MEMORANDUM   *Corrected September 21, 2020 (see p. 46), modified September 24, 2020, by the addition of Table 4 and related discussion toward the end of Section B in accordance with points in September 22, 2020 email to participants.   Further corrected on October 16, 2020.*

TO: Participant in the Department of Housing and Urban Development's Panel on Housing Discrimination Against LEP Individuals

FROM: James P. Scanlan, Panelist

DATE: September 19, 2020

SUBJECT: Measurement of Discrimination Issues

INTRODUCTION

The purpose of this memorandum is to bring to the attention of all persons involved with the Department of Housing and Urban Development's (HUD's) September 22, 2020 expert panel on housing discrimination issues certain statistical concepts that it is crucial for persons analyzing demographic differences to know but that few people analyzing demographic differences in fact know.

While I hope that the comprehensiveness of the memorandum will make it a useful resource going forward, I recognize that participants may not be able to read or fully understand all of it prior to the panel.[1] For that reason, I set out immediately below certain key points that are addressed in some manner in the memorandum and that will be reflected in my own views during the panel discussion with respect to any issues to which the matters are pertinent.

1. Contrary to the belief of almost every federal government agency and the overwhelming majority of social scientists, generally reducing an adverse outcome tends to increase, not reduce, relative (percentage) racial and other demographic differences in rates of experiencing the outcome (though it tends to reduce relative differences in rates of experiencing the corresponding favorable outcome). This holds whether the adverse outcome is rejection of a loan application, foreclosure on a mortgage, eviction from a rental property, denial of an opportunity to view a property, suspension from school,

---

[1] Because of the haste in preparation of the memorandum in order to provide it to participants at least a couple of days in advance of the panel discussion, it is certain to have a substantial number of clerical errors and may have some other errors as well. When corrections are made they will noted on a copy made available by a link on the Measurement Letters page of jpscanlan.com, which contains links to many similar items.

arrest or incarceration, infant or maternal mortality, mortality from COVID-19, failure to receive a test or appropriate therapy for COVID-19, or any other adverse outcome.

2. Promotion of the mistaken belief that reducing adverse outcomes will tend to reduce, rather than increase, relative racial differences in rates of experiencing the outcomes materially contributes to beliefs that racial bias and racism are increasing.

3. Failure to understand patterns by which measures tend to be affected by the prevalence of an outcome has undermined virtually all studies of racial and other demographic differences regarding outcome rates in the law and the social and medical sciences, as well as the interpretation of data in clinical trials. None of the measures commonly used to quantify differences between outcome rates of advantaged and disadvantaged groups provides a sound means of determining whether discrimination or any other factor causing outcome rates of advantaged and disadvantaged groups to differ has increased or decreased over time or is larger in one setting than another because each measure tends to be affected by the overall prevalence of an outcome.

4. Only a measure that does not change solely because there occurs a change in the overall prevalence of an outcome akin to that effected by altering a test cutoff can effectively quantify demographic differences involving outcome rates.

5. The government should neither conduct nor fund research into demographic differences unless such research attempts to distinguish between the extent to which observed patterns are solely function of the varying prevalence of an outcome and the extent to which the patterns reflect the effects of policies or other factors on the differences in the circumstances of the groups being studied.

6. Paired tester studies suffer from the same measurement problems as other studies but with the additional problem that decisions about the tester plan systematically affect the size of the measures used to quantify treatment differences. With a sound measure, it is possible to effectively quantify differences in treatment of two groups on the basis of the rates at which testers from each group receive or fail to receive a desired outcome. But it may not be possible to effectively quantify differences in treatment of the two groups on the basis of rates at which members of tester pairs appeared to be favored of disfavored.

Some panelists already are or may be familiar with certain of the issues addressed here. As discussed further towards the end of this Introduction, Panel Chair Gail Heriot and panelist Peter Kirsanow, who are both members of the U.S. Commission on Civil Rights, are familiar with some of the issues, among other reasons, because of my December 8, 2017 testimony at the Commission's December 8, 2017 hearing on school discipline disparities issues. Mr. Kirsanow quoted from that testimony in his dissent to the Commission's July 2019 report on the hearing and related investigation. As also explained toward the end of the Introduction, Panelist Robert Santos, may be familiar with the issues because of his longstanding involvement with the American Statistical Association (ASA), an organization of which is he currently President-Elect. I have many times addressed these issues in ASA forums and have addressed them at great length in communications to ASA leadership, including Mr. Santos when he was Vice-President

of the organization. Mr. Santos may also be familiar with the issues through his role as Vice President and Chief Methodologist of the Urban Institute and more specifically because of a 2014 Urban Institute report addressing some of the issues.

The issues addressed below involve patterns by which all standard measures of differences between outcome rates tend to be systematically affected by the prevalence of an outcome. That is, measures tend to change solely because the prevalence of an outcome changes and different measures tend to change in opposite directions as the prevalence of an outcome changes. The measures thus cannot provide useful information regarding (a) how policies affect differences in the circumstances of advantaged and disadvantaged groups, (b) the ways different decisionmakers treat persons of different demographic groups, and (c) whether discrimination or other forces causing outcome rates of advantaged and disadvantaged groups to differ are growing stronger or weaker over time or are greater in one setting than another setting, unless those studying the subjects have a complete understanding of the patterns by which the measures tend to change solely because the prevalence of an outcome changes (or employ a measure that is unaffected by the prevalence of an outcome). That is the matter I refer to below as the broader problem in analyses of demographic differences with respect to discrimination issues as well as all other issues involving differing favorable or adverse outcome rates of advantaged and disadvantaged groups.

But an important element of the failure to understand patterns by which measures tend to be affected by the prevalence of an outcome – and one that is responsible for some sorely misguided civil rights law enforcement policies and profound misunderstandings about whether discrimination or racism in the United States is growing stronger or weaker – is the mistaken belief, shared by the overwhelming majority of persons analyzing demographic differences and all government agencies involved in the enforcement of civil rights laws, that generally reducing an adverse outcome like rejection of a loan application, suspension of students from school, arrest or incarceration or the use of unnecessary force by law enforcement officers, or death from COVID-19 will tend to reduce, rather than increase, (a) relative (percentage) racial differences between the rates at which advantaged and disadvantaged groups experience the outcome and (b) the proportion the disadvantaged group makes up of persons experiencing the outcome.

This mistaken belief may be regarded as one part of the failure to understand the pattern whereby reducing the frequency of an outcome tends to increase relative differences in rates of experiencing the outcome while reducing relative differences in rates of experiencing the corresponding opposite outcome. By way of the simplest of examples, where two groups differ in average test scores, lowering a test cutoff will tend to increase relative differences between the failure rates of higher-and lower-scoring groups while reducing relative differences between the groups' pass rates. Or, as Urban Institute researchers put the matter with respect to an insurance coverage issue in the 2014 report that I will discuss at various places in this memorandum, "[w]hen two groups differ in their uninsurance rate, if the overall uninsurance rate (i.e., the rate of uninsurance across both groups) decreases without any change in the underlying difference between the groups, then (1) the relative difference in uninsurance rates between the groups tends to be greater, and (2) the relative difference in rates of having health coverage tends to be smaller."

3

While the pattern was recognized by National Center for Health Statistics (NCHS) as early as 2004 and is discussed in the agency's most significant disparities measurement guide, few people analyzing demographic differences yet understand that it is even possible for the relative difference in a favorable outcome and the relative difference in the corresponding adverse outcome to change in opposite directions, much less that this tends to occur systematically. And the great majority of such persons, like all government agencies enforcing civil rights laws, believe that reducing an adverse outcome will tend to reduce relative differences in rates of experiencing the outcome. That belief is exactly the opposite of what NCHS found and, more important, exactly the opposite of reality.

Discussions of varying comprehensive and complexity of both the broader and narrower issues may be found in my "Race and Mortality Revisited," *Society* (July/Aug. 2014)[2] (to which I will make repeated references below), "The Mismeasure of Health Disparities," *Journal of Public Health Management and Practice* (July/Aug. 2016), my amicus curiae brief in *Texas Department of Housing and Community Development, et al. v. The Inclusive Communities Project, Inc*., Supreme Court No. 13-1731 (Nov. 17, 2014) (TDHCD Brief), "The Perverse Enforcement of Fair Lending Laws," *Mortgage Banking* (May 2014), "Measuring Health and Healthcare Disparities," Federal Committee on Statistical Methodology 2013 Research Conference (FCSM Paper), "The Mismeasure of Discrimination," Faculty Workshop, University of Kansas School of Law (Sept. 20, 2013), "Can We Actually Measure Health Disparities?," *Chance* (Spring 2006), "Race and Mortality," *Society* (Jan./Feb. 2000) "Divining Difference," *Chance* (Fall 1994), "The Perils of Provocative Statistics," *Public Interest* (Winter 1991), as well as "Measuring Discipline Disparities," Testimony for U.S. Commission on Civil Rights Briefing "The School to Prison Pipeline: The Intersection of Students of Color with Disabilities" (Dec. 8, 2017). Very extended discussions of the issues may be found in my Comments for the Commission on Evidence-Based Policymaking (Nov. 14, 2016) (CEP Comments), Letter to the American Statistical Association (Oct. 8, 2015), and Letter to Harvard University (Oct. 9, 2012).

Most of the above materials include tabular or graphical illustrations of pertinent statistical patterns based on actual and hypothetical data. Many more such graphical and tabular illustrations may be found in seven methods workshops given at American universities between 2012 and 2015,[3] as well twenty-two conference presentations given in Europe and North

---

[2] I have not attempted any consistency of formatting for the various references. In some cases, items references are identified solely by links.

[3] "The Mismeasure of Health Disparities in Massachusetts and Less Affluent Places," Quantitative Methods Seminar, Department of Quantitative Health Sciences, University of Massachusetts Medical School (Nov. 18, 2015) (Abstract), "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).

America between 2001 and 2011 (links to which may be found here). Other treatments of these issues as they bear on the methodological failings of particular research institutions and governmental entities may be found in letters to such entities between 2010 and 2020, links to which are collected on my Measurement Letters webpage. The page also contains links to letters to school districts and law enforcement authorities whose activities have been scrutinized by public interest groups or law enforcement agencies that mistakenly believe that reducing adverse criminal justice and school discipline outcomes tends to reduce relative racial differences in rates of experiencing the outcome.

Readers may get an overview of the issues fairly quickly by browsing the slides in the methods workshop given for arms the University of Minnesota in September 2014 (two of which will be discussed further in Section A with respect to perceptions about demographic differences in certain housing/neighborhood-related issues in the Minneapolis, Minnesota area and other disparities issues in that area) or by browsing materials I used in connection with a July 24, 2017 Federalist Society teleforum titled "Is Federal Civil Rights Enforcement Based on an Understanding of Statistics That is the Opposite of Reality?" (July 24, 2017) or the handout I distributed for a March 22, 2018 meeting with staff of the U.S. Department of Education.[4]

It would be also useful for readers to browse the eight tables in "Race and Mortality Revisited" (several of which I will reference below, including Table 7, which is of particular pertinence to analysis of tester studies and which is discussed in Section B of this memorandum) and the twelve tables in the 2013 FCSM Paper (which is among the sources on which the 2014 Urban Institute report relied for its recognition of several key issues).

The broader issues are also summarized very briefly (508 words apart from references) in a response to a 2012 *Health Affairs* article by Urban Institute researchers, two of whom would co-author the 2014 Urban Institute report discussed here.[5] This response, and communications related to the response, were the principal bases for the researchers to recognize the issues.

---

[4] The Federalist Society teleforum materials dealt solely with misunderstandings of the effects of policies on the aforementioned (a) and (b). The Department of Education materials also discuss the broader issues, as I had done with regard to a range of that agency's policies in "Innumeracy at the Department of Education and the Congressional Committees Overseeing It." Federalist Society Blog (Aug. 24, 2017).

[5] The response, titled "How Measures Are Affected by the Prevalence of an Outcome," responded to Clemans-Cope L, Kenney GM, Buettgens M, et al. The Affordable Care Act's coverage expansions will reduce differences in insurance rates by race and ethnicity. Health Aff (Millwood) 2012;31(5):920-930). As explained in the response made available by the link in the text above, as a result of revisions to the *Health Affairs* website, the response no longer appears on that website and is no longer available by the link provided in note 15 (at 45) of the 2014 Urban Institute report. The title is an editor-shortened version of the submitted title, "Appraisals of Effects of the Affordable Care Act on Healthcare Disparities Must Consider the Ways Measures Tend to be Affected by the Prevalence of an Outcome." The submitted title, and many like titles among about 150 similar online responses to medical or health policy journals collected on my Journal Comments webpage, are more suggestive of the near universal failure of health and healthcare disparities research to reflect an awareness that measures other than the ones employed in a study can, and typically do, yield conclusions about things like directions of changes in the size of disparities that are the opposite of the conclusions being reported.

One key point of the above references is summarized fairly briefly (about 1200 words) in my December 2012 Statistician's View column titled "Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies" in *Amstat News*, the membership magazine of the American Statistical Association. The column explains that, contrary to the beliefs underlying federal civil rights enforcement laws pertaining to lending and school discipline, relaxing standards and thereby reducing the frequency of adverse borrower outcomes and adverse school discipline outcomes tends to increase, not reduce, relative differences between the rates at which advantaged and disadvantaged groups experience the outcomes (while reducing relative differences between rates at which the groups experience the opposite outcomes). The column illustrates the point with test score data showing how, as mentioned above, lowering a cutoff, and thereby generally increasing pass rates and generally reducing failure rates, while tending to reduce relative differences between pass rates of higher- and lower-scoring groups, tends to increase relative differences between the groups' failure rates. The column also explains that credit score data would similarly show that lowering a credit score requirement to receive a loan, while tending to reduce relative differences in rates of meeting the requirement, will tend to increase relative differences in rates of failure to meet the requirement.

The test score example employed in the column is illustrated in Table 1 of Section A to this memorandum, which is the same as Table 1 of "Race and Mortality Revisited," and versions of which appear in many of my articles and workshops.[6] Versions also appear as Tables 1 and 2 of my written statement for the December 2017 Commission on Civil Rights hearing. The tables are also reproduced in the Commission's July 2019 report *Beyond Suspensions: Examining School Discipline Policies and Connections to the School-to-Prison Pipeline for Students of Color with Disabilities* (at 215) as part of the dissenting statement of Commissioner Kirsanow. Section A of the memorandum will use Table 1 to illustrate a variety of things concerning patterns by which measures tend to be affected by the prevalence of an outcome and implications of the near universal failure to understand such patterns among persons analyzing demographic differences.

Tabular illustrations of the *Amstat News* column's point about the effects of lowering a credit score requirement on relative racial differences in rates of meeting and failing to meet the requirement, and a like point that can be made about the effects of lowering an income requirement, may be found in Tables 1 and 2 of my December 13, 2018 letter to HUD's General Counsel and certain other agency officials, as well as in tables in earlier letters to leadership of the Government Accountability Office (Apr. 17, 2018) and Department of Justice (Apr. 13, 2017). Descriptions of the effects of particular changes may be found in my "What the

---

[6] A graphical illustration of the example at a level of sophistications beyond my capabilities may be found in slides 18 and 19 of Mark Kiemele, "Data Science, Design of Experiments, and Predictive Analysis," Tutorial at the 2018 Systems Engineering Test and Evaluation Conference 2018, Sydney, Australia (Apr. 30, 2018). While I do not always agree completely with treatments by others of the patterns I describe, such treatments often illustrate the patterns in ways that are more familiar or more useful to persons with statistical training than my methods. See Lambert PJ, Subramanian S (Disparities in Socio-Economic outcomes: Some positive propositions and their normative implications. Soc Choice Welf 2014;43:565-576), 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)); Thomas H, Hettmansperger TP (Risk Ratios and Scanlan's HRX. J Stat Distr and Appl 2017;4:27).

government gets wrong about fair lending," *American Banker* (Apr. 9, 2018), and my Comments on Consumer Financial Protection Bureau Request for Information: Equal Credit Opportunity Act and Regulation B (Aug. 7, 2020).

The *Amstat News* column also briefly discusses problems in research into demographic differences in the law and the social and medical sciences, including the failure to recognize that improvements in health and health care, while tending to reduce relative racial and other differences in the increasing, favorable outcomes (e.g., survival, receipt of appropriate care), will tend to increase relative differences in the corresponding, decreasing adverse outcomes (e.g., mortality, nonreceipt of appropriate care).[7] The column only touches upon the more complicated patterns by which absolute (percentage point) differences between rates and differences measured by odds ratios tend to be affected by the prevalence of an outcome and does not discuss how one might quantify demographic differences by a measure that is unaffected by the prevalence of an outcome – subjects I had discussed in many other places since 2005 and would discuss at length in "Race and Mortality Revisited" and the various other post-2012 works listed at the beginning of this memorandum.

As discussed earlier and as will be made clear in Section A below, the mistaken belief that generally reducing an adverse outcomes tends to reduce racial and other demographic differences in rates of experiencing the outcome is only a part of the larger problem in the analysis of demographic differences arising from the failure to understand the ways all standard measures of differences between outcome rates tend be affected by the prevalence of an outcome. But the mistaken belief – in addition to being highly pertinent to many HUD activities and many issues HUD studies – is of great consequence with regard to discussions or race in America (the subject of Panel Agenda item 5). For those reasons, I give the matter substantial attention in this Introduction.

In my recent "COPAA v. DeVos and the Government's Continuing Numeracy Problem," Federalist Society Blog (Sept. 12, 2019), and "Usual, But Wholly Misunderstood, Effects of Policies on Measures of Racial Disparity Now Being Seen in Ferguson and the UK and Soon to Be Seen in Baltimore," Federalist Society Blog (Dec. 4, 2019),[8] I discuss the following perverse and potentially inflammatory situation. By promoting the mistaken belief that generally reducing an adverse outcome will tend to reduce relative racial differences in rates of experiencing the outcome (or measures that are functions of relative differences), governments

---

[7] The column does not discuss the fact that, as discussed would be discussed in n "Race and Mortality Revisited" (at 331-335) and "The Mismeasure of Health Disparities" (passim), statisticians at the National Center for Health Statistics (NCHS) recognized this pattern as early as 2004 (though, as discussed in those references, did not act on the recognition in a sensible manner). The matter is discussed briefly supra and will be discussed further in Section A.

[8] The former addresses, inter alia, the Departments of Education and Justice's continuing mistaken belief that reducing adverse outcomes for students in public schools will tend to reduce, rather than increase, relative racial differences in rates of experiencing the outcomes. The latter addresses mistaken beliefs of various governmental entities about effects of reducing adverse criminal justice outcomes on measures of racial/ethnic differences in rates of experiencing the outcomes with a focus on the way that recent reductions in those outcomes have in fact been accompanied by increases in those measures.

affirmatively promote racial mistrust. For, when measures of racial disparity increase in the face of actions that the government has led the public to believe should reduce the measures, observers who believe that racial bias plays an important role in racial differences will reasonably believe that racial bias must be increasing. See also "The Pernicious Misunderstanding of Effects or Policies on Racial Differences in Criminal Justice Outcomes," Federalist Society Blog (Oct. 12, 2017).

The inflammatory possibilities inherent in this situation would seem especially great when the focus is on racial disparities in adverse school discipline outcomes and adverse criminal justice outcomes (including being subjected to the use of force), matters that have lately received great attention in provocative discussions about racial disparities and racism in America.[9] That is particularly so with regard to places like Ferguson, Missouri, and Baltimore, Maryland, that have been the scenes of great civil unrest, and where the Department of Justice enforces consent decrees premised on mistaken beliefs about the effects of reducing adverse criminal justice outcomes on the measures of racial disparity that underlay the perceived need for the decrees and that are employed in monitoring the decrees. This is a subject of the aforementioned April 13, 2017 letter to the Department of Justice (and many email communications to Department of Justice attorneys directly involved with the Ferguson and Baltimore decrees) urging the agency to explain this matter to the courts handling the Baltimore and Ferguson decrees. So far, however, there is no indication than any person within the Department of Justice is aware that the agency's understanding of the effects of policies on measures of racial disparity is incorrect. See also my "Compliance Nightmare Looms for Baltimore Police Department," Federalist Society Blog (Feb. 8, 2017), "The misunderstood effects of the Baltimore police consent decree," *The Daily Record* (Feb. 15, 2018), as well as my Letter to the Honorable James K. Bredar (Feb. 14, 2017) (the judge handling the Baltimore Policy consent decree). See also letters to faculties of the University of Maryland Department of Criminology and Criminal Justice (June 13, 2018) and Johns Hopkins University Departments of Sociology and Applied Mathematics & Statistics (June 11, 2018), urging them to understand the issue and then to explain it to Judge Bredar.

Further, so far, in the face of continuing calls to defund the police and many actual reduction in police resources in response to such calls, no one discussing or dealing directly with such issues has shown an awareness that reducing police resources will tend to increase the measures of racial disparity most commonly cited as evidence of police bias or racism.[10]

Possibilities for mistaken understandings about effects of policies on measures of racial disparity to inflame discussions of racial bias and racism may also be especially great when the focus is on

---

[9] As discussed in the Appendix to the September 2019 Federalist Society Blog post on the *COPAA v. DeVos* case, the U.S. Commission on Civil Rights July 2019 *Beyond Suspensions* report (discussed further below with regard to panel chair Gail Heriot and panelist Peter Kirsanow who are members of the Commission and who both dissented from the report) increased the potentially inflammatory consequences of leading the public to believe incorrectly that reducing public school suspensions would tend to reduce relative racial differences in suspension rates by suggesting that racial differences in suspensions were entirely the result of racial bias.

[10] The particular situation in Minneapolis, Minnesota, another scene of recent great civil unrest, and possibly the first place where calls to defund police were taken seriously by local government officials, is treated in Section A.2

racial disparities in COVID-19 outcomes, as has been very much the case in recent months and may continue to be the case indefinitely. For, just as observers for decades have mistakenly believed that reducing adverse health outcomes ought to reduce relative differences in rates of experiencing the outcomes, they are likely to believe – and government officials and social scientists may well lead them to believe – that improvements in care for COVID-19 will be expected to reduce relative racial differences in rates at which person infected with COVID-19 die from the disease. In fact, as available data already make clear – and that numerate observers should in any case understand – improvements in COVID-19 care will tend to increase relative racial and other demographic differences in rates at which infected persons die from the disease (while reducing relative differences in rates at which infected persons survive the disease), as I recently discussed and illustrated in BMJ comments of April 28, 2020, May 13, 2020 (which specifically addresses on the proposed COVID-19 Racial and Ethnic Disparities Task Force Act), and June 29, 2020, as well as on my Obesity Illustration webpage.

The Public Health England study discussed in in the June 29, 2020 BMJ response includes data showing that among seriously ill COVID-19 patients, relative racial/ethnic and gender differences in mortality rates are greater among younger patients than older patients, while also showing that relative racial/ethnic and gender differences in survival are greater among older patients than younger patients. It thus should be evident to those giving the matter a little thought that improving care such as to give older patients the same chances of survival as younger patients will tend to increase relative racial/ethnic and gender different in mortality rates, while reducing relative racial/ethnic and gender differences in survival rates, both among the older patients (where the overwhelming majority of deaths are concentrated) and among all patients combined. The Obesity Illustration page shows how improving weight control will tend to increase relative racial differences in obesity, while reducing relative racial differences in rates of avoiding obesity. It correspondingly shows that improving COVID-19 outcomes for infected persons who are obese will tend to increase relative racial differences between rates at which obese COVID-19 patients dies from disease, while reducing relative racial differences between rates at which such persons survive the disease. Data on systolic blood pressure of blacks and whites would illustration the same patterns.

Notably, on September 15, 2020, several news sources reported in a provocative manner the release of a Centers for Disease Control and Prevention (CDC) study that had reported that "although Hispanic, Black, and AI/AN persons represent 41% of the U.S. population aged <21 years, these groups accounted for approximately 75% of deaths in persons aged <21 years." From those data, one can divine that the ratio of the rate at which persons in those groups died from COVID-19 is 4.32 times the rate of persons not in those groups. While data were not presented on other age groups, one can be confident that the 4.32 ratio among persons under 21 is much higher than among persons 21 and older, and vastly higher than it would be among persons over 70. But possibly not a single person analyzing such data at the CDC will recognize that such pattern is to be expected because mortality from COVID-19 is extremely rare among young persons. As discussed above, such pattern, along with contrasting pattern for failing to die from the disease, is one of the reasons informed observers should understand that improving care will tend to increase relative racial/ethnic differences in mortality while reducing relative racial/ethnic differences in survival. Few observers, however, do understand these things. Thus, observers may well find evidence of racism or genetic difference in the comparatively large

relative racial/ethnic differences in COVID-19 death among young persons just as they have found such evidence in comparatively large relative racial differences in infant mortality where parents are well educated or comparatively large mortgage rejection rates among higher-income loan applicants – as discussed, for example, in "Race and Mortality," Society (Jan./Feb. 2000) – without any understanding that relative differences in adverse outcomes tend to be comparatively large, while relative differences in the corresponding favorable outcomes tend to be comparatively small, wherever the adverse outcomes are comparatively uncommon.

This subject, along with the subject of the way perceptions about the effects of increases in rates of screening, vaccination, or receipt of any other useful intervention depend on the measure an observer happens to favor is discussed further in Section A. That discussion includes an explanation of why the NCHS and the Agency for Healthcare Research and Quality (AHRQ) would tend to reach opposite conclusions regarding whether disparities are larger before or after such increase and that AHRQ, remarkably, will tend to find that some disparities decreased over a period even though it also finds that the disparities are larger at the end of a period than at the beginning the period.

With regard to matters that are within the particular province of HUD, as in the *Amstat News* column, I have usually emphasized the way the misunderstanding of this issue has led to the anomaly whereby lenders that comply with government encouragements to relax lending standards in order to reduce relative racial differences in loan rejection rates increase the chances that the government will accuse them of discrimination. This anomaly, along with the government's obligation to explain to affected entities that the government's explicit or implied guidance regarding the effects of policies on measures of racial disparity was incorrect, was also the focus of the December 13, 2018 letter to HUD officials, as well as my September 11, 2019 Comments on HUD's Proposed Rule Implementation of the Fair Housing Act's Disparate Impact Standard.

I am uncertain whether misleading the public with respect to the effects of policies on relative differences in adverse borrower outcomes has the same potential for inflaming discussions about racism as misleading the public with respect to the effects of policies on relative differences in adverse school discipline, criminal justice, or health outcomes, or whether the matter would be any different with regard to disparities in the foreclosure/eviction rates that are receiving such attention in light of the adverse economic effects of COVID-19-related shutdowns. In any case, it essential that HUD and all persons discussing the last subject recognize that the increases in foreclosures and evictions that will result from the shutdowns will tend to reduce relative demographic differences in foreclosure and eviction rates while increasing relative demographic differences in rates at which homeowner and renters avoid these outcomes – just as raising (rather than lowering) lending requirements would tend to reduce relative differences in loan rejection rates while increasing relative differences in loan approval rates.

It may be even more important that HUD and others understand that policies that reduce foreclosures and evictions – whether mandated by governments or voluntarily implemented by lenders and landlords themselves – will tend to increase relative demographic differences in foreclosure and eviction rates while reducing relative differences in rates of avoiding these outcomes. And, with respect to the civil rights law enforcement issues that were the particular

subject of the *Amstat News* column, it will be crucial that entities involved with such enforcement understand that, other things being equal, the more lenient are particular lenders and landlords in response to missed mortgage and rental payments, the greater will tend to be relative racial differences in rates of foreclosures and evictions by those lenders and landlords, while the smaller will tend to be relative racial differences in rates of avoiding foreclosure and evictions by such lenders and landlords. See my "What the government gets wrong about fair lending," *American Banker* (Apr. 9, 2018), and my Comments on Consumer Financial Protection Bureau Request for Information: Equal Credit Opportunity Act and Regulation B (Aug. 7, 2020). See also my "Bias Data Can Make the Good Look Bad," *American Banker* (Apr. 27, 1992).

It will also be important for HUD and others to understand the following with regard to the frequently-noted circumstances where, for lack of knowledge of the programs or for other reasons, disadvantaged groups benefit less often than advantaged groups from particular programs aimed at reducing foreclosures or evictions. Increasing the availability and awareness of such programs, while tending to reduce relative differences in rates of being aware of and benefitting from such programs, will tend to increase relative difference in rates of being unaware of and not benefitting from such programs.

<div align="center">***</div>

With regard to panelist familiarity with the measurement issues I raise, as noted, Professor Heriot and Mr. Kirsanow are familiar with certain issues at least as a result of my testimony in the December 2017 Commission on Civil Rights hearing "The School to Prison Pipeline: The Intersection of Students of Color with Disabilities." In written and oral testimony, I sought to explain to the Commission that, contrary to the views promoted by the Department of Education and Justice in a January 2014 Dear Colleague Letter and other places, generally reducing adverse school discipline outcomes will tend to increase relative racial differences in rates of experiencing the outcomes (a subject treated in the *Amstat News* column and articles like "The Paradox of Lowering Standards," *Baltimore Sun* (Aug. 5, 2013) even before the Dear Colleague letter was issued). In oral testimony (Tr. 52-53), I discussed the obligations of the Departments of Education and Justice to explain to the public and others the way the agencies have misled them regarding the effect of policies on measures of racial disparity. I also suggested that, in light of the possible inability of those agencies to understand the issues, the Commission on Civil Rights should assume that responsibility. I also suggested that the Commission assume the responsibility of informing the public about the broader measurement issues such as are addressed in my CEP Comments.   Mr. Kirsanow excerpted significant parts of my written statement his dissent from the Commission's July 2019 *Beyond Suspensions* report (at 214-216), while criticizing the Commission's failure to adequately address my testimony in the report.[11]

---

[11] As discussed in the Appendix to the Federalist Society Blog post on COPAA v. Devos, the *Beyond Suspensions* report continued to promote the mistaken view that generally reducing suspensions would tend to reduce relative racial and other differences in suspension rates. While probably not evident to most readers, the report's treatment of my testimony (at 145-46) may be regarded as rejecting my testimony that reducing suspensions will tend to increase relative differences in suspension rates solely on the basis that it was possible for absolute differences between suspensions rates to decrease at the same time that relative differences increased.   That it is not merely possible for general reductions in suspensions to reduce absolute differences in suspensions at the same time that the reductions increase relative differences in suspensions, but that this will typically occur, is something that I have often pointed out or illustrated, as for example, in the "The Paradox of Lowering Standards," *Baltimore Sun* (Aug. 5,

Professor Heriot, who also dissented from the *Beyond Suspensions* report, was already somewhat familiar with the issues at least as a result of her participation in a panel at the Federalist Society Fourth Annual Executive Branch Review Conference (May 17, 2016), in which I gave a presentation titled "The Mismeasure of Disparate Impact," and in conjunction with which I published "Is the Disparate Impact Doctrine Unconstitutionally Vague?," Federalist Society Blog (May 6, 2016). Like the December 13, 2018 letter to HUD officials, those materials addressed constitutional issues arising from the fact that actions that are commonly regarded as less discriminatory alternatives to a practice with a disparate impact, such as lowering a lending requirement, tend to increase relative differences in loan rejection rates at the same time that the actions reduce relative differences in loan approval rates. Both the presentation and the blog post also addressed broader issues regarding the measurement of discrimination without consideration of the ways the measures used tend to be affected by the prevalence of an outcome.

Mr. Santos, who, as noted, has long been involved in American Statistical Association (ASA) activities and who is currently the organization's President-Elect, may be familiar with the issues simply because of my discussion of the issues in varied ASA forums since 1994.[12] But he also may be familiar with the issues because, as a Vice-President of the organization, he was recipient of my letters to the agency's leadership dated October 8, 2015, December 2, 2015, and July 25, 2016. The October 8, 2015 letter, which was mentioned above as an extensive treatment of the broader issues addressed here, urged the organization (a) 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 long-term undertaking), and (b) to advise arms of the United States government that generally reducing the frequency of an adverse outcome will tend to increase, not reduce, relative differences between rates at which advantaged and disadvantaged groups experience the outcome (something that the organization should do immediately). The December 2, 2015 letter discussed the statistical issues further in connection with a specific request for recusal of the then President of ASA from consideration of issues raised in the October 8, 2015 letter.[13] The July

---

2013). I have also explained why such pattern will be commonly occur whenever adverse outcome rates are in the ranges that are usually found for suspension rates of all racial/ethnic groups in "Race and Mortality Revisited" and many scores of other places since 2006.

[12] I had previously discussed the key point of the *Amstat News* column in article titled "Divining Difference" and a guest editorial titled "Can We Actually Measure Health Disparities?" in the Fall 1994 and Spring 2005 issues of the ASA publication *Chance*. That latter item also discussed the broader problems in the analysis of demographic differences by measures that tend to be affected by the prevalence of an outcome. I had also treated various related issues in presentations seven ASA conferences between 2001 and 2011 and would discuss the issues in many posts on an ASA member discussion forum beginning in 2014. Milo Schield, President of the National Numeracy Network and a Fellow of the American Statistical Association, gave a presentation on this subject at the organization's 2020 Joint Statistical Meetings titled "Statistical Literacy; Scanlan's Paradox."

[13] The request for recusal was based partly on matters discussed at pages 2-4 of the earlier letter in connection with a suggestion that the ASA President consider recusing himself. It was also based partly on the fact that on October 13, 2015, a statistician who apparently worked under supervision of the ASA President in the President's non-ASA position had submitted to the editor of *Society* a response to my "Race and Mortality Revisited," *Society* (July/Aug. 2014) (a principal reference in the October 8. 2018 letter as it is here) which I maintained mischaracterized the

26, 2016 letter discussed, inter alia, mistaken beliefs in recent statements by President Barrack Obama regarding the effects of generally reducing adverse criminal justice outcomes on measures of racial/ethnic disparity,[14] and urged the organization to explain the issue to President Obama.[15]

It is also possible that Mr. Santos would be familiar with the issues due to his role as Vice President and Chief Methodologist of the Urban Institute and the work or Urban Institute researchers Lisa Clemans-Cope, Matthew Buettgens, and Hannah Recht in a December 2014 Urban Institute titled "Racial Differences in Uninsurance Rates under the ACA – Are Differences Projected to Narrow?" The paper, which is discussed further in Section A, recognized (at 45 n.16) that general increases in insurance rates, while tending to reduce relative racial/ethnic differences in insurance rates, would tend to increase relative differences in uninsurance rates, as well as certain other things about differences in conclusions resulting from reliance on different measures, and appeared to recognize the need for a measures of demographic difference between outcome rates that did not change solely because the prevalence of an outcome changes.

But I have not seen awareness of the pattern by which the relative differences in favorable and adverse outcomes tends to change in opposite directions, or even an awareness that different measures will commonly yield opposite conclusions about directions of changes in disparities, in other Urban Institute research. See discussion of a more recent Urban Institute work on uninsurance disparities in Section A.4. As discussed further in Section A, absence of awareness that measures different from the one researchers happens to be examining, can, or commonly will, or in fact do, yield opposite conclusions from those being reported is something that might be said of the disparities research of almost every research institution in the world, as discussed with a particular focus on research of arms of Harvard University at pages 344-345 of "Race and

---

content of "Race and Mortality Revisited." The request specifically sought appointment of another ASA officer to oversee ASA's consideration of the issues raised in the October 8, 2015 letter. ASA's Executive Director subsequently advised that the ASA President was recusing himself from any role in the matter but that there would be no appointment of another officer to oversee it.

[14] President Obama's remarks were also the subject of my "Things the President Doesn't Know About Racial Disparities," Federalist Society Blog (Aug. 5, 2016).

[15] In September 2016, ASA staff, without apparent involvement of ASA elected leadership, advised that it would not take actions requested in my communications on the basis that the organization believed that I was effectively highlighting the issues raised in the communications and that it did not see an additional role ASA could take in disseminating information on the issues. I discuss the ASA refusal at pages 5-6 of my Comments for the Commission on Evidence-Based Policymaking (Nov. 14, 2016) – comments similar in many respects to my October 8, 2015 letter to ASA. The CEP Comments also noted that, notwithstanding the refusal of ASA to take the actions I requested, its leadership should be in a position to advise the Commission on Evidence-Based Policymaking on the validity of the points raised in my comments. I also discuss the ASA response, as well as the response to a similar request of the Population Association of American and the Association of Population Centers, towards the end of "The Pernicious Misunderstanding of Effects or Policies on Racial Differences in Criminal Justice Outcomes," Federalist Society Blog (Oct. 12, 2017).

Mortality Revisited," and as reflected in many of the letters collected on my Measurement Letters page.[16]

In any case, in my experience, an awareness of the aforementioned understanding of the effects of increasing insurance coverage such as reflected in the Urban Institute 2014 report does not automatically lead to an understanding that, for example, reducing police resources will tend to increase relative racial differences in adverse criminal justice outcomes, or that (as will be shown in Section B), increasing the level of paired tester qualifications, or sending the testers to areas of greater rather than lesser opportunity, will tend to reduce relative differences between rates at which testers of different demographic groups experience favorable outcomes while increasing relative differences between rates at which they experience corresponding adverse outcomes. That is the reason for the length of this memorandum.

***

Section A discusses the broader problems in the analysis of differences between outcomes rates of advantaged and disadvantaged groups without consideration of the ways measures tend to be affected by the prevalence of an outcome, with a focus on analysis of discrimination issues. It explains that that there have been no sound quantification of demographic differences involving outcomes rates – and hence no sound appraisal of the effects of policies on differences in the circumstances of advantaged and disadvantaged groups reflected by their adverse or corresponding favorable outcome rates and no sound drawing of inferences based on the comparative size of demographic differences in different setting or with respect to different issues. Section A also shows why it is when the identification or quantification of discrimination that is at issue that the unsoundness of standard of standard measures of demographic differences involving outcome rates, and the futility of choosing between or among such measures, is most evident. The section also explains more fully the pattern described in the *Amstat News* column, as well as the following corollary to the pattern: any general change in the prevalence of an outcome for two groups, or any factor that generally changes the groups' outcome rates in the same direction, will tend to cause a larger proportionate change in the rate for the group with the lower baseline rate for the outcome while causing a larger proportionate change in the opposite outcome rate for other group. That is, for example, anything that generally improves one's situation, like having high income rather than low income or receiving a beneficial health procedure, will tend reduce adverse outcomes rates proportionately more for whites than blacks,

---

[16] See, e.g., letters to Justice Research and Statistics Association (Feb. 18, 2020), National Quality Forum (Mar. 15, 2019), American Association for Cancer Research (July 10, 2018), National Institute for Minority Health and Health Disparities (July 6, 2018), National Quality Forum (Aug. 29, 2017), American Institutes for Research (Aug. 25, 2017), University of Oregon Institute on Violence and Destructive Behavior and University of Oregon Law School Center for Dispute Resolution (July 5, 2016), University of Oregon Institute on Violence and Destructive Behavior and University of Oregon Law School Center for Dispute Resolution (July 3, 2016), New York City Center for Innovation through Data Intelligence (June 6, 2016), Population Association of America and Association of Population Centers (Mar. 29, 2016), Council of Economic Advisers (Mar. 16, 2016), Stanford Center on Poverty and Inequality (Mar. 8, 2016), Agency for Healthcare Research and Quality (July 1, 2015), Institute of Medicine (May 28, 2014), Mailman School of Public Health of Columbia University (May 24, 2013), The Commonwealth Fund (June 1, 2010), Institute of Medicine (June 1, 2010), Robert Wood Johnson Foundation (Apr. 8, 2010), National Quality Forum (Oct. 22, 2009).

while tending to increase the corresponding favorable outcomes proportionately more for blacks than whites. Section A also explains an approach for quantifying demographic differences between outcome rates that is unaffected by the prevalence of an outcome that I have discussed and illustrated in many places since 2008 and illustrates that approach with respect to appraisals of changes in demographic differences in uninsurance rates among recent mothers in a 2019 Urban Institute brief.

Section B discusses measurement issues of special pertinence to paired tester studies. The section first explains that, while there yet been no sound quantification of treatment differences in paired tester studies that compared rate at which members of the tester pairs received or failed to receive the desired outcome, it is possible to soundly quantify treatment differences based on such rates if there exists complete information on the extent to which the members of the tester pairs were actually considered by decisionmakers. But it also explains that, for reasons that include but are not limited to issues I raised with respect to prominent Urban Institute paired tester study of hiring practices in "[Measuring Hiring Discrimination,](#)" *Labor Law Journal* (July, 1993), it may not be possible to soundly quantify differences in treatment of testers based on comparisons of percentages of cases where testers in the groups being compared appeared to be favored or appeared to be disfavored.

A. Problems in the Analyses of Demographic Differences in Outcome Rates by Measures That Tend to Be Affected by the Prevalence of an Outcome and Description of a Measure That Is Not Affected by the Prevalence of an Outcome

> **1. Description of the Patterns by Which Standard Measures Tend to Be Affected by the Prevalence of an Outcome and the Way Failure to Understand the Patterns Has Undermined Virtually All Analyses of Demographic Differences**

For reasons related to the shapes of most risk distributions, as the prevalence of an outcome and its opposite change, relative differences in the increasing outcome tend to decrease while relative differences in the decreasing outcomes tend to increase. Directions of changes in absolute (percentage point) differences between rates and differences measured by odds ratios are unaffected by whether one examines favorable or adverse outcome rates. But these measures tend also to change as the frequency of an outcome changes, though in a more complicated way that the two relative differences. Roughly, as an uncommon outcome (rates below 50 percent for both groups being compared) becomes more common, absolute differences tend to decrease; as common outcome (greater than 50 percent for both groups) becomes even more common, absolute differences tend to decrease. The prevalence/distribution-driven pattern is more difficult to predict when an outcome goes from being below 50 percent for both groups to being above 50 percent for one group or to being above 50 percent for both groups or from being above 50 percent for one group to being above 50 percent for both groups.[17] As the prevalence of an

---

[17] Of course, movements in the opposite direction have the opposite effects on the absolute difference between rates. It should be recognized that the described patterns of movement for one outcome reflect opposite patterns of movement for the opposite outcome. Thus, for example, when a common outcome becomes even more common and thus tends to reduce the absolute differences between rates, the opposite outcome is an uncommon outcome that is decreasing thus also reducing the absolute difference between rates. Declines in most types of mortality, for example, involve situations where survival is a common outcome that is becoming even more common, while mortality is an uncommon outcome that is becoming even less common.

outcome changes, differences measured by odds ratios tend to change in opposite directions from the absolute differences between rates.[18]

The described pattern regarding absolute differences means, for example, that improvements in care for conditions with low survival rates (either generally or for particular subgroups like older patients compared with younger patients) will tend to increase absolute differences between survival (and mortality) rates of different racial/ethnic groups, while improvements in care for conditions with high survival rates (either generally or for particular subgroups like younger patients compared with older patients) will tend to decrease absolute differences between survival (and mortality) rates of different racial/ethnic groups   Similarly, the increased availability or utilization of preventative or therapeutic procedures where rates for such procedures are low (either generally or for particular subgroups) will tend to increase absolute racial/ethnic differences between rates of receipt (and nonreceipt) of the procedures, while the increased availability or utilization of preventative or therapeutic procedures where rates for such procedures are high (either generally or for particular subgroups) will tend to reduce absolute racial/ethnic differences between rates of receipt (and nonreceipt) of the procedures.

One may better understand these patterns with reference to Table 5 of "Race and Mortality Revisited" while considering movement from the first to the second row as reflecting increases in uncommon outcomes and movement from the third to the fourth row as reflecting increases in common outcomes, or by reference to Table 22 (slide 66) of the University of Minnesota workshop, which presents the matter in terms of treatment rates. Graphical illustrations of changes in the patterns of all four measure may be found in Figure 4 (slide 39) of the workshop or on page 31 of the CEP Comments.

Any increase or decrease in the forces causing outcome rates of advantaged and disadvantaged group to differ will tend to increase or decrease all measures of difference between the groups' rates. Such changes thus will counter somewhat the tendency for a general change in the prevalence of an outcome to increase certain measures of difference and bolster somewhat the tendency for the general change to reduce other measures of difference. The change in the strength of the forces causing the rates to differ, however, may or may not be sufficient to cause departure from the usual pattern of changes in directions of measures arising from changes in the prevalence of an outcome. When there occurs a very large change in the prevalence of an outcome – as would occur, for example, with respect to unemployment in a major recession or infant mortality from over the course of a century – it would be rare for the prevalence-related patterns not to be observed notwithstanding significant changes in the strength of the forces causing the groups' outcome rates to differ. The matter could be quite different with regard to year-to-year changes, though in such cases the departures from the usual patterns are as likely to

---

[18] I use the phrase "differences measured by odds ratios" rather than simply "odds ratios" because every pair of rates for two groups can yield four odds ratios, two of which are above 1 and two of which are below 1, and the former of which are the same as each other and the reciprocal of the latter which are also the same as each other. See discussion in Section A.3 infra regarding the ratios of rates in Table 1 and the relative difference those ratios represent.

reflect irregularities in the underlying distributions as they are to reflect changes in the forces causing the groups' outcome rates to differ than.

All measures may change in the same direction as the prevalence of an outcomes changes, in which case one may cautiously infer that there occurred a genuine change in the comparative situation of two groups reflected by their favorable or corresponding adverse outcome rates. But in the common situation where all measures do not change in the same direction and the relative differences in the favorable and corresponding adverse outcome rates change in opposite directions, the absolute difference will tend to change in the same direction as the smaller relative difference. Observers who rely on relative differences tend to rely on whichever of the two relative difference (i.e., in the favorable or the corresponding adverse outcome) is larger. Thus, there is a tendency for observers who rely on absolute differences to systematically reach opposite conclusions about directions of changes in disparities from observers who rely on relative differences.

Discussions of changes in racial differences in poverty and unemployment are illustrative. Observers who rely on relative differences to discuss these matters always rely on relative differences in the adverse outcomes rather than relative differences in the corresponding favorable outcomes (i.e., avoidance of poverty or unemployment). But increasingly observers discuss demographic differences regarding these matters in terms of changes in absolute differences, especially in the case of unemployment. Yet, there has probably never been a major change in poverty or unemployment rates such as have occurred in major recessions or recoveries where the relative racial difference in poverty and unemployment rates and the absolute racial difference between such rates did not change in opposite directions.

It remains rare for researchers examining demographic differences to notice (or, in any case, to mention) that the relative difference they happen to be examining and the absolute difference have changed in opposite directions or show opposite patterns as to the comparative size of disparities in different settings, even when the researchers mention both the relative difference and absolute difference in the same study. In the cases where researchers have mentioned such a pattern – an occurrences largely limited to health and healthcare disparities research and given substantial attention in "Race and Mortality Revisited" (at 335-336) and "The Mismeasure of Health Disparities" (passim) – the researchers have almost never mentioned the relative difference for the opposite outcome.[19] This occurs even though anytime the relative difference being examined and the absolute difference have changed in opposite directions, 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. See CEP Comments at 15 n. 26. The same holds for comparisons of the size of demographic differences in different settings where the prevalence of the outcome differs from setting to setting.

_____

[19] Possibly the only instance where the second relative difference has been mentioned is found in the 2014 Urban Institute report by Clemans-Cope et al. That mention, however, was limited to a footnote (note 16 at 45). The body of the report is similar to the usual discussion of situations where the relative difference being examined and the absolute difference have changed in opposite direction without mention of the second relative difference. The paper does, however, importantly mention (at 6) that "as uninsurance rates for all groups decline, the relative difference between the groups will tend to increase (all other factors remaining equal)." That is an implied recognition that the premise of a vast body of research and the numerous civil rights law enforcement policies that reducing adverse outcomes would be expected to reduce relative differences in rates of experiencing the outcomes is incorrect.

In should be borne in mind that the point supported by the reference to the CEP Comments necessarily means that every case where a relative and absolute difference have changed in opposite directions, whether the pattern has been noted or not, involves a case where the relative difference in the favorable outcome and the relative difference in the corresponding adverse outcome in fact have changed in opposite directions.

Researchers who study racial difference in cancer outcomes commonly refer to disparities in survival even when they are in fact analyzing relative differences in morality. When cancer outcomes disparities research states or suggests that it is examining patterns of demographic differences in survival from cancer, a substantial majority of the time (though not always) the research is actually analyzing relative differences in mortality. And the authors involved in such research, like the editors and peer reviewers responsible for its publication, are invariably unaware that it is even possible for the relative difference in survival to show an opposite pattern as to the comparative size of disparities from the relative difference in mortality, much less that such is the usual case. Anytime it is reported that general improvements in cancer care were accompanied by increased racial or other demographic disparities in survival, or that survival disparities were greater among younger subjects than older subject or for more treatable cancers than less treatable cancers, there is a high probability that the researchers in fact examined relative differences in mortality and that the patterns for relative differences in survival are in fact the opposite of those being described. To a lesser degree, researchers studying infant outcomes have discussed increasing survival disparities in the face of general improvements in infant outcomes when they in fact examined relative differences in mortality and when the relative difference in survival actually decreased (as did the absolute difference between rates). See my Mortality and Survival webpage.

All the most provocative discussions of racial differences in infant mortality, such as in recent discussions about increases in such differences between 1850 and 2016 or between 1916 and 2017 involve situations where relative differences in infant survival and absolute differences between infant mortality (and survival) rates had decreased. Yet sometimes, as in a 2019 *Pediatrics* study,[20] researchers will discuss racial differences in infant mortality in absolute terms, and find reductions in disparities, while showing no awareness that the long line of provocative research it sites that found increases in racial disparities in infant mortality would instead have found decreases in such disparities if that research had measured disparities in terms of absolute differences between rates. See my Spurious Contradictions webpage regarding a

---

[20] See Boghossian NS, Geraci M, Lorch SA, et al. Racial and Ethnic Differences Over Time in Outcomes of Infants Born Less Than 30 Weeks' Gestation. Pediatrics. 2019; 144(3):e20191106. Reflective of the unawareness of implications of choice of measure on the part of the authors (and editors and peer reviewers) is that the study discusses decrease in disparities or larger reductions in rates for blacks and Hispanics than whites without mention of the measure employed. An understanding of the patterns described here would alert a reader to the fact that either the study was measuring racial/ethnic differences in absolute terms or that it had found a notable departure from the usual pattern. But it would require careful reading further into the article for clarification (as was done for me by the study's Figure 3 on the seventh page).

study that read two prior studies as showing opposite patterns of effects of increasing rates for uncommon procedures on changes in racial disparities among Medicare patients, and called for more research to resolve the differing findings, when in fact the prior studies produced similar results but simply reported them using different measures (relative differences in receipt of care versus absolute differences between rates).

Often when summarizing disparities research by others observers will discuss changes in disparities found in such research without mentioning the measure employed, reflecting the fact those summarizing the matters are unaware that different measures can (and typically do) show different patterns of changes in directions of disparities. An extreme example of this failure of understanding is the massive Agency for Healthcare Research and Quality (AHRQ)-funded study conducted by researchers from Vanderbilt University that examined thousands of studies to determine the effects of improvements in healthcare on racial disparities in healthcare while giving no attention whatever to the measures employed in the studies examined. See "Race and Mortality Revisited" (at 333) and my AHRQ's Vanderbilt Study webpage.

Some researchers, especially when discussing public school suspensions, have read decreases in absolute differences to mean that relative differences also decreased even though the data they cite show that the relative differences being discussed in fact increased. Most situations where researchers have stated or implied that a general reduction in school suspensions has been accompanied by a decrease in relative racial differences in suspension rates have involved situations where the relative difference in fact increased (and the researchers mistakenly assumed that the reduction in absolute differences was accompanied by a reduction in the relative difference). These situations are discussed on my Allegheny County (PA) Disparities, Denver Disparities,Oakland (CA) Disparities, Massachusetts Disparities, and Virginia Disparities pages).

And, as discussed in note 11 supra and the Appendix to the COPAA v. DeVos Federalist Society Blog post, the Commission on Civil Rights rejected my testimony that generally reducing suspensions would tend to increase relative racial differences in suspensions rates – and went on to further promote the mistaken belief that reductions would tend to reduce those differences – not the basis of any suggestion that I was incorrect on that point, but entirely on the basis of the fact that the absolute difference could decrease even as the relative difference increased. While that would seem a remarkable example of innumeracy, one observes comparable examples in peer-reviewed literature and in actions of putatively expert components of the federal health and healthcare establishment.

One important comparable example may be found in AHRQ's yearly National Healthcare Disparities Report's (NHDRs), which since 2013 are part of the National Healthcare Quality and Disparities Reports (NHQDRs). That matter, which can only be fully understood in the context of actions of NCHS, is addressed to a degree on my NHDR Measurement Issues page. It is also addressed in the pages immediately below, which may provide the reader some insight as to the prospects that a COVID-19 disparities task force such as proposed by the COVID-19 Racial and Ethnic Disparities Task Force Act, or any of the like task forces already implemented by state and local authorities, will provide useful insight into the ways policies affect differences in the circumstances of racial/ethnic groups with respect to COVID-19. The same applies to things the

Centers for Disease Control and Prevention (CDC) and other federal agencies are already doing regarding these issues.

As discussed in "Race and Mortality Revisited" and "The Mismeasure of Health Disparities," beginning in 2004, on the basis of my "Race and Mortality," *Societ*y (Jan./Feb. 2000)), and "Divining Difference," *Chance* (Fall 1994), beginning in 2004, NCHS recognized that as health and health care improved, relative differences in the increasing favorable outcomes tended to decrease while relative differences in the decreasing adverse outcomes tended to increase. Rather than do anything sensible in response to this recognition, however, NCHS simply determined that, for the purpose of evaluating the health disparities reduction goals of Healthy People 2010 both health and health and healthcare disparities would be measured in terms of relative differences in adverse outcomes. Thus, improvements in health and healthcare would tend to be associated with increasing demographic disparities as to both health and healthcare. In the case of situations where healthcare indicators had been cast in terms of receipt of care, however, the indicators would continue to be cast in favorable terms, even though the disparity would be quantified in terms of relative differences in the corresponding adverse outcome. Thus, for example, if the relative difference in nonreceipt of mammography or immunization increased, it would be reported that relative difference in receipt of mammography and immunization had increased, even though the relative difference in receipt of mammography and immunization had actually decreased. The size of changes in the relative differences in the adverse outcomes were quantified in percentage point differences between prior and current relative differences. That is, an increase in a relative difference from 150 percent (as would be reflected by a rate ratio of 2.5) to a relative difference of 160 percent (as would be reflected by a rate ratio of 2.6) would be treated as involving a 10 percentage point increase in the relative difference. See note clarifying terms in Table 1 in Section A.3 *infra*.

As discussed in ""The Mismeasure of Health Disparities," however, in 2016, NCHS changed its recommendation with respect to which relative difference in healthcare outcomes would be examined for purposes of appraising progress in meeting Healthy People 2020 disparities reduction goals. That is, healthcare disparities would now be measured in terms of relative differences in favorable outcomes, though disparities in health outcomes would continue to be measured in terms of relative differences in adverse outcomes. In illustrating the point, the guide (at 12) would specifically explain that disparities relating to insurance would be measured in terms of relative differences in receipt of insurance. So, while improvements in health would still tend to be associated with increased disparities, improvements in healthcare would tend to be associated with reduced disparities. It warrants note, however, that, like all NCHS disparities material since about 2005, the explanation was not presented in a way that readers would necessarily understand that the change could result in a reversal of identified changes in directions of disparities. Unless a reader at least understands that is possible for relative differences in favorable outcome and relative differences in corresponding adverse outcome to change in opposite directions, the reader will likely regard the revision as a cosmetic matter of no particular consequence.[21]

---

[21] The reader should keep in mind that if there is a relative difference adverse to a particular group as to one outcome there will also be relative difference adverse to that group for the opposite outcome (though quantification of the relative difference will vary), something that all researchers and others readily understand. It is the fact that when the prevalence of an outcome changes the sizes of the relative differences in the two outcomes tend to change

The confusion arising from NCHS efforts to provide guidance on the measurement of disparities is well reflected in Table 3 of the FCSM Paper (at 16-17). That table illustrates a situation where a peer-reviewed article by leading authorities on the measurement of health and healthcare disparities reported in an abstract that major increases in mammography rates were accompanied by dramatic increases in relative socioeconomic difference in mammography rates, when in fact there had occurred dramatic decreases in relative socioeconomic difference in mammography rates. The article eventually made clear it was actually measuring relative differences in nonreceipt of mammography, while also indicating that it was specifically relying on NCHS guidance in doing. But it did not do so in a way that would likely alert readers that reporting on nonreceipt of mammography rather than receipt of mammography reversed the conclusion about directions of changes in the size of disparities. As a result of the change in policy in 2016, however, though it would remain unknown to readers of the study, reliance on NCHS guidance would now result in the reportage of a dramatic decrease in relative socioeconomic differences in mammography over the period covered in the study.[22]

See also CEP Comments (at 25-26) regarding the way that in a 2016 report of the Stanford Center on Poverty and Inequality, researchers, relying on previous NCHS guidance to measure disparities in insurance coverage in terms of relative differences in uninsurance rates, drew inferences about underlying processes on the basis of the comparative size of relative differences in uninsurance in different settings. The authors were likely unaware that NCHS had already reversed the guidance, and, like virtually all disparities researchers, were unaware that comparative size of relative difference in insurance rates could, or would usually, support very different inferences about processes from those supported by the comparative size of relative differences in uninsurance rates.

From 2005 on, in the NHDRs, AHRQ followed the NCHS Healthy People 2010 guidance, and also measured both health and healthcare disparities in terms of relative differences in adverse outcomes. Thus, like NCHS, AHRQ would tend to find improvements in health and healthcare to be associated with increased disparities for both health and healthcare. It also followed NCHS guidance to report indicators in terms of a favorable outcomes even though the NHDR was actually measuring the relative differences in the corresponding adverse outcome. AHRQ also

---

in opposite directions that few understand. Any discussion of differences in quantification of the two relative differences tends to leader reader away from, rather than toward, an understanding that the two relative differences can, and usually will, change in opposite directions as the prevalence of an outcome changes. The occasional discussions of the fact when rates are very high for an outcome and very low for the opposite outcome the relative difference for the former is small and the relative differences for the latter is large ought to alert readers to the fact that altering the prevalence of the outcome will tend to increase one relative difference while reducing the other relative difference. But, so far as is revealed in published literature, that seems never to have happened.

[22] See also the discussion of Table in Table 3 (at 17) of "The Mismeasure of Health Disparities" regarding the way that since 2016, NCHS would agree with findings of a 2008 study in *Pediatrics* that, ignoring or unaware of NCHS guidance issued in 2004-05, found that increase in Hepatitis B vaccination following a school-entry vaccination requirement were associated with reduced racial/ethnic disparities in vaccination rates (whereas at the time of study NCHS would have reached an opposite conclusion).

followed NCHS guidance for measuring changes in disparities in terms of percentage point changes in relative differences.

But at least by the time of 2012 NHDR, while continuing to quantify health and healthcare disparities in terms of relative differences in adverse outcomes, AHRQ analyzed directions of changes in disparities on the basis of the comparative size of the absolute changes in each group's rate (which, with respect to determining directions of changes in disparities, is the equivalent of measuring the absolute differences between rates at the two points in time). Given the ranges in which one finds most (though not all) adverse health outcomes, the AHRQ approach would result in a situation whereby, in circumstances where adverse outcomes rates decreased, the NHDR would tend to find that the disparities decreased on the basis of larger absolute reductions in rates for the disadvantaged group than the advantaged group – even though, on the basis of an increase in the relative differences in the adverse outcome, AHRQ would also find that the disparity was larger at the end of the period than at the beginning of the period.[23] In the case of the aforementioned appraisals of changes in infant mortality disparities between 1850 and 1916 and between 1916 and 2017, AHRQ would find the disparities to be larger at the end of the periods than at the beginning of the periods; but it would also find that the disparities declined substantially over those periods.

Given the way patterns of changes in absolute differences turn on the rate ranges at issue, the matter is more complicated with regard to healthcare disparities. Again, regardless of the rate ranges involved, AHRQ would tend to find improvements in care to result in larger disparities at the end of a period than at the beginning of the period (on the basis of increases in relative differences in nonreceipt of care). And when uncommon favorable outcomes increased (though not enough to become no longer uncommon), AHRQ would tend also to find an increase in the disparity on the basis of a larger absolute decrease in the adverse outcome rates for the advantaged group than the disadvantaged group (or the corresponding increase in the absolute difference, which, again, holds whether one examines the favorable or the adverse outcome).

But when a common favorable healthcare outcome generally increased and absolute differences between rates tend to decrease, AHRQ would tend to find that the disparity had decreased on the basis of a larger absolute decrease for the adverse outcome for the disadvantaged group than the advantaged group. In fact the 2012 NHDR highlight as some of the largest decreases in disparities (on the basis of reductions of absolute differences between rates, though calculated by comparing the size of absolute reductions in adverse outcome rates for the groups being compared) situations where the agency would also regard the disparity to be much larger at the end of the period than at the beginning of the period. See Table 6 (at 21) of my July 1, 2015 letter to the Director of AHRQ. The fourth and fifth rows reflect a situation where, with respect to the indicator described as "Hispanic-white disparity for hospital patients age 65+ with pneumonia who received a pneumococcal screening or vaccination," the 2012 NHDR highlighted the reduction in disparity between 2005 and 2010 on the basis of the 13 percentage point greater decline in the adverse outcome for Hispanics than whites (reducing a 19 percentage

---

[23] For adverse health outcomes like mortality from certain cancers and deaths among extremely vulnerable persons infected with COVID-19, decreases in mortality would tend to result in smaller absolute reductions for the disadvantaged group than the advantaged group. In those cases, AHRQ would tend to find both that the disparity increased and that the disparity was larger after the improvement than before the improvement.

point difference to a 6 percentage point difference), even though the agency would regard the relative difference between adverse outcome rates to be 53 percent in 2005 and 113 percent in 2010 (which NCHS would describe as a 60 percentage point increase in the relative difference while others might term a 113 percent increase in the relative difference[24]).

As reflected on my NHDR Measurement Issues page, following receipt of my July 1, 2015 letter, the agency made an effort to understand this problem. But, as of the 2018 National Healthcare Quality and Disparities Report, the agency was still measuring the size of disparities in terms of relative difference in adverse outcome but still measuring changes in disparities in terms of the comparative size of absolute changes in each group's adverse outcomes rate (though with Appendixes that make in clearer, at least to me, that such was the approach than the 2012 NHDR had done).

It took me a long time to figure out how AHRQ was measuring healthcare disparities in the NHDRs. This was partly due to the statement of indicators in terms of favorable outcomes (and many bar charts comparing favorable outcome rates) when the disparities were in fact being measured in terms of relative differences in the corresponding adverse outcome, partly due to the agency's use of the term percent when it meant percentage points, and partly due to the agency's failure to accurately describe what it was doing or my own failure to understand those descriptions. Possibly, the change in approach may been related to confusion about the measures of changes in the size of relative differences in absolute terms and changes in each group's rate in absolute terms.

It should be kept in mind, however, that AHRQ has never indicated an awareness that it is even possible for relative differences in the favorable outcome and relative difference in the corresponding adverse outcome to change in opposite directions. And, while the 2005 NHDR did mention that the relative difference and the absolute difference could change in opposite directions (though noting that such occurrences were uncommon), the reports have never since indicated such an awareness even as the reports use one measure to quantify the size of disparity at any point in time and another measure to determine whether the disparity had increased or decreased between two points in time.

The yearly NHDRs/NHQDRs presumably consume great resources that provide little information of value and a great deal of information that is misleading. Probably, however, the greater waste of resources due to AHRQ failures of understanding is the agency's funding of research like the Vanderbilt study, where AHRQ personnel awarding and monitoring such funding, like the recipients of funding, are unaware of the measurement issues that must be addressed for such studies to be useful. This problem applies to research funded by all federal agencies, including presumably, HUD. See discussion at page 4 of my Letter to Departments of Education, Health and Human Services, and Justice (July 17, 2017) regarding a moratorium on funded research that fails to consider how measures employed tend to be affected by the

---

[24] AHRQ itself would have measured the change on the basis of the 60 percentage point decrease though it would have termed it a "percent" increase. See my Percentage Points webpage.

prevalence of an outcome. See also my recommendations about requirement for recipients of federal funds at pages 45-46 of my CEP Comments.

The only correction for this situation would seem to be a requirement that parties seeking federal funds for research into demographic difference explain the measures they intend to use to quantify such differences and to explain as well how they intend to address issues arising from the way the measure tends to be affected by the prevalence of an outcome.

2. Some Illustrations of the Misunderstanding of Racial Disparities Issues in the Minneapolis, Minnesota Area

In the Introduction I suggested browsing the slides of my 2014 workshop at arms of the University of Minnesota as a means of gaining familiarity with the issues to be discussed, noting that it contains some slides of particular relevance to interpretations regarding housing/neighborhood issue. The workshop arose out my commentary "It's easy to misunderstand gaps and mistake good fortune for a crisis," *Minneapolis Star Tribune* (Feb. 8, 2014), which pertained to the failure to understand that advantaged places like Minnesota, Norway, and Sweden, where adverse outcome rates were comparatively uncommon, tended to have comparatively large relative racial or socioeconomic differences in adverse outcomes, but comparatively small relative racial and socioeconomic difference in the corresponding favorable outcomes.[25] The same point could as well be made about the failure to recognize that extremely large relative differences in adverse outcome (along with the extremely small relative differences in the corresponding favorable outcomes) are commonly functions of the fact that the adverse outcome is extremely rare (as discussed with regard to racial/ethnic differences in COVID-19 mortality among young people in the Introduction and in Section A.4 *infra* regarding racial differences in maternal mortality).

Slide 72 discusses a Minneapolis Urban League study of the comparative effects of location on blacks and whites with respect to both (a) home ownership and (b) rejection of mortgage loan applications. The study analyzed the former in terms of the comparative size of relative effects in a favorable outcome (where blacks had the lower baseline rates) and the latter in terms of the comparative size of relative effects on an adverse outcomes (where whites had lower baseline rates). But it did so without recognizing that a factor that affects an outcome rate tends to cause a larger relative effect on the rate for the group with the lower baseline rate for the outcome (while causing a larger relative effect on opposite outcome for the other group). That is, a study that measured things as this study did would usually show larger location effect on home ownership for blacks than whites but larger location effects on mortgage rejection rates for whites than blacks (something that will be also reflected in Table 3 infra with regard to the effects of indicating a criminal conviction on the resumes of one of the applicants in employment tester applicant pairs comprised or two black applicants and two white applicants).   So there would be no basis for drawing inferences about processes based on the observed patterns. See

---

[25] As reflected title of the University of Massachusetts Medical School seminar ("The Mismeasure of Health Disparities in Massachusetts and Less Affluent Places"), the pertinence of that pattern to a particularly advantaged place like Massachusetts was also the focus of that seminar.

slide 20 regarding fact that the pattern of relative effects on lower and higher baseline outcomes rates is corollary to the pattern whereby the rarer an outcome the greater tends to be the relative difference in rates of experiencing and the smaller tends to be the relative difference in rates of avoiding it.   See also the discussions of subgroup effects in "Race and Mortality," "Divining Difference," *Chance* (Fall 1994) and my Subgroup Effects and Subgroup Effects – Nonclinical. pages.

The misinterpretation of data on subgroup effects is a quite serious problem that I barely touch upon here.[26]   Thus, I suggest careful reading of the section titled "Illogical Expectations and Unfounded Inferences" at pages 339-341 of "Race and Mortality Revisited" and careful consideration of the point in the April 28, 2020 BMJ response regarding the way that data on the effects of COVID-19 therapies on different races will be misunderstood.

Slide 73 discusses a 2010 study of housing issues in St. Paul where the author regarded the fact that a very high proportion of vacant buildings was in poor neighborhoods as a consequence of a recession, without understanding that recessions tend to reduce such proportions while recoveries tend to increase them. As a result of the economic downturns in the Minnesota area due to COVID-19-related shutdowns, the proportion vacant buildings in poor neighborhoods makes up of total vacant buildings in the Minnesota area will tend to decrease. But, as in all situations, many factors apart from changes in the prevalence an outcome can affect these patterns. And in Minneapolis and St. Paul, in particular, the locations of vacant buildings will be much affected by the location of recent or future destructive riots.

I add here the following points of particular pertinence to the Minneapolis area and Minnesota generally, which, due to recent and ongoing events, are likely to be continuing focal points for discussions of race and racism in American, as well as debates about consequence of limiting police resources. Slide 73 of the Minnesota workshop discussed the way that in Minneapolis and St. Paul recent general reductions in suspensions had in fact been accompanied by increased relative racial differences in suspensions. Since that presentation, however, such pattern has emerged in many more jurisdictions than indicated in the slide or as indicated in my December 2017 testimony for the Commission on Civil Rights (as most recently summarized in the introductory material to my Richmond Disparities page). Notably, as discussed on my Minnesota Disparities page, some of the most striking evidence of the usual effect of reducing suspensions on relative racial differences in suspension rates may be found in a Hechinger Report study's finding that in all 73 districts in Minnesota where the matter could be analyzed

---

[26] The analysis of subgroup effects (also termed interaction or effect heterogeneity) in clinical trials is an important issue about which a great deal is written has been written in medical and health policy journals. None of the literature has yet recognized that it is even possible for a factor to cause a larger proportionate effect on one outcome (e.g., mortality) for one group and a larger proportionate effect on the opposite outcome (e.g., survival) for the other group. Anytime a study states that it found that an interventions increased survival from cancer more for a younger group than an older group, the study will almost certainly have in fact examined relative effects on mortality rather than survival and study results will show that the intervention increased survival proportionately more for the older group than the younger group.

general reductions in suspensions were accompanied by increases in the ratio of the black suspension rate to the white suspension rate.[27]

Further, immediately after death or George Floyd, the Minnesota Department of Human Rights (MDHR) was ordered to study racial disparities in actions of the Minneapolis Police Department (MPD), which, for reasons already stated, can only be useful if conducted with an understanding of the effects of the prevalence of an outcome on measures of racial disparity. In a June 8, 2015 letter to the Minneapolis Mayor and Chief of Police, I explained, with respect to a recent ACLU study calling for reductions in arrests for minor offenses because of racial differences in arrests for such offenses, that the study's premise that reducing arrests would tend to reduce (rather than increase) relative racial differences in arrests and the proportion blacks made up of arrested persons was incorrect. In letters of August 27, 2018, and May 14, 2018, I attempted to explain to MDHR that generally reducing suspensions would tend to increase, not reduce, relative racial differences in suspension.

It is doubtful, however, that anyone in the Minneapolis Mayor's Office or MPD understood these issues in 2015 or understands them now. And MDHR actions regarding school discipline disparities subsequent to receipt of my letters makes clear that the agency is continuing to base its activities regarding school discipline on the mistaken premise that reducing suspensions will tend to reduce relative racial differences in suspensions.[28] Thus, if the study that was subject of my 2015 letter in fact led to general reductions in arrests or less aggressive policing, relative racial differences in adverse policing outcome may well have increased precisely for that reason. And MDHR might well inflame racial tensions by reporting that relative racial differences in adverse outcomes increased despite general reductions in those outcomes, when in fact the increases occurred because of the general reductions.

Many similar interpretative issues will exist going forward where calls for defunding of police in Minneapolis and elsewhere result in reductions in police resources and modifications to practices

---

[27] The report merely stated that the suspensions rates were reduced more for whites than blacks, which I take for granted meant a larger relative reduction for whites than blacks and thus an increase in the relative difference between black and white suspension rates. If there had been a larger absolute reduction for whites than blacks, however, that that would necessarily also mean larger relative reduction for whites than blacks (assuming that whites originally had lower suspension rates). But, as discussed in note 11, I assume that in high proportion of the 73 case, blacks had a larger absolute reduction in suspensions and that the absolute difference between black and white rates decreased. The Hechinger Report study be compared to the study of suspension disparities in Maryland discussed bn my "Maryland Discipline Study Shows Usual – But Misunderstood – Effects of Policies on Measures of Racial Disparity," *Gunpowder Gazette* (Dec. 16, 2019). An Appendix in the Maryland study showed general decreases in suspension in Maryland school districts were accompanied by increased relative differences between the suspension rates of blacks and other students in 20 of the 23 districts where data could be analyzed. The Maryland study shows that, while reductions in suspensions will not always be accompanied by increased relative racial differences in suspension rates, that typically will happen. In the three cases where the relative difference decreased, the absolute difference would necessarily have decreased. Presumably, the absolute difference also decreased in a high proportion of the 20 cases where the relative difference increased.

[28] See my letters to Walker Hackensack Akeley Public Schools (Sept. 10, 2018), and Rochester Public Schools (Sept. 9, 2018).

that then lead to fewer arrests and prosecutions. That will tend to result in increased relative racial differences in rates of experiencing such outcomes, though virtually everyone discussing the situation, including MDHR, will expect the opposite.   In such circumstances, we will observe the interaction of the effects of general decreases in police activities on measures of racial disparity and the effects of any changes in the severity of any bias involved in police activities and any race-conscious action taken aimed at reducing racial differences irrespective of any role of bias in such differences. But only persons who understand the effects of changes in the prevalence of an outcome on measures of disparity will be able usefully analyze data on these issues. [29]

Meanwhile, MDHR will presumably continue to examine relative racial differences in school suspensions. In doing so, it may discuss increased relative racial differences in suspensions as things that are occurring despite general in suspension rather than as a result of MHDR's aggressively pressuring schools to generally reduce suspensions as well as activities voluntarily undertaken by school administrator who have been led wrongly to believe that reducing suspensions will tend to reduce, rather than increase, relative racial differences in suspensions.

3. A Simple Illustration of Problems in Identifying Which of Two Decisionmakers Is More Likely to Have Engaged in Discrimination Using Measures that Yield Opposite Conclusions

In the Introduction I made the point that it is when discrimination is at issue that the unsoundness of standard measures of differences between outcome rates is most evident. I make the point with reference to in "Race and Mortality Revisited" (at 335-36) with regard to the illustration in the article's Table 5, while explaining the fallacy that a value judgment is involved in choosing between or among measures that yield opposite conclusions about changes in disparities over time or about the size of a disparity in one setting compared with the other. In doing so, I also explain the varying mistaken inferences observers draw about processes, including discrimination in the process, based on the size of whichever of the two relative differences the observer happens to be examining. I encourage readers to review that material carefully, as well as the discussion of Table 2 (at 339-340, 343) regarding the absence of any in value in exploring

---

[29] As I discuss in such places as "United States Exports Its Most Profound Ignorance About Racial Disparities to the United Kingdom," Federalist Society Blog (Nov. 2, 2017), and "Compliance Nightmare Looms for Baltimore Police Department," Federalist Society Blog (Feb. 8, 2017), in the case of criminal justice outcomes, there are reasons why one might commonly see departures from the patterns I describe because of the way reductions may be focused on particular types of crimes or limiting police presence in certain neighborhoods. The same hold for school discipline. An illustration may be found in the data in the study discussed in my Oakland (CA) Disparities page that was responsible for leading Oakland Unified Schools District mistakenly to believe that restorative justice program reduced relative racial differences in suspensions. Table 7 of the study showed that reductions in suspensions led to increased relative racial differences in suspension for offenses termed disruption/defiance, offenses termed serious, and for all offenses combined. But that table would also show that if the school district eliminated all suspensions for offenses termed disruption/defiance, relative racial differences for total offenses would decrease. A much larger reduction in the disruption/defiance offense than actually occurred could similarly shift the overall pattern notwithstanding than the pattern continues to be observed within each offense type category. While such things can occur and all sorts of things can occur with any aggregation of disparate data components. It nevertheless remains a great mistake to operate on the belief that reducing any outcome will tend to reduce relative racial differences in rates of experiencing the outcome.

reasons for changes in measures of racial differences relating to poverty without an understanding of what census data make clear about the way general changes in poverty affect these measures.[30]

I discuss Table 5 with regard to four levels of prevalence, thus illustrating somewhat the complicated ways that absolute differences between rates and differences measured by odds ratios tend to be affected by the prevalence of an outcome.[31]   But the simpler illustration in Table 1 below, which is a version of Table 1 of "Race and Mortality Revisited," can satisfactorily illustrate the key points for instant purpose. The table may be compared to Table 1 of my written testimony before the Commission on Civil Rights, which is one of the tables that Commissioner Kirsanow included in his dissenting opinion to the *Beyond Suspensions*, except that the instant table includes that absolute difference between rates in the final column.[32]

---

[30] However clear census data make these issues, the issues are entirely unknown to the Census Bureau just as they entirely unknown to every government agency, except possible a few people at the NCHS.

[31] The values chosen, however, poorly illustrate the point with regard to the odds ratio. In the prevalence ranges shown in the table, the odds ratios show little variation from row to row. But when overall rates are very high for any outcome, with corresponding very low rates for the opposite outcome, differences measured by odds ratios operate just like the larger of the two relative differences. See the extreme rows or Table 1 (at 24) of "The Misinterpretation of Health Inequalities in the United Kingdom," British Society for Population Studies Conference 2006 (Sept. 18-20, 2006), as well as Figure 3 (slide 35) of the of Minnesota workshop, which involve the same specifications as Table 5 of "Race and Mortality Revisited."

As it happens, odds ratios are often criticized because other than when rates for the outcomes being examined are very low (which means that the opposite outcome rates are very high) and the odds ratio approximates a risk ratio, odds ratio tend to exaggerate relative differences. See, e.g., this September 15, 2020 JAMA Network article, which, in addition to restating the common but mistaken belief that ratios tend to be constant across different baseline rates, confusingly uses "%" to reflect both percent effects and percentage point effects (see my Percentage Points webpage). It is actually when the rates are in middle ranges, as in the second and third rows of Table 5, that odds ratios are useful for quantifying the strength of an association (and where the standardized mean difference derived from the odds ratio yields a value very close to probit d'). It is when outcomes are very low (with the corresponding opposite outcomes very high), that by approximating the relative difference for the rarer of the two outcomes, the difference measured by odds ratios merely approximates an unsound measure of association. If underlying risk distributions are poisson distributions, however, odds ratios, or measures that are functions of odds ratios like the standardized mean difference, would seem to provide the best way of quantifying a strength of association involving outcome rates (though I have little understanding of the poisson distribution or of what properties of distributions cause either odds ratios of logarithms to operate as they do).

[32] Mr. Kirsanow also presented Table 2 from my testimony. The table included rows showing that lowering the cutoff increased the proportion DG made up persons who passed the test and the proportion DG made up of persons who failed the test. Discussions of effects of changes in the prevalence of an outcome on the proportion a group makes up of persons experiencing a favorable and corresponding adverse outcome complicates explanations of effects of policies on measures of differences between outcome rates. Thus, in less extensive discussions of this subject, I avoid treatment of such proportions other than in contexts when demographic differences are commonly discussed with reference to such proportions (as in the school discipline context on which the Commission on Civil Rights briefing was focused). My December 13, 2018 letter to HUD does not mention such proportions because demographic differences regarding borrower outcomes are not commonly discussed with references to such proportions.

The table shows the pass and fail rates of an advantaged group (AG) and a disadvantaged group (DG) at two cutoff points in a situation where the groups have normally distributed test scores with means that differ by half a standard deviation (a situation where approximately 31 percent of DG's scores are above the AG mean) and both distributions have the same standard deviation. The table also shows (in columns 5 through 7) three measures that might be used to quantify differences in test outcomes of AG and DG.

Table 1. Illustration of effects of lowering a test cutoff on measures of differences in test outcomes

Row	(1) AG Pass Rate	(2) DG Pass Rate	(3) AG Fail Rate	(4) DG Fail Rate	(5) AG/DG Pass Ratio	(6) DG/AG Fail Ratio	(7) Abs Df
1	80%	63%	20%	37%	1.27	1.85	17 PP
2	95%	87%	5%	13%	1.09	2.60	8 PP

Column 5, which presents the ratio of AG's pass rate to DG's pass rate,[33] shows that at the higher cutoff, where pass rates are 80 percent for AG and 63 percent for DG, AG's pass rate is

---

[33] While I commonly refer to patterns of relative differences in this memorandum, the table actually presents ratios of two rates that I commonly refer to as rate ratios (and that are also termed, probably more precisely, risk ratios or relative risks). 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. In the former case, the larger the rate ratio, the larger the relative difference; in the latter case, the smaller the rate ratio, the larger the relative difference.   It may be more common to employ the disadvantaged group's rate as the numerator for ratios regarding the favorable as well as the adverse outcome, which is the approach as to favorable outcomes of the "four-fifths" or "80 percent" rule for identifying disparate impact under the Uniform Guideline for Employee Selection Procedures. I have sometimes employed this approach, as in "Can We Actually Measure Health Disparities?," *Chance* (Spring 2006). More recently, however, I have usually used the larger figure as the numerator for both rate ratios, in which case, as to both favorable and adverse outcomes, the larger the ratio, the larger the relative difference. Choice of numerator in the rate ratio, however, has no bearing on the patterns by which as the prevalence of an outcome changes, the two relative differences tend to change in opposite directions. Further, I never refer to the patterns by which changes in the prevalence of an outcome affects ratios because the accuracy of the statement turns on which figure is used as the numerator in the ratios. Instead, I refer to patterns of changes in the relative differences that the ratios represent.

Further regarding terminology relating to differences between rates, it is common to refer to ratios of two rates (however characterized) and absolute differences as measure of association in very authoritative in places. Sometimes they are referred to as risk ratios and risk differences in such places. But only the absolute difference (or risk difference) is actually a measure of difference between two rates. The ratio of two rates is a means of quantifying or representing the relative difference between two rates. Confusion about this subject leads some observers to refer to relative risks, relative disparities, and relative differences interchangeably (as in the reference on which the 2014 Urban Institute paper relies (at 6) to term the "relative risk" a measure of association, without recognizing that the relative risk was simply the ratio used to calculate the relative difference that the report was already discussing). Conflation of these terms that mean different things causes a good deal of confusion and sometime leads observers to add an additional 100 percentage points to any relative difference (see my Times Higher webpage). It also sometimes leads observers to quantify a change in disparities in terms of the percentage change in the ratio that represents the relative difference rather than in the relative difference itself, as, for example, by characterizing a reduction of a rate ratio of 2.0 to a rate ratio of 1.0 as a 50 percent reduction in the disparity, when in fact such a reduction would entirely eliminate the disparity by reducing the relative difference by 100

1.27 times (27 percent greater than) DG's pass rate. If the cutoff is lowered to the point where AG's pass rate is 95 percent, DG's pass rate would be about 87 percent. At the lower cutoff, AG's pass rate is only 1.09 times (9 percent greater than) DG's pass rate.

That lowering a cutoff tends to reduce relative differences in pass rates is well understood and underlies the widespread view that lowering a cutoff tends to reduce the disparate impact of tests on which some groups outperform others.

But, whereas lowering a cutoff tends to reduce relative differences in pass rates, it tends to increase relative differences in failure rates. As shown in column 6, initially DG's failure rate was 1.85 times (85 percent greater than) AG's failure rate. With the lower cutoff, DG's failure rate is 2.6 times (160 percent greater than) AG's failure rate.

Thus, lowering the cutoff and making test passage more common and test failure less common decreased the relative difference in the increasing outcome but increased the relative difference in the decreasing outcome. This pattern hold across the entire range of possible test scores.

Column 7 shows that the absolute difference decreased from 17 to 8 percentage points. The decrease, however, is a function of the rate ranges at issue and would not be observed when, for example, a high cutoff is lowered slightly (and where the absolute difference would tend to increase), as reflected by movement from the first to the second row of Table 5 of "Race and Mortality Revisited," But the rates shown in the table are in the ranges usually found for many matters where demographic differences, including loan rejections, foreclosures and evictions, unemployment, and poverty, are commonly quantified in terms of relative differences in the adverse outcomes (though sometimes in terms of absolute differences between rates).

The following things must be kept in mind in appraising the significance of the pattern illustrated in Table 1 and other illustrations based on hypothetical test score data. Agencies like the Departments of Education and Justice have not reasoned as follows: while lowering a test cutoff and thus decreasing test failure will tend to increase relative differences in test failure rates, for this or that reason, one should not expect a similar result from reducing adverse school discipline or criminal justice outcomes, and in fact should expect relative differences in those outcomes to decrease. Rather, despite decades of dealing with racial differences in test outcomes, the agencies have yet to show even an understanding that lowering a test cutoff – or improving test performance or allowing the retaking of a test – will tend to increase relative differences between the failure rates of higher- and lower-scoring groups. Similarly, notwithstanding the connection between test scores and student proficiency, observers who analyze proficiency disparities in

percent (and making it zero). The same issue applies to odds ratios, which are not measures of differences between two odds, but means quantifying, or representing, the relative difference between the two odds.

I might be faulted for presenting rate ratios in tables and charts when I am discussing relative differences. But I do so because observes are used to seeking disparities presented in terms of ratios. And discussions (and illustrations of) whether the ratios increase or decrease is a simple way of illustrating changes in directions of relative differences, even if not everyone will immediately understand that a reduction in of a ratio of 3.0 to a ratio of 2.0 is a reduction in the 200 percent relative difference represented by the first ratio to a 100 percent relative difference represented by the second ratio.

terms of relative difference in the favorable outcomes or the corresponding adverse outcomes have yet to realize that general improvements or general worsening of education, or changing to an easier or harder test, will tend to cause relative differences in meeting standards to change in opposite directions from relative differences in failure to meet the standard. This occurs, even though the test score data on which they rely make should make such patterns evident. See my [CUNY ISLG Equality Indicators](#) webpage.[34]

In the case of illustrations based on income and credit score data, such as in Tables 1 and 2 of the December 13, 2018 letter to HUD officials, there is no issue of whether the patterns observed in hypothetical data would be expected to occur in other contexts. Rather, the tables directly demonstrate that lowering income and credit score requirements will tend to increase relative racial differences in rates of failure to meet the requirements. Nevertheless, for more than a quarter century the numerous agencies enforcing federal fair lending laws have proceeded according to the belief that lowering the requirements would tend to reduce those differences.

It also warrants mention that in the hundreds or thousands of instances where, in peer-reviewed journals and elsewhere, observers have reported that a relative demographic difference in rates of experiencing an adverse outcome increased despite a general decrease in the outcome – and where "despite" reflects the view that the reduction should have reduced the relative difference – no one has ever posited a reason why one would expect reduction in the outcome to reduce the relative difference. Rather, the expectation has simply been taken for granted. Meanwhile, no one regarded the myriad instances where the opposite occurred as a reason for questioning the validity of the expectation. Each departure from expectation was instead deemed to make the finding more significant.

It is also important to understand that neither the increases in one measure of disparity nor the decreases in another measure of disparity that occur solely because of a change in the prevalence of an outcome means that disparity has increased or decreased in any meaningful sense. Rather, the change in the prevalence of an outcome merely caused one unsound measure of disparity to change in one direction while causing another unsound measure of disparity to change in another direction. One must look further to determine whether there was any actual change in the

---

[34] The page discusses the Equality Indicators project of the Institute for State and Local Governance (ISLG) at the City University of New York, which, measuring racial/ethnic differences in proficiency in New York City schools in terms of relative differences in failure to meet proficiency standards, and thus would tend to reach opposite conclusions about changes in racial/ethnic differences in proficiency from those the New York City government's Center for Innovation through Data Intelligence would reach (based on relative differences in meeting proficiency standards). The ISLG Equality Indicators project has provided guidance on adapting its 96 equality indicators for use in five other cities involved in a project funded by the Rockefeller Foundation. Four of the cities measure racial/ethnic proficiency disparities in terms of relative differences in meeting proficiency standards and thus tend to reach opposite conclusions about changes from those ISLG would reach, while one measures racial/ethnic proficiency disparities in terms of relative differences in failure to meet proficiency standard and thus tends to reach the same conclusions ISLG would reach. The ISLG Equality Indicators project might be compared to the health and healthcare disparities research activities of the National Quality Forum discussed in my letter to that organization of [March 15, 2019](#), and many aspect of the NHDRs/NHQDRs, where entities involved with disparities issues find more and more outcomes respecting which to examine racial/ethnic differences while having no understanding of the problems of measuring those difference and being unaware even that different measures can yield opposite conclusions about directions of changes.

differences in the circumstances of the groups being compared and then whether it is worthwhile to identify the reasons for such change.

For purposes of considering the points made with regard to Table 5 of "Race and Mortality," it is useful to regard the pass and fail rates in the two rows of Table 1 as the favorable and adverse outcome rates resulting from the subjective judgments of two decisionmakers for any of a wide of subjects and one must examine the data to determine which of the decisionmakers is more biased or more likely to be biased. The favorable and adverse outcomes could involve any of the following subjects (some of which, for simplicity, I cast simply in terms of the outcome more commonly discussed while leaving the opposite outcome implied) or countless other subjects: decisions concerning pass or fail grades on a subjectively graded test; decisions concerning hiring, promotion, or termination in employment; decisions regarding approval or rejection of loan applications or regarding whether a loan recipient qualifies or fails to qualify for a particular rate; decisions regarding forbearing or not forbearing in taking actions against homeowners and renters who are behind in payments; decisions about discipling students or resorting to restraints in order to control agitated students; decisions to arrest or not arrest an apparent lawbreaker or to resort to a level of force in executing an arrest; decisions regarding recommending or not recommending a particular therapy for a patient: and anything else where an analysis of outcome rates might be used to appraise the likelihood that the decisionmaker was biased.

The relative differences in the favorable and adverse outcomes provide opposite conclusions about which row reflects the greater likelihood of bias or greater severity of bias. Yet, given that the forces causing favorable outcome rates to differ are exactly the same forces causing adverse outcome rates to differ, the contrasting conclusions based on the two measures cannot both be correct. In fact, there is no rational basis for relying on one measure rather than the other, and thus neither can usefully distinguish between the two decisionmaker with respect either to the severity or to the likelihood of bias. For example, it should be obvious that it makes no sense to say that one decisionmaker is more biased with respect to favorable outcomes like hiring job applicants or approving mortgage applications while the other is more biased with respect to adverse outcomes like rejecting job applicants or rejecting mortgage applications. Nor would it make sense to say that one decisionmaker is more biased in absolute terms while the other is more biased in relative terms (or, more precisely, in relative terms as to the outcome that yields an opposite conclusion from the absolute difference rather than the one that yields the same conclusion as the absolute difference).

Only one decisionmaker can be more biased or more likely to be biased than the other, though, as in the specific situation reflected in the two rows, they can be equally biased or equally likely to be biased. The goal of measurement is to identify the reality underlying the two rows of data with respect to the issue at hand and value judgments about choices of measure provide no assistance in the achieving that goal. Understanding the ways measures tend to be affected by the prevalence of an outcome, however, is essential to fulfilling that goal. While I am inclined to think the point is most evident in a situation when the issue is bias, the point applies as well to any effort to determine whether the forces causing the adverse and corresponding favorable outcome rates to differ are increasing or decreasing or are larger in one setting than another, regardless of the nature of the forces or the nature of the outcomes.

What can be inferred from the values in the two rows is that, with respect to the listed outcome combinations, decisionmakers who are more lenient and more circumspect with regard to imposing the adverse outcomes rather than the favorable outcomes or who are better skilled in techniques that avoid resort to imposition of the adverse outcome, or who simply are more responsive to instructions or pressures to limit the adverse outcomes, will tend to show larger relative differences in adverse outcomes but smaller relative differences in the corresponding favorable outcomes than other decisionmakers – that is, to show a pattern more like that in Row 2 than Row 1.[35]

Of course, in situations where there is a finite number of favorable outcomes, as in a hiring or promotion process, the two rows would simply reflect the more opportunities available in the second situation than the first situation.

The above points should not be read to suggest that the results in Row 2 invariably reflect a more desirable situation. It can simply reflect the results of pressure to lower standards when the higher standard is more desirable or reflect unjustified reluctance to impose the adverse outcomes. But it remains important to recognize that in the contexts where there is near universal reliance on relative differences in adverse outcome to identify decisionmakers whose actions should be scrutinized or sanctioned, it will the decisionmakers who are most compliant with the guidance to avoid resort to the adverse outcome or pressure or encouragements to relax standards or otherwise reduce adverse outcomes who are most likely to be singled out for such scrutiny or sanctions.

The two rows may also be regarded as the outcomes of decisions by decisionmaker where Row 1 involves decisions about sanction for more serious or objectively-identified offenses and Row 2 involves decisions about sanction for less serious of subjectively-identified offenses. It would thus reflect the situation where observers draw inferences about processes on the basis of the larger relative difference in adverse outcomes for Row 2 offenses, while unaware of or ignoring the smaller relative difference for the favorable outcome in that row, and while failing to recognize the extent to which such pattern is a function of greater caution in imposing sanctions for less serious or subjectively-identified offenses than for more serious and objectively-identified offenses. See my Offense Type Issues webpage.

The two rows may also be regarded as the situations of less qualified (Row 1) or more qualified (Row 2) job or loan applicants or home buyers, where observers will commonly draw inferences about processes on the basis the comparative size of whichever of the two relative differences they happen to be examining.

---

[35] See my "Getting it Straight When Statistics Can Lie," *Legal Times* (June 23, 1993), regarding the mistaken belief that adding procedural safeguards to limit arbitrary terminations will tend to reduce, rather than increase, relative racial differences in terminations. The article also discusses the mistaken belief that generally improving training such as to better enable employees to succeed at their jobs will tend to reduce, rather than increase, relative racial differences in terminations for failure to succeed in the jobs. It also addresses a court decision that was premised on the mistaken belief that higher performance standards tended to result in greater relative differences in failure to meet the standards than lower performance standards. None of these issues is any better understood today than it was in 1993, by courts or by any other entities.

All of these considerations apply as well when the outcome rates in the two rows reflect the actions of the same decisionmaker during different periods. But in no circumstance can either of the two relative differences provide a sound basis for such inferences just as neither can provide a sound basis for appraising the likelihood that one decisionmaker is more likely to be biased than another. The same issues apply to the absolute difference, even if it does not provide contradictory interpretations of patterns of data in the way the two relative differences do. The simple fact that the absolute difference tends to change as the prevalence of an outcome changes renders it an unsound measure of the strength of the forces causing the two outcome rates to differ.

All these points apply as well to efforts to determine whether there is more bias or racism in one jurisdiction than another or whether bias and racism are increasing or decreasing over time. Increasing numbers of studies in the United States and the United Kingdom observe that relative health inequalities (for adverse outcomes) have been increasing while absolute inequalities have been decreasing and then refer to value judgments or normative judgments in choosing a measure, as in the BMJ study that was the subject of my [May 9, 2016 response]. It ought to be clear that such studies provide nothing useful with respect to evaluating the effects of policies on differences in the circumstances of advantaged and disadvantaged group. But, it any case, it should be obvious that it make no sense to say that bias and racism are increasing in relative terms but decreasing in absolute terms or decreasing as to favorable outcomes but decreasing as to adverse outcomes or any similar such things about differences in bias or racism from place to place.

4. An Explanation of a Method of Quantifying a Demographic Difference Reflected by Favorable or Adverse Outcome Rates With Some Illustrations Based on Data Recently in the News and Recent Study of Changing Patterns of Disparities in Uninsurance Rates

"Race and Mortality Revisited" discusses a method of quantifying a demographic difference involving outcome rates that is unaffected by the prevalence of an outcome. The method involves deriving from a pair rates for a favorable or adverse outcome the difference between means of underlying normal risk distribution in terms of units of a standard deviation. I developed the method in 2008 terming the resulted "EES" for "estimated effect size." But I later came to understand that the method already existed with the results termed probit d'.

There are numerous illustrations of the method using outcome rates found in reality in the materials discussed in the Introduction, including in all table of "Race and Mortality Revisited" except Tables 1 and 5, as well as in many scores of webpages on jpscanlan.com. The values for Tables 1 and 5 of "Race and Mortality Revisited" would necessarily be .50 because hypothetical data are premised on a situation of normal distributions with means that differ by .50 standard deviations (though, due to the rounding of the figure in the two tables, those values would not yield probit d' values of exactly .50). As previously noted, the .50 standard deviation difference between mean scores underlying the test score hypothetical (which also underlies my graphical illustrations of the various patterns on place like page 31 of the CEP Comments and Figure 3 of University of Minnesota workshop), reflects a situation where about 31 percent of the lower-scoring group would score above the mean for that higher-scoring group. For additional

perspective, a probit d' value of .30 reflects a situation where about 38 percent of the risk distribution for the disadvantaged group is above the mean of the distribution for the advantaged group. With the .30 difference between means on the test, at the point where the 95 percent of the advantaged group passes the test, the pass rate for the disadvantaged group would be about 91 percent (which may be compared with the 87 percent figure shown in Table 1).

The method may be easily implemented with the downloadable ES-Calculator.xls file made available on a webpage of Professor David B. Wilson of George Mason University. On the Proportions tab, one merely enters the favorable or adverse outcome rates for the groups being compared and the probit d' value appears as the "d" value in the Probit Method box It does not matter whether one enter enters the favorable or adverse outcome rates or which group's rate is entered in each row, as each approach will yield the same value (though a negative sign will appear in front of the value when the lower rate is entered in the treatment group box).

In the case of the infant mortality rates between 1850 and 2016 and between 1916 and 2017 discussed earlier – and which show the usual pattern whereby major reductions in infant mortality over the two periods were accompanied by increases in relative racial differences in infant mortality rates but decreases in the relative racial differences in infant survival rates and decreases in absolute racial differences between rates – the method would produce the following results: the black and white infant mortality rates of 340 and 217 per thousand live births in 1850 and 11.4 and 4.9 per thousand live births in 2016 would yield values of .370 in 1850 and .306 in 2016; the black and white infant mortality rates of 184.9 and 99.0 per thousand in 1916 and 10.8 and 4.9 per thousand in 2017 would yield values of .387 in 1916 and .285 in 2017.[36]

A July 19 Urban Institute brief titled "Racial Disparities in Uninsurance among New Mothers Following the Affordable Care Act," by different researchers from those authoring the 2014 Urban Institute report discussed above, can illustrate several of the matters discussed above about research employing standards measures of differences and discussions about large relative differences in adverse outcomes, while also offering perspective on the utility of probit d'.

It first warrants note that the brief suggests it was prompted by recent attention both to increasing maternal mortality in the United States and to very large relative racial differences in maternal mortality (whether cast in terms of the relative difference between rates or the ratio the black rate to the white rate, which, as noted, is the relative difference plus 1, as in reference 7 and 8 to the brief). In May 2019 attention to both the overall increases and the size of the relative racial difference led Senator Kamala Harris to introduce the Maternal Care and Access and Reducing Emergencies Act to address both subjects. Possibly the subjects, like racial disparities in COVID-19 outcomes, will receive further attention in presidential and vice-presidential debates.

But an important thing to understand about racial difference in maternal outcomes is that relative racial differences between rates of dying in childbirth tend to be comparatively large, while relative racial differences in rates of surviving childbirth tend to comparatively small, because, even with recent increase, maternal mortality is an extremely rare outcome (as reflected by the

---

[36] In Professor Wilson's calculator the 340 per thousand rate black infant mortality rates would be entered as .340. Rates like the 42.8 per 100,000 black maternal mortality rate discussed infra would be entered as .000428.

fact that it is counted in terms of deaths per 100,000 while infant mortality is counted in terms of deaths per 1,000). Further, improvements in care that generally reduce maternal mortality tend to increase relative racial differences in maternal mortality (while reducing relative racial differences in rates of maternal survival).   On the other hand, increase in maternal mortality such as the 26 percent increase between 2000 and 2014 referenced in legislation proposed by Senator Harris commonly reduce relative racial differences in maternal mortality while increasing relative differences in maternal survival.[37] Discussions of epidemics or crises in black infant or maternal mortality are commonly prompted by the size of the relative racial differences rather than the actual rates, and thus references to epidemics or crises are likely to be more common when adverse outcomes among black, as among whites, are at low points rather than when they are at high points. See my discussion of perceptions about large or increasing relative differences in mortality in prior decades in my "The Perils of Provocative Statistics," *Public Interest* (Winter 1991), "Race and Mortality," *Societ*y (Jan./Feb. 2000), and "Understanding Racial Difference in Infant Mortality," *PrenatalEd Update* (October 2000).

Senator Harris's bill cited recent (apparently 2014) maternal mortality rates of 42.8 deaths per 100,000 live births for black women, compared to 13 deaths per 100,000 live births for white women, which would translate into a ratio of 3.3 (which means that the black maternal mortality rate is 230 percent greater than the white rate). The 3.3 black-white ratio is sometimes discussed in terms of an order of magnitude greater than the ratio of the black infant mortality rate to the white infant mortality rate, which is commonly somewhat above 2.0. The larger relative difference for maternal mortality than infant mortality is something that it would be remarkable not to find, given the far lower prevalence of maternal mortality than infant mortality. But relative differences in survival rates and absolute differences between mortality (and survival) rates are smaller in the case of maternal outcomes than infant outcomes. And the figures cited in the legislation yield a probit d' value of .318, which is fairly comparable to the aforementioned recent values for infant mortality.[38] Thus, whatever the forces causing black and white infant outcomes to differ and black and white maternal outcome to differ (and one would expect the

---

[37] I have not seen data on the rates between the two years. But such pattern in fact is shown in a 2017 Obstetrics & Gynecology study of racial/ethnic trends in maternal outcomes for 27 states and the District of Columbia during periods of increases in maternal mortality between 2008-09 and 2013-14.

[38] The values underlying the statement in the brief (at 2) that "[m]aternal mortality disproportionally affects non-Hispanic black women, who experienced 43.5 deaths per 100,000 live births from 2011 to 2013, a mortality ratio 3.4 times higher than white women," including the white rate of 12.7 per 100,000 for non-Hispanic white women found in the source, would yield a probit d' value of .329. Inasmuch as the black rate is higher and white rate is lower than the rates cited in Senator Harris's bill, the probit d' value (and every other measure of difference) would necessarily be larger than would be derived from the rates in the bill. I note that this is an instance where using the phrase "times higher" rather than times as high, where both groups' rates are not reported, could cause a precise reader to infer that the white rate was 9.7 per 100,000. See my Times Higher web page regarding the way the misusage predominates in the leading medical and scientific journals with the exception of the *New England Journal of Medicine*. I am not sure about all the reasons for the widespread use of times higher when one means times as high and the like formulations discussed on the Times Higher webpage. But in the discussion of disparities issues there may be a tendency to say times higher to stress the size of disparity, while those using the phrase fail to recognize that the usage in fact states that the relative difference being stressed is 100 percentage points higher than it actually is. While the usage at page 2 is borrowed from the peer-reviewed source, it is also employed three other places in the brief.

forces to involve related factors), there is little basis for maintaining that they are greater for one type of outcome than the other. In any case, any study of such issue can provide little of value, though potentially much that is misleading, if uninformed by an understanding of the statistical patterns just described.

In endeavoring to describe the effects of the Affordable Care Act on racial/ethnic differences in uninsurance, the 2019 Urban Institute brief begins by discussing that uninsurance rates fell for all groups between 2010 and 2017, and then shows figures for percentage reductions that reflect the usual pattern whereby the percentage reduction was greater for whites (whose initial uninsurance rate was lowest) than for Hispanics and blacks. The brief then notes that Hispanic-white disparity decreased by 26 percent between 2013 and 2017 and that black-white disparity decreased by 37 percent between 2013 and 2016. These figures are percentage reductions in the absolute difference between rates, which pattern also would typically occur when there is a general decline in an outcome where rates are in the ranges commonly observed for insurance in the United States.[39]

Thus, at the outset the article discusses the size of relative decrease in the adverse outcome (which are of the nature to increase relative differences between the rates of whites and each minority group), and then discusses decreases in absolute differences between rates that are the functions of the larger absolute reductions for minorities than whites, essentially things that could be as well said about the patterns in Table 1. Like many articles on subjects like this, in discussing disparities issues the brief mentions relative measures and absolute measures without recognition that the measures yield opposite conclusions about directions of change. And, like some of the works that suggested that ratios of suspensions rates of blacks to whites had decreased when in fact they had decreased, the statement in the conclusion that "Hispanic new mothers remained three and a half times [as likely as] white new mothers to be uninsured," suggests that ratio of the Hispanic uninsurance rate to the white uninsurance rates is down, rather than up, from the earlier years.

Table 2 presents the pertinent insurance and uninsurance figures in rates shown in the brief for Hispanics and whites for the years, 2010, 2013, and 2017, and for blacks and whites for the years 2010, 2013, and 2016, along with measures of difference employed in Table 1, as well as the probit d' value (termed EES as in my tables in the references). The favorable and adverse outcome rates for the advantaged and disadvantaged groups are arrayed in the same order as in Table 1.   Figures are presented for both Hispanic-white comparisons and black-white comparison with comparison indicated in the first column. The figures for 2013 and 2017 or 2013 and 2016 are highlighted because those are the figures on which the brief focuses in describing its conclusion that uninsurance disparities decreased based on the reduction in the absolute differences between rates (or larger percentage point reductions in uninsurance for Hispanics and blacks than for whites).

---

[39] In developing countries where insurance coverage may be very low (and the corresponding uninsurance rates very high) modest reductions in uninsurance coverage would tend to increase absolute differences.

Table 2 White and Hispanic insurance and uninsurance rates for 2010, 2013, and 2017, and white and black insurance and uninsurance rates for 2010, 2013, and 2016, with measures of difference

Comp	Year	Wh Ins Rt	Min Ins Rt	Wh Un Ins Rt	Min Un-- Ins Rt	W/M I Ratio	M/W UI UI Ratio	Abs Df	EES
H-W	2010	86.7%	59.8%	13.3%	40.2%	1.45	3.02	26.90	.864
H-W	2013	87.3%	63.8%	12.7%	36.2%	1.37	2.85	23.50	.788
H-W	2017	93.0%	75.6%	7.0%	24.4%	1.23	3.49	17.40	.782
B-W	2010	86.7%	79.6%	13.3%	20.4%	1.09	1.53	7.10	.284
B-W	2013	87.3%	80.6%	12.7%	19.4%	1.08	1.53	6.70	.277
B-W	2016	93.2%	89.0%	6.8%	11.0%	1.05	1.62	4.20	.264

The first thing warranting note in the table is that between 2010 and 2017 one observes the standard patterns I described above. But with respect to the Hispanic-white comparison, during this period of declines in the adverse outcome between 2010 and 2013, the relative difference in the adverse outcome decreased (along with the relative difference in the favorable outcome). From that one can infer that there in fact occurred a reduction in the forces causing Hispanic and white rates to differ, something that is also reflected in the reduction of the EES from .864 in 2010 to .788 in 2013. The failure of the relative difference between black and white adverse outcome rates to increase between 2010 and 2013 might support a like inference of a decrease in the forces causing black and white rates to differ, though probably a very small decrease. That inference is also supported by the decline in the EES from .284 to .277.

But over the 2013 to 2017 and 2013 to 2016 periods on which brief focuses to draw its conclusions about decreasing disparities we observe the usual patterns of changes in standard measures during a period of general declines in uninsurance. But the changes in EES values are negligible suggesting little occurred with regard to disparities that was not a function of the general change in the prevalence of uninsurance.

The brief does not cite the 2014 Urban Institute report that showed recognitions of the ways different measures can yield opposite conclusions and that the relative difference in the decreasing outcome tend to increase while the relative difference in the corresponding increasing outcome tend to decrease. Rather, as is the case with most discussions of demographic differences, the brief does not reflect a recognition that such things are possible or that they in fact occurred.

Many other examples seems to me to support the utility of probit d' especially in circumstances where, given advantaged and disadvantaged groups' rates at one point in time and either group's rate at another point in time, the method can predict closely the rate for the other group at the end of the period. The EES figures in Table 27 (slide 71) of the University of Minnesota workshop

38

reflect the changes in white, black, and Hispanic unemployment rates between 2007 and 2011. [40] The black and white EES figures mean that, given the group's rates in 2007, on the basis of the white rate in 2011, one could closely predict the black rate in the 2011 and thus little basis for saying either group was more harmed by the recession with respect to unemployment. One the other hand, the Hispanic EES figure suggests that Hispanic were particularly hard hit by the recession compared with blacks and white.[41] See also the EES values in Table 2 of "Race and Mortality Revisited" at three different levels of poverty. But there are scores of pages on jpscanlan.com that allow one to appraise the plausibility of probit d' (EES) quantifications of change in disparities as the prevalence of an outcome changes in circumstances where standard measures typically show the usual, and contrasting patterns, as well as when there are departures from those patterns.

In "Race and Mortality Revisited" (at 337), I discuss the potential weaknesses of the method since it relies on assumptions about the shapes of underlying distribution. [42]   But I also explain the reasons why an approach of that nature, though ideally better informed by the actual shapes of the underlying distributions, is vastly superior to the approaches commonly employed in the analysis of demographic difference.[43]

Probit d' has since been recommended for measurement of demographic differences in suspension rates in a 2019 *Educational Psychologist* article (Girvan et al., "Tail, Tusk, and Trunk: What Different Metrics Reveal About Racial Disproportionality in School Discipline") which refers to it as a robust method.[44] A similar approach is recommended in April 2020 article

---

[40] Probit d' can be used to quantify differences the changes in each group's rate over time in the same way that it can quantify the difference between each group's rate at two points in time. The difference between the probit d' values for each group's change will be the same as the difference between the probit values for the differences between the two groups' rates before the change and after the change. See on my Educational Disparities the way the measures can compare the changes in proficiency/nonproficiency rate for many groups at one time.

[41] See also the California Reading Instruction Competence Assessment webpage. The page shows how the various groups' pass rates after multiple opportunities to take a test are almost exactly what would be expected based on pass rates on the first taking of the test and any group's rate after multiple testing (while also showing that allowing multiple testing resulted in smaller relative differences in meeting the standard but larger relative difference in failure to meet the standard than would be the case if a person were allowed only one opportunity to take the test.

[42] See also the issues I raise at page 51 of "Can We Actually Measure Health Disparities?," *Chance* (Spring 2006).

[43] I do not regard the method employed in the Urban Institute 2014 report as an effective means of measuring a difference between two rates in a way that is unaffected by the prevalence of an outcome. For example, in a situation where there are 100 AG and 100 DG test takers, the measure described in the report would yield a value of .0345 for the first row and a value of .0195 for the second row, rather than the same values for each row.

[44] This study was the first study by education researchers to recognize that suspensions tend to increase relative differences in suspensions rates, while reducing absolute differences. But in discussing varied measures, the study mistakenly concluded that the measures are functional equivalents of one another. The mistaken conclusion was based on the authors' observing the pattern whereby the larger any measure of difference, the larger tends to be all other measure of difference. But changing the prevalence of an outcome tends to cause the various measures to change in different directions, though leaving the probit d' unchanged. Thus, in evaluating changes over time, and in comparing setting with different prevalence of an outcome, the measures are hardly equivalent; more important, none of the measures apart from probit d' provides a theoretically sound means of determining whether forces causing outcome rates to differ have increased or decreased.

in American Journal of Epidemiology by statisticians at the Centers for Disease Control and Prevention and National Center for Health Statistics who were authors or the paper to which my "The Mismeasure of Health Disparities" responded (Talih, et al. ("Measuring the magnitude of health inequality between two population subgroup proportions"). The approach can be fairly regarded as an attempt to inform the method employed in "Race and Mortality Revisited" by a more refined understanding of the shapes of the underlying distribution as I suggested should be done. Whether the Talih approach will prove to be in any way superior to probit d', or yield results that are at all different from probit d' – and whether its greater complexity will cause it to be ignored – remains to be seen. As explained in the April 28, 2020 BMJ response,[45] I do regard the paper as potentially an important step in the right direction by employees of agencies that so far have done a great disservice to the public in promoting or allowing the expenditure or vast resource on research that has produced very little of value but a great that is misleading.[46]

Resolving these issues was what I urged the American Statistical Association to do in the October 5, 2015 letter, as well as in the CEP Comments. Until some entity undertakes such a project (and performs it with a complete understanding of the failings in the analysis of demographic differences to date), I suggest probit d' provides the best available foundation for addressing the issues regarding measuring discrimination in housing or anything else.

B. The Particular Measurement Problems of Paired Tester Studies

Table 7 of "Race and Mortality Revisited" presents data from a tester study that sought to determine the comparative effects of having a criminal conviction on the hiring prospects of black applicants and white applicants. The study used tester pairs of two black and two white applicants where the members of the pairs were matched as to qualifications and manner of presenting themselves but would be distinguished by the fact that the resume for one member of each tester pair indicated a criminal conviction. The study was intended not to quantify differences between the treatment of similarly situated black and white testers, but to quantify for each racial group the difference in the treatment between members of the tester pairs whose resumes did and did not reflect a conviction. One useful feature of the study that is obviated concerns about whether members of tester pairs were in fact effectively matched by alternating from test to test the member of the tester pair who had a resume indicating a criminal conviction.

Results are set out in Table 3 below, which is similar to Table 7 of "Race and Mortality Revisited." The table treats the advantaged and disadvantaged groups as tester applicants

---

[45] Notwithstanding its potential importance, the paper can be criticized for obscuring the problems in standard measures of differences and for merely suggesting that its provide a way of avoiding making value judgments, when in fact value judgments have no place whatever in efforts to determine whether the forces causing rates of advantaged and disadvantaged groups to differ have increased or decreased and whether particular policies cause them to increase or decrease. Many readers may avoid the method because they will be willing to make the value judgments rather than deal with the complexities of the method while others may avoid the method because they think it is useful to include value judgments in measurement.

[46] See also the discussion of the work of Andrew Ho of Sean Reardon in the CEP Comments (at 38-39).

without a conviction indicated on their resumes and tester applicants with a criminal conviction indicated on their resumes. But, rather the leaving advantaged and disadvantaged groups' adverse outcome (no callback) rates to be inferred by the reader (1 – callback rate), the table includes the adverse outcome rates in the table in the same way in which the information is presented in Tables 1 and 2 above. In order to be consistent with the ordering of columns in Tables 1 and 2 above, the table presented favorable and adverse outcomes rates of the advantaged and disadvantaged groups in the same columns as in those tables. In addition, to be consistent with Table 1, the situation where the favorable outcome is less common and the adverse outcome is more common (as in the case of the black tester pairs) is put in the first row rather than as in Table 7 of "Race and Mortality Revisited."

Table 3. Callback and no callback rates of members of black and white tester pairs without and with a criminal conviction on their resumes, with measure of difference.

Race	(1) No Conv CB Rt	(2) Yes Conv CB Rate	(3) No Conv NCB Rt	(4) Yes Conv BCBB Rt	(5) No/Yes CB Ratio	(6) Yes/No NCB Ratio	(7) Abs Df	(8) EES
B	14%	5%	86%	95%	2.80	1.10	9 PP	.56
W	34%	17%	66%	83%	2.00	1.26	17 pp	.54

The study analyzed the results based on the proportions of cases where testers of each race with and without convictions on their resumes received callbacks from interviews. The study found that having a conviction caused a larger percentage reduction in callback rates for black tester applicants than white tester applicants and discussed implications of such pattern.

For reasons discussed in Section A, the finding is the same thing as a finding that having no conviction caused a larger percentage increase in callback rates for blacks than whites, and it is with respect to effects of having no conviction that the matter is presented in Table 3 (that is, with the advantaged group's rate as the numerator of the favorable outcome ratio and the disadvantaged group's rate as the numerator of the adverse outcome ratio).

In column 5 the table shows the percentage effect of having no conviction on blacks and whites, with the effect cast in terms of the ratio of the callback rate for the group with no conviction to the callback rate for the group with a conviction. The ratios of 2.80 among blacks and 2.00 among whites that having no conviction increased callback by 180 percent for blacks compared with an increase of 100 percent for whites.

But, for reasons explained above, given that blacks without a conviction had lower callback rates than whites without a conviction, one would expect a larger proportionate effect of having no conviction for blacks than whites. For the same reason, one would expect a larger proportionate effect on no callbacks for whites than blacks. Such was in fact that case. The ratios in Column 6 mean that having no conviction reduced no callback rates by 26 percent for whites compared with 10 percent for blacks.

Column 7 shows the absolute difference between the callback/no callback rates of tester applicant with and without convictions for both blacks and whites – 9 percentage points for blacks compared with 17 percentage points for whites. Those using this measure, as was done in

the Urban Institute 2019 study whose results are shown in Table 2, would find a larger effect of having (or not having) a conviction for whites than blacks whether evaluated from the perspective of effect on the favorable outcome or on the adverse outcome. Each of these patterns is of a nature to be expected given the generally lower callback rates of blacks and, with respect to the percentage point changes, the rate ranges at issue.

The probit d' values in the EES column – .54 for blacks and .56 for whites – suggest that to the extent we can measure the effects of having a criminal record on each group, the effect was approximately the same for blacks as for whites.

It is possible that discrimination had a role in the generally higher success rates for white than blacks. But since there was no matching of black testers with white tester for particular tests, there is not tester evidence for that.

There are nevertheless things to be learned about tester studies from the values in rates in the two rows and the measures of difference between the rates, if we simply view the rows of as reflecting favorable and adverse outcome rates of any advantaged and disadvantaged demographic group rather than persons without and with convictions. For one thing, the values in the table illustrate the problems with using standard measures to compare the degree of perceived discrimination in one locale with another locale, or with respect to one type of job compared with another, where one locale or one type of job may have more openings than another.

These same points, of course, could be made based on the normally distributed data underlying Table 1. But it is useful to appraise the significance of the patterns to paired tester studies with actual paired tester data.

The patterns in the table similarly illustrates the fact that decisions by tester designers to increase the fabricated qualifications of the members of the tester pairs or enhance the interview skills of those persons, or to perform the tests in contexts with greater opportunities, will tend to increase some measures of disparity while reducing other measures. Those decisions would tend to reduce the relative differences in the favorable outcome on which the study under discussion relied, though they would tend to increase the relative differences in the adverse outcomes on which other studies might rely. Further, when such choices or market conditions cause favorable outcome rates to be very high, and thus relative difference in favorable outcomes tend to be quite small, some observers may decide to rely on relative differences in adverse outcomes to measures disparities. Thus, the decisions that increase favorable outcomes, and reduce the corresponding adverse outcomes, will tend to increase the relative difference in new measure of disparity (i.e., the relative difference in the adverse outcome).

In note 18 of "Measuring Hiring Discrimination," *Labor Law Journal* (July, 1993), I discuss with regard to the somewhat different measurement issues involved with the 1991 Urban Institute tester study addressed in the article the way that decision to send testers to employers who had not advertised (assuming those employer had comparatively few openings) would tend to cause a study to find little discrimination. But in a study that examined favorable outcome rates of the two groups and measured disparities in terms of relative difference in favorable outcomes, as in

the study whose results are reflected in Table 3, a study conducted among employers that did not advertise would tend to find comparatively large disparities because fewer favorable outcomes tends to result in a larger relative differences in the favorable outcome, just as would be the situation in the case of a test with a very high cutoff.

Meanwhile, such decisions that generally increase favorable outcome rates, with contrasting effects on relative differences in the favorable outcome and relative differences in the corresponding adverse outcome, will for time increase absolute differences as the rates rise above those in the either row of Table 3. But when the favorable outcome rates reach a certain level, further general increases in the outcome rates will tend to cause absolute differences between rates to decrease.

I do not yet know enough about paired tester studies with respect to housing issues to know that favorable/adverse outcome rates commonly at issue. My impression is that in many contexts at least the tester from the advantaged group experiences what is deemed to be the favorable outcome the great majority of the time when the favorable outcome is merely being provided with appropriate assistance in securing desired housing. But the matter might be very different when, for example, the desired outcome is an offer of a rental unit that many people are seeking.

In any case, even apart from that the fact that decisions regarding the testing plan necessarily will affect the prevalence of favorable and corresponding outcome in ways that affect standard measures of difference between such rates, the fact that such measures tend to change because the prevalence on outcome change renders the measures unsound means of quantifying differences in treatment in the paired tester context just as it does in every other context.

Paired tester studies like that in the study underlying Table 3 that yield the overall favorable (or adverse) outcome rates for the groups being compared can, however, be soundly analyzed, just as an any pair of rates observed in reality can be soundly analyzed, albeit inexactly, as discussed in Section A.4 And it does not matter how decisions about the tester scheme affect the overall favorable and adverse outcome rates because the method is theoretically unaffected by the prevalence of an outcome.

But that is only so in the situation when the study produces information on the number of instances where decisionmakers made decisions about the members of the tester pairs. That would seem to include all the cases involved in the study underlying Table 3. For each member of the tester teams was interviewed and a decision was made to call the interviewee back or not to call the interviewee back.

The situation would be quite different, however, when part of the applicant universe receives no attention from decisionmakers. The point is illustrated in Table 4, which is based on the situation in Table 1, and where, apart from rounding issues, each row would reflect an EES of .50. All rows can be regarded as reflecting a situation where paired tester apply for some desired outcome and the pass and fail rates reflect the proportions of total tests in which the testers received or failed to receive the desired treatment. The first of each pair of rows contains the same information as in Rows 1 and 2 of Table 1 and each row (as well as the EES that would be derived from the rounded rates using the table rather than the .50 figure that underlies the

hypothetical). The second or each pairs of rows (1a and 2a) presents information on how the same situation would appear when the decisionmakers only examined half of the applications.

Table 4. Illustration of effects on indicators in Table 1 (and EES) of fact that half the persons in each group received no attention whatever from any decisionmaker

Row	(1) AG Pass Rate	(2) DG Pass Rate	(3) AG Fail Rate	(4) DG Fail Rate	(5) AG/DG Pass Ratio	(6) DG/AG Fail Ratio	(7) Abs Df	EES
1	80%	63%	20%	37%	1.27	1.85	17 PP	.510
1a	40%	31.5%	60%	68.5%	1.27	1.14	8.5 PP	.228
2	95%	87%	5%	13%	1.09	2.60	8 PP	.519
2a	47.5%	43.5%	52.5%	56.55%	1.09	1.08	4 PP	.101

One will observe that the ratio of AG's favorable outcome rates to DG's favorable outcome rate (Column (5)) is the same in the second of each pair of rows as in the first of each pair of rows. Some might mistakenly regard that as a useful feature of that ratio. It would be if the ratio (or the relative difference it represents) is a sound measure of association. But, putting the differing interpretations yielded by the other standard measures, the EES values vary dramatically. That is, the true situation is reflected in Rows 1 and 2, but the fact that only half the universe was actually considered by decisionmakers would lead to the mistaken perceptions that Rows 1a and 2a would support (including the mistaken perception that the forces causing outcome rates of the two groups to differ to would vary substantially between Rows 1a and 2a).[47]

The 1991 Urban Institute tester study of hiring discrimination in Washington, DC, and Chicago that I addressed "Measuring Hiring Discrimination" examined different types of rates from those examined in the study underlying Table 3. The Urban Institute study made comparisons of differences in proportions of total tests in which either of the members of the tester pair appeared to be favored over the other. My criticism of the methodology raised issue about identifying numerators and denominators when one does not know what proportion of situations where either tester appeared to be favored actually involved such a situation and implications of the fact that much of the time one did not know whether either tester received any consideration.

As reflected in note 23 of that article, I had already described the pattern by which the rarer an outcome the greater tends to be the relative difference in rates of experiencing the outcome and and the smaller tends to be the relative difference in rates of avoiding the outcomes a number of places. But I suggested the pertinence of that pattern to aspect of the study's finding in a way I would likely do differently today and that would in any case apply to different issues.[48] And the

---

[47] I discuss this existence of this complication with regard to most analyses of discrimination claim in "The Mismeasure of Discrimination," Faculty Workshop, University of Kansas School of Law (Sept. 20, 2013) (21-22) and on my Case Study II webpage.

[48] Part of the criticism of the study involved the fact that the more opportunities were available the more findings of perceived discrimination there would be, according to the methodology being employed. I believe I discussed the issue with respect to the study's finding that there was more discrimination in Washington, DC than Chicago in "Study Misses Mark on Discrimination in D.C.," *Washington Business Journal* (May 18-25, 1992), but I have not seen that article in many years.

suggestions of how one might make comparisons between rates once one identified an appropriate numerator and denominator to calculate each group's rate reflects approaches to comparing the difference between outcomes rates that I no longer believe are sound.

In any case, I think that, while the points made in the article must be considered carefully, it is improbable that examinations of instances of perceived favoritism will yield rates of the type that can be soundly analyzed by a method like probit d'. That is, I believe I know a method of measuring an effect based on rates at which two groups experience either the favorable or the corresponding adverse outcome that is theoretically sound regardless of the level of qualifications of different tester pairs and regardless of prevalence of favorable and corresponding adverse outcome in the settings where the tests are conducted. But I do not know a sound way of measuring differences between rates at which one member of tester pair is favored, disfavored, or treated the same as the other member of the tester pair. And persons who feel they know how to do that ought to reexamine their views view in light of the things discussed in Section A and points raised in the *Labor Law Journal* article.

The same holds for paired tests in housing that seek to quantify discrimination based on rates of perceived favoritism rather that rates at which members of the groups being compared experience an outcome.[49] But I believe the utility of a particular measure of any difference in any context can only be appraised after an examination of data on what happened to various tester pair combinations in particular situations, and with an understanding of the implications of the fact that it commonly may be impossible to know what part of the relevant universe actually received any attention from decisionmakers.

---

[49] The criticism of the study also included a questioning, based on patterns shown respecting rates for particular tester pairs, the study authors' confidence that members of tester pairs were appropriately matched.

CORRECTION

Corrections through September 21, 2020:[50]

Page 1, second paragraph, last line: "to which matters pertinent" change to "to which the matters are pertinent"

Page 2, point 4, line 2: "to the" change to "to that"

Page 4, last paragraph, line 7: "2018" change to "2020"

Page 5, first full paragraph, fourth last line: "opinion" changed to "statement"

Page 8, first full paragraph, fourth last line: "and letters" change to ". See also"

Page 8, note 10, second line, "Section A" change to "Section A.2"

Page 9, first full paragraph, eighth line: "reducing relative differences" change to "reducing relative racial/ethnic and gender differences"

Page 9, second full paragraph, lines 6 and 7: "3.89" changed to "4.32"

Page 10, second full paragraph, second last line" "bigger" changed to "larger"

Page 11, second full paragraph, fifth last line: "suggested it assume" changed to "suggested that the Commission assume"

Page 11, second full paragraph, second last line" "opinion to" changed to "dissent from"

Page 13, second full paragraph, fifth line: "A.3" changed to "A.4"

Page 15, last paragraph, seventh line: "below percent" changed to "below 50 percent"

Page 20, first line: "decreasing adverse outcomes tended to decrease" changed to "decreasing adverse outcomes tended to increase"

Page 25, second full paragraph, third line: 'as debates" changed to "as well as debate"

Page 32, paragraph from previous page, third last line" "particular for" changed to "particular therapy for"

Page 33, fourth full paragraph, last line: "looking at" changed to "examining"

Page 34, first full paragraph, first line, "determine to determine" changed to "efforts to determine"

Page 36,first full paragraph, fifth and fourth last line: "whatever the forces causing black and white infant outcomes and maternal outcome to differ" changed to "whatever the forces causing black and white infant outcomes to differ and black and white maternal outcome to differ"

Page 38, third paragraph after Table 2, second and third lines, "that relative difference in the decreasing outcome tend to decrease" changed to "that the relative difference in the decreasing outcome tend to increase"

Page 39, first paragraph, fourth last line: "plausibility of quantifications" changed to "plausibility of probit d' (EES) quantifications"

Page 39, second full paragraph, third line: "informed by the" changed to "informed as to"

Page 41, third paragraph after Table 3, third line: "with no conviction" changed to "with a conviction"

---

[50] Very minor corrections are not identified.