanding that materially undermines important policies. For example, as noted above, the Individuals with Disabilities Education Act has a provision that requires actions that typically will have effects that are the opposite of what that the legislation intends.

State and local educational authorities devote immense resources to examining demographic differences in a range of educational outcomes, including such things as “significant disproportionality” by racial/ethnic group in assignment to special education. Like the federal agencies monitoring the actions of those authorities, the authorities do so while thinking that reducing assignment to special education will tend to reduce relative differences in assignment rates. Indeed, for more than 35 years large relative racial differences in assignment to special education (or differences between the proportion racial minorities make up of school children and the proportion they make up of children assigned to special education) have been debated without an understanding that any efforts to address those differences that generally reduce assignment rates will tend to increase relative racial differences in assignment rates (and the proportion racial minorities make up of persons assigned). Consequently, much of the discussion of such issues has made little sense.

One recent development in this area, which is touched upon in the HHS/DOE letter (at 10-11), involves a Government Accountability Office (GAO) report released in March 2013 titled “[Standards Needed to Improve Identification of Racial and Ethnic Overrepresentation in Special Education](#)” and the probable DOE response thereto. The GAO report urged the DOE to provide guidance to the states on how to measure “significant disproportionality” in assignment of particular racial/ethnic groups to special education programs, something the states usually

measures by some ratio of assignment rates. DOE issued a [Request for Information](#) on the subject in June 2014.⁵⁰

The report makes evident that GAO, like all other federal entities save NCHS, has no understanding of the ways such ratios tend to be affected by the frequency of an outcome or any awareness that reducing the number of special education assignments will tend to increase those ratios. The same can likely be said of every document GAO has issued that touches on racial or other differences in outcome rates. See the above-mentioned September 9, 2014 [letter](#) to the GAO's Financial Markets and Community Investment Program. It is true that the GAO report on "significant disproportionality" cannot be read as specifically leading readers to believe that reducing assignment rates will tend to reduce relative racial differences in assignment rates. But the report's failure to recognize the relationship between the frequency of assignment and relative differences in assignment rates nevertheless contributes to the general lack of understanding of the matter. Whether any response DOE ultimately makes to the GAO report will specifically lead readers erroneously to believe that generally reducing special education assignment rates will tend to reduce relative differences in assignment rates, it is safe to assume that, unless an entity like ASA educates DOE on the matter, such response also will contribute to the misunderstanding of this matter.

Further, the anomalies specifically discussed in the *Amstat New* column have only grown more severe in recent years. The column discussed that Colorado had already enacted legislation to generally reduce discipline rates, and that other jurisdictions were considering similar legislation, based on the belief that reducing the frequency of public school adverse discipline outcomes would reduce relative racial/ethnic differences in discipline rates. But, as reflected in the referenced Discipline Disparities subpages bearing names of states or local authorities, and the February 26, 2015 [letter](#) to the Vermont Senate Committee on Education, many similar actions have been proposed or taken by legislative bodies or state or local school authorities. See also the *Baltimore Sun* [commentary](#) cited in note 2 *supra* regarding the planned action of the Maryland State Board of Education to relax discipline standards in order to reduce relative racial differences in adverse discipline outcomes. The proposed action became final several months later.

Similarly, public interest groups frequently call for reducing adverse school discipline outcomes in order to reduce relative differences in those outcomes or the proportion disadvantaged groups make up of persons experiencing those outcome (of which the subjects of the April 7, 2015 [letter](#) to Texas Appleseed and the August 31, 2015 [letter](#) to McKinney, Texas Independent School District are merely examples).

Some of the recent actions by DOE and DOJ, now joined by HHS, that encourage the reducing of discipline rates in order to reduce relative differences in discipline rates are discussed in "Race and Mortality Revisited" (at 342) and the HHS/DOE letter. In addition to issuance,

⁵⁰ It was this Request for Information that prompted my creation of the [IDEA Data Center Disproportionality Guide](#) subpage of the [Discipline Disparities](#) page and the August 11, 2014 [letter](#) to the IDEA Data Center.

with HHS, of the “Policy Statement on Expulsion and Suspension Policies in Early Childhood Settings” discussed in the introduction, DOE recently launched a “[Rethinking Discipline](#)” campaign. The web page for the campaign, including its Know the Data section, will contribute to the erroneous belief that generally reducing discipline rates tends to reduce relative differences in discipline rates.

In fact, it is difficult to keep up with all the recurring iterations of misinformation on the subject of racial disparities in school discipline by DOE and various public interest groups. Meanwhile, the data discussed on the Discipline Disparities subpages mentioned above consistently show that recent reductions in discipline rates have been accompanied by increased relative differences in discipline rates. And, with rare exception if not universally, state and local authorities are confounded by these results because the government has for so long led them to expect the opposite.

At the same time, the DOE is investigating large relative differences in discipline rates in various school districts and entering into agreements that will call for both general reductions in discipline rates and reductions in relative differences in discipline rates. See the [Oakland Agreement](#) subpage of the Discipline Disparities page. The record to date suggests that, in appraising compliance with such agreements, the DOE, the local authorities, and experts retained to monitor the authorities’ compliance will all proceed on the mistaken belief that reducing discipline rates will tend to reduce relative differences in discipline rates and/or the proportion racial minorities make up of persons disciplined.

Significant developments regarding enforcement of fair lending laws are discussed in the letter to Chief Data Scientist Patil. These include (a) the February 2013 issuance by the Department of Housing and Urban Development (HUD) of its final rule titled “[Implementation of the Fair Housing Act’s Discriminatory Effects Standard](#),” which specifies that a practice with a disparate impact on a protected group can be upheld only if there exists no less-discriminatory alternative that equally serves the covered entity’s legitimate business interest, and (b) the Supreme Court’s upholding that rule in its June 2015 [decision](#) in *Texas Department of Housing and Community Affairs et al. v. The Inclusive Communities Project, Inc.* (and upholding the general applicability of the disparate impact doctrine to the Fair Housing Act). Such developments will increase pressure on lenders to seek ways of reducing disparate impacts, which, in consequence of the government’s mistaken guidance on the matter, commonly involve actions that generally reduce adverse outcomes and, in doing so, tend to increase disparate impacts according to the way the government measures such things.⁵¹

As discussed with regard to the [letter](#) to the Department of Justice and the City of Ferguson, Missouri, the same misunderstanding that has undermined the government’s

⁵¹ In July 2015 HUD issued its final rule titled “[Affirmatively Furthering Fair Housing](#),” which specifically applies the disparate impact doctrine to public housing location issues. Because such issues tend to be analyzed in terms different from comparisons of favorable or adverse outcome rates of advantaged and disadvantaged groups, I am at this time uncertain of the specific pertinence of the subject of this letter to such issues. But see the [LIHTC Approval Disparities](#) subpage of the [Scanlan’s Rule](#) page of [jpscanlan.com](#).

David R. Morganstein, President, *et al.*
American Statistical Association
October 8, 2015
Page 40

enforcement of laws prohibiting discrimination regarding lending and educational matters has been imported into its efforts to ensure nondiscrimination in local law enforcement. Here, too, in addition to incongruous enforcement actions resulting from the government's mistaken belief that reducing adverse outcomes in the criminal justice system will tend to reduce relative differences in rates of experiencing those outcomes, countless authorities may be taking actions based on the assumption that the government's belief must be correct, with the same anomalous consequences as in the school discipline context. Here, too, public interest groups are contributing to these pressures with their own flawed analyses of disparities issues. See the above-mentioned June 8, 2015 [letter](#) to the City of Minneapolis regarding an American Civil Liberties Union study of Minneapolis police practices that reflects the same misunderstandings found in the DOJ report on the racial impact of police and court practices of Ferguson, Missouri.

In sum, there are vast areas where the failure of the government and others to understand that reducing the frequency of an outcome tends to increase relative differences in rates of experiencing the outcome undermines or has the potential to undermine civil rights enforcement and other important government policies and appraisals of policies. But even if there were no consequences of the government's failure to understand this matter, it would be incumbent upon an organization like ASA to correct that failure.

Ultimately, I suggest, ASA will see reason to make clear to a wide range of entities that relative differences in rate of experiencing an outcome, like other measures that tend to change solely because the frequency of an outcome changes, do not provide a useful basis for evaluating processes affecting outcome rates. But, however, long it may take ASA to come to that view, ASA should immediately advise the government that the belief that reducing the frequency of an outcome will tend to reduce relative differences in rates of experiencing the outcome is incorrect.

Please contact me if you have questions about any matter discussed in this letter.

Sincerely

/s/ James P. Scanlan

James P. Scanlan