

Table of Contents

	Page
Statement of Interest	1
Introduction and Summary	4
A. Patterns by Which the Two Relative Differences and Related Disparity Indicators Tend To Be Affected by the Frequency of an Outcome	7
B. Implications of the Patterns by Which Measures Tend to Be Affected by the Frequency of an Outcome With Respect to Particular Requirements of the Consent Decree	24
1. The Fundamental Unsoundness of Analyses of Discrimination Issues Based on the Proportion a Group Makes Up of Persons Potentially Experiencing an Outcome and the Proportion the Group Makes Up of Persons Actually Experiencing the Outcome	24
2. The Common Inadequacy of Adjustment for Differences in Characteristics and the Implications of the Statistical Significance of Residual Disparities	27
3. The Drawing of Inferences on the Basis of the Comparative Size of Relative Differences	28
4. Inferring Bias on the Basis of the Magnitude of Disparities	32
5. Disparities in Favorable Outcomes	33
C. Suggested Disposition of this Matter	34
Table 1. Illustration of effects on relative differences in pass and fail rates, and proportions DG makes up of persons who pass and fail, of lowering a cutoff from a point where 80% of AG passes to a point where 95% of AG passes (when mean scores differ by approximately half a standard deviation and DG comprises 50% of test takers)	9
Figure 1. Ratios of (1) DG Fail Rate to AG Fail Rate and (2) AG Pass Rate to DG Pass Rate at Cut Points Defined by AG Fail Rates	11
Table 2. Illustration of effects on relative differences in poverty rates and rates of avoiding poverty, and proportions African Americans make up of the poor and the non-poor, of reducing poverty such as enable everyone living on incomes above 75% of the poverty line to escape poverty	12

Statement of Interest

I am a lawyer who has dealt extensively with the statistical analysis of discrimination issues while representing both the government and private parties. I represented the Equal Employment Opportunity Commission (EEOC) in the largest employment discrimination case ever fully tried (*EEOC v. Sears, Roebuck and Co.*, 504 F. Supp. 241 (N.D. Ill. 1986), *affirmed*, 839 F.2d 302 (7th Cir. 1988)), which was a case based almost entirely on statistics. At the time of leaving the EEOC in 1995, I was the agency's Assistant General Counsel for Expert Services, with responsibility for overseeing litigation in the agency's field offices involving statistics.

Since 1987, I have published over 30 articles about misunderstandings of statistics in the law and the social and medical sciences. I have also given over thirty presentations on this subject at statistics, demography, epidemiology, and public health conferences in North America and Europe and in methods workshops at American universities. My work on statistics has given particular attention to implication of the patterns (and the failure to understand the patterns) by which standard measures of differences between outcome rates tend to be systematically affected by the frequency of an outcome. These patterns include that whereby the rarer an outcome the greater tends to be the relative (percentage) difference in rates of experiencing it and the smaller tends to be the relative difference in rates of avoiding it. On the basis of my work, the pattern was recognized by statisticians at the National Center for Health Statistics in 2004 and it has since been recognized by scholars in the United States, Europe, and Asia. But the pattern remains unknown to the Department of Justice and other agencies enforcing federal civil rights laws, and policies of those agencies consistently reflect the mistaken belief that reducing the frequency of an outcome will tend to reduce, rather than increase, relative differences in rates of

experiencing the outcome as well as associated measures of disproportionality regarding the outcome.

Since 2012 I have given substantial attention to situations where, as the result of the government's mistaken belief about the statistical implications of reducing the frequency of adverse outcomes, it has encouraged entities covered by civil rights law to engage in conduct that increases the chances that the government will sue them for discrimination. The key issues regarding this matter are explained fairly succinctly in my "Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies,"¹ *Amstat News* (Dec. 2012), and more elaborately in my "Race and Mortality Revisited," *Society* (July/Aug. 2014), and my *amicus curiae* brief in *Texas Department of Housing and Community Affairs et al. v. The Inclusive Communities Project, Inc.*, Sup. Ct. No. 13-1371 (Nov. 2014) (TDHCA brief).

I took a special interest in the Department of Justice's March 4, 2015 Investigation of the Ferguson Police Department (Ferguson Report) because of the report's mistaken premise that unwarrantedly aggressive policing practices and unjustifiably harsh court practices caused African Americans to make up a higher proportion of persons experiencing adverse outcomes at the hands of the police and the courts than would be the case with less aggressive policing practices and less harsh court practices. I brought the fact that the premise was mistaken to the attention of the Department of Justice and the City of Ferguson by letter dated March 9, 2015.

The mistaken premise of the Ferguson Report played a significant role in prompting me to write letters to the Chief Data Scientist of the White House Office of Science and Technology

¹ Underlinings in print copies of this letter reflect links to referenced materials in electronic copies of the letters that may be found by means of the Institutional Correspondence subpage of the Measuring Health Disparities page of jpscanlan.com. I do not include links with every mention of an item, but attempt to include them often enough to make online items readily available without the reader's having to search for an earlier link. Since the online copy of this letter may be corrected or annotated, I include here a link to the most recent version (which, if corrected or annotated, will provide a link to a copy of the letter in the form filed with the Court).

Policy (Sept. 8, 2015) and the American Statistical Association (Oct. 8, 2015) urging the recipients, among other things, to explain to arms of the federal government that reducing the frequency of an outcome tends to increase, not decrease, relative differences in rates of experiencing the outcome and the proportions disadvantaged groups make up of persons experiencing the outcome. I have also written to the City of Minneapolis, Minnesota (June 8, 2015), City of Boulder, Colorado (Mar. 5, 2016), and City of Madison, Wisconsin (Mar. 12, 2016) with respect to perceptions about their policing practices that have, or may, reflect the mistaken understanding of the statistical relationship between the frequency of an outcome and measures of racial disparities found in the Department of Justice's Ferguson Report. See also my letter to the House Judiciary Committee (Oct. 19, 2015) discussing the Ferguson Report among several types of actions of the Department of Justice and other agencies where an entity's compliance with the government's express or implied guidance increases the chance that the entity will be sued for discrimination.

The filing of the instant action, with a complaint reflecting the same mistaken premise as that underlying the Ferguson Report, prompted me to write "Things DoJ doesn't know about racial disparities in Ferguson," *The Hill* (Feb. 22, 2016), which explains the mistaken premise fairly succinctly. Also, in recent letters to the Council of Economic Advisers (Mar. 16, 2016), Population Association of America and Association of Population Centers (Mar. 29, 2016), and Consortium of Social Science Associations (Apr. 6, 2016) urging the recipients to take actions similar to that sought in the letters to the Chief Data Scientist and the American Statistical Association, I have cited the filing of the instant suit as a development giving particular urgency to the request that the recipients explain to the Department of Justice and other federal agencies that their understanding of the statistical effects of generally reducing adverse outcomes is

incorrect. I also brought the filing of the suit to the attention of the American Statistical Association committee that is considering the recommendations of my October 8, 2015 letter to that organization.

When sending the above letters, I was not aware of the parties' motion of March 16, 2016, requesting entry of the Consent Decree or the Court's Order of March 26, 2016, scheduling a hearing for April 19, 2016, and requesting comments by April 12, 2016. I became aware of the order when checking the status of the case on PACER on April 7, 2016. Thus, this submission is somewhat hurriedly drafted and may suffer from such fact. I note in particular that the Consent Decree's Paragraph 74 (discussed *infra* at 17-18) and the subject of the last two paragraphs of Section B.1 of this submission (*infra* at 26) are matters of greater importance than their placement or the attention they are given might suggest. Nevertheless, I believe this submission should facilitate the Court's consideration of the issues before it.

It is my hope that this submission will cause the resolution of this matter to be based on a sound understanding of pertinent statistical principles by the Court and both parties and that actions taken regarding this matter may cause the Department of Justice to reform its practices with regard all matters involving the statistical issues addressed here.

Introduction and Summary

For more than twenty years the Department of Justice and other agencies enforcing federal fair lending laws have encouraged lenders to relax lending standards and otherwise reduce the frequency of adverse borrower outcomes in order to reduce relative (percentage) racial/ethnic differences in rates of experiencing those outcomes. For at least several years, the Departments of Justice and Education have encouraged public schools to relax discipline

standards and otherwise reduce the frequency of adverse school discipline outcomes in order to reduce relative racial and other demographic differences in rates of experiencing those outcomes.

The belief that reducing the frequency of such outcomes will reduce relative differences in experiencing them, however, is the exact opposite of reality. Reducing the frequency of an adverse outcome will tend to reduce relative differences in rates of experiencing the corresponding favorable outcome. But reducing the frequency of an adverse outcome will tend to increase relative differences in rates of experiencing the outcome itself.² Unaware that reducing the frequency of an outcome tends to increase relative differences in rates of experiencing it – indeed, believing just the opposite – the government has continued to monitor the fairness of lending and school discipline practices on the basis of relative differences in adverse outcomes. Thus, lenders and schools that reduce the frequency of adverse outcome in compliance with government encouragements increase the chances that the government will accuse them of discrimination.

A corollary to the pattern of relative differences just described is a pattern whereby reducing the frequency of an outcome tends to increase the proportions groups most susceptible to the outcome make up of persons (a) experiencing the outcome and (b) failing to experience the outcome. Here, too, as reflected by the Department of Justice's actions regarding Ferguson, Missouri, the government has consistently acted on the mistaken belief that reducing the frequency of an adverse outcome will tend to reduce the proportions groups most susceptible to the outcome make up of persons experiencing the outcome. Thus, here, too, the government has

² The factor at issue here is the direction of the change in the frequency of an outcome and the corresponding change in the opposite direction of the opposite outcome. That is, reducing the frequency of any outcome tends to reduce relative differences in the corresponding adverse outcome while increasing relative differences in the outcome itself. I refer to the adverse and favorable outcome here simply to facilitate reference to the two outcomes.

been encouraging entities covered by civil rights law to take actions that increase the chances that the government will accuse them of discrimination.

As discussed earlier, a succinct treatment of these issues with a particular focus on the instant litigation may be found in my “Things DoJ doesn’t know about racial disparities in Ferguson,” *The Hill* (Feb. 22, 2016).

As will be shown below, that reducing the frequency of an outcome tends to increase, not decrease, relative difference in rates of experiencing the outcome and the proportions more susceptible groups make up of persons experiencing the experiencing, while not widely known, is hardly debatable. Confronted with this submission and required to address the points it makes, the Department of Justice should acknowledge that it has been mistaken on the matter.

But whether the Department acknowledges its mistaken understanding or not, the Consent Decree in its current form – with its various provisions requiring the City of Ferguson to monitor racial differences and its various provisions aimed at generally reducing adverse outcomes, all in a context where the government itself does not understand the statistical consequences of reducing the frequency of an outcome – would place the City in an untenable situation. Thus, approval should be deferred until various statistical issues are addressed and resolved.

Section A of this submission briefly describes the above-mentioned statistical issues, while referencing materials that address the matters in greater detail. Section B addresses certain statistical issues having particular bearing on the City of Ferguson’s obligations under the decree that, while implicit in the discussion in Section A and explicitly addressed in many of its references, may not be evident to all readers. Section C suggests a framework for resolving this matter in light of the issues explained in Sections A and B.

A. Patterns by Which the Two Relative Differences and Related Measures of Disparity Tend To Be Affected by the Frequency of an Outcome

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

The above measures most pertinent to disparate impact and other discrimination issues are the two relative differences and I will limit the discussion here to those measures (though many of the references discussed below treat the other measures at length). The pertinent statistical pattern with respect to the two relative differences is that whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative differences in avoiding it.³

In some settings, including that involving the Department of Justice's action regarding police and court practices of Ferguson, Missouri, appraisal of demographic differences for

³ A more precise description of the pattern would state, rather than "the rarer an outcome," "the more the outcome is restricted toward either end of the overall distribution." But I have characterized the pattern in the manner done in the text above for some time and those discussing it have not been confused by the usage. Thus, I am not at this time inclined to depart from the usage in the text. Similar considerations apply to my use of the term "frequency" used here or the term "prevalence" that I have used in other places. See Section A.8 of the Scanlan's Rule page of jpscanlan.com.

purposes of analyzing discrimination or other issues are based, not on a comparison of outcome rates for different groups, but on a comparison of the proportion a group makes up of persons potentially experiencing an outcome and the proportion the group makes up of persons actually experiencing the outcome. Such comparisons are affected by the frequency of an outcome in a manner related to the above-described pattern regarding relative differences. Specifically, a corollary to the pattern whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it is a pattern whereby the rarer an outcome the greater tend to be the proportions groups most susceptible to the outcome make up of both persons experiencing the outcome and persons avoiding the outcome.⁴

These patterns can be easily illustrated with normally distributed test score data. Table 1 below, which is the same as Table 1 of the letter to the American Statistical Association mentioned in the Introduction and Summary and which appears in some form in each of the recent letters urging the recipients to explain this issue to the Department of Justice and other federal government entities, is based on a situation where the means of normal test score distributions of an advantaged group (AG) and a disadvantaged group (DG) differ by half a standard deviation and both distributions have the same standard deviation. In addition to showing the pass and fail rates of each group at two cutoff points, the table shows the ratio of AG's pass rate to DG's pass rate and the ratio of DG's fail rate to AG's fail rate at each cutoff

⁴ The pattern whereby reducing the frequency of an adverse outcome tends to increase the proportions disadvantaged groups make up of persons experiencing it and failing to experience it could just as well be termed the cause of the pattern of relative differences. For it is the fact that the disadvantaged group tends to make up a larger proportion of persons both below and above each increasingly lower cut point on a continuum that underlies the described pattern of relative differences. See Table 1 of the 2006 *Chance* editorial discussed *infra*. My earliest treatments of this subject principally focused on perceptions about disproportionate representations of certain groups among persons experiencing an adverse outcome. See "The 'Feminization of Poverty' is Misunderstood," *Plain Dealer* (Nov 11, 1987); "An Issue of Numbers," *National Law Journal* (Mar. 5, 1990); and "The Perils of Provocative Statistics," *Public Interest* (Winter 1991).

(the third and fourth last columns).⁵ Based on a situation where AG and DG each make up half of the test takers, the final two columns show the proportion DG makes up of persons who pass and persons who fail at each cutoff.

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

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

According to the specifications underlying the table, at the cutoff where 80% of AG passes the test, approximately 63% of DG would pass the test. Thus, the ratio of AG's pass rate to DG's pass rate would be 1.27. When the cutoff is lowered to the point where the pass rate for AG is 95%, the pass rate for DG would be approximately 87%. Thus, lowering the cutoff reduces the ratio of AG's pass rate to DG's pass rate from 1.27 to 1.08.

The fact that lowering a test cutoff tends to reduce relative differences in pass rates is widely known and underlies the common understanding that lowering a test cutoff reduces the disparate impact of employment and other tests where some demographic groups outperform

⁵ While I commonly refer to patterns of relative differences in this submission, the tables I use actually present rate ratios. The relative difference is the rate ratio minus 1 where the rate ratio is above 1 (in which case the larger the rate ratio the larger the relative difference) and 1 minus the rate ratio where the rate ratio is below 1 (in which case the smaller the rate ratio the larger the relative difference). One should be careful not to mistakenly refer to the rate ratio (which can also be termed "risk ratio" or "relative risk") as the relative difference. But the distinction between the rate ratio and the relative difference is not pertinent to the discussion here of patterns by which relative differences tend to be affected by the frequency of an outcome, save with respect to the fact that one may not make the same points about patterns of changes in rate ratios that one may make about patterns of changes in relative differences. In recent years I commonly present the rate ratios for both outcomes with the larger figure in the numerator. Thus, as to both outcomes, the larger the rate ratio, the larger the relative difference. Choice of numerator in the ratio, however, has no bearing on the described patterns by which the two relative differences tend to be affected by the frequency of the outcome.

others.⁶ Such fact may also underlie the government's longstanding (though mistaken) belief that relaxing standards will tend to reduce relative differences in adverse outcomes.

But, while lowering a cutoff tends to reduce relative differences in pass rates, it tends to increase relative differences in failure rates. With the higher cutoff, DG's failure rate was 1.85 times AG's failure rate (37%/20%). Lowering the cutoff increases that ratio to 2.60 (13%/5%).

That is, lowering the cutoff and thereby generally increasing pass rates and generally reducing failure rates, decreased the relative difference in the outcome that increased in frequency while increasing the relative difference in the outcome that decreased in frequency.

The final two columns of the table show how lowering the cutoff increases the proportions DG makes up of persons passing the test and persons failing the test. Assuming, DG makes up half of the test takers, the former proportion would increase from 44% to 48% and the latter proportion would increase from 65% to 72%.

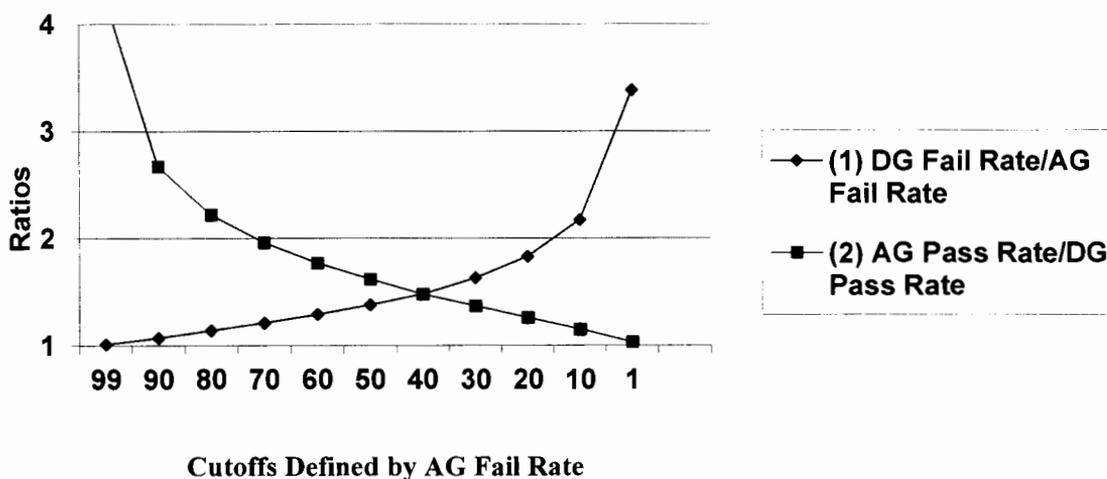
The pattern of directions of changes in these proportions is unaffected by the actual proportions DG makes up of persons taking the test. As discussed in the March 2016 *Hill* article mentioned above, if the proportion DG made up of persons taking the test was 67% (corresponding to the proportion African Americans make up of the population of Ferguson), lowering the cutoff would cause the proportion DG makes up of persons failing the test to increase from 79% to 84%. Though not pointed out in the article, when DG makes up 67% of test takers, lowering the cutoff would cause the proportion DG makes up of persons passing the test to increase from 62% to 65%.⁷

⁶ Whether lowering (or raising) a standard in fact increases or decreases the impact of a standard is a complex issue and the answers may vary from setting to setting. See Section E (at 27-32) of the Kansas Law paper discussed *infra*.

⁷ The proportion the disadvantaged groups make up of persons potentially experiencing an outcome or its opposite is, however, pertinent to the issues addressed in Section B.1 *infra*.

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

Figure 1. Ratios of (1) DG Fail Rate to AG Fail Rate and (2) AG Pass Rate to DG Pass Rate at Cutoff Points Defined by AG Fail Rates



A graphical illustration of the pattern of changes in the proportion DG makes up of person who pass and persons who fail across the full range of cutoffs may be found in Figure 1a (at 18) of the applied statistics workshop at Harvard’s Institute for Quantitative Social Science identified in note 10 *infra*.

Another useful illustration of the above-described patterns may be found in income data, as in Table 2 below (to which I will make further references in Section B.2 *infra*). The table is a portion of Table 1 of my guest editorial in the Spring 2006 issue of the American Statistical Association magazine *Chance* titled “Can We Actually Measure Health Disparities?” The table

presents the percentages of whites and African Americans living on incomes below and above 100% and 75% of the poverty line in 2004, along with the same quantitative indicators as in Table 1 of this submission.⁸

Table 2. Illustration of effects on relative differences in poverty rates and rates of avoiding poverty, and proportions African Americans make up of the poor and the non-poor, of reducing poverty such as enable everyone living on incomes above 75% of the poverty line to escape poverty

Perc of Poverty Line	White Perc Below	AA Perc Below	White Perc Above	AA Perc Above	AA/W Below Ratio	W/AA Above Ratio	AA Prop of Persons Below	AA Prop of Persons Above
100%	10.8%	24.7%	89.2%	75.3%	2.29	1.18	26.2%	11.6%
75%	7.2%	17.8%	92.8%	82.2%	2.48	1.13	28.0%	12.1%

Moving from the first row to the second row, we can observe the consequences of a reduction in poverty whereby everyone living on an income above 75% of the poverty line is enabled to escape poverty. And we see that with the substantial decrease in poverty, the ratio of the African American poverty rate to the white poverty rate would increase from 2.29 to 2.45, while the ratio of the white rate of avoiding poverty to the African American rate of avoiding poverty would decrease from 1.18 to 1.13. At the same time, the proportion African Americans make up of the combined African American and white poor population would increase from 26.8% to 28.0%, while the proportion African Americans make up of the combined African American and white non-poor population would increase from 11.6% to 12.1%.

⁸ Table 1 in the *Chance* editorial used the ratio of the African American rate of falling above an income level to the white rate of falling above the income level as an indicator of the relative difference in the favorable outcome, whereas here I use the ratio of the white rate of falling above the level to the African American rate of falling above the level. Among the reasons for the latter approach (see also note 5 *supra*) is that it facilitates the graphical illustration of the pattern by which the two relative differences tend to change in the opposite direction as the frequency of the outcome changes. See note 9 *infra*.

Graphical illustrations of the pattern of relative differences with respect to falling below and above the full range of income levels may be found in the top part of Figure 6 (at 35) of the Harvard workshop (which also includes the odds ratio).⁹

For the same reason that reducing the frequency of an outcome tends to increase relative differences in rates of experiencing it while reducing relative differences in rates of avoiding it, relative differences in adverse outcomes will tend to be comparatively large, while relative differences in the corresponding favorable outcomes will tend to be comparatively small, in areas or among populations and subpopulations where the adverse outcomes are comparatively uncommon than in areas or among populations and subpopulations where the adverse outcomes are comparatively common. The same holds for the other measures of disproportionality discussed above. That is, in areas where the adverse outcome are comparatively uncommon, disadvantaged groups will tend to make up a larger proportion of persons experiencing those outcomes and avoiding those outcomes, than in areas where the adverse outcomes are comparatively common.

The statistical patterns discussed in this submission are addressed at much greater length in the above-mentioned October 8, 2015 letter to the American Statistical Association. Other recent extended treatments of the issues may be found in my “Race and Mortality Revisited,” *Society* (July/Aug. 2014); “The Perverse Enforcement of Fair Lending Laws,” *Mortgage Banking* (May 2014); “Measuring Health and Healthcare Disparities,” Federal Committee on Statistical Methodology 2013 Research Conference (March 2014) (FCSM paper); “The Mismeasure of Discrimination,” Faculty Workshop, University of Kansas School of Law (Sept.

⁹ Graphical illustrations of the pattern of relative differences in falling above and below the full range of income levels may also be found in Figures 2 and 3 of the *Chance* editorial itself. But because the African American rate is used as the numerator for the rate ratio of falling above the line in the editorial’s Figure 3, the fact that the two relative differences change in opposite directions may be less evident.

2013) (Kansas Law paper); and the TDHCA brief mentioned in the Statement of Interest. See also my forthcoming “The Mismeasure of Health Disparities,” *Journal of Public Health Management and Practice* (July/Aug. 2016) (available online in May 2016).

The above items, and most other materials I have created since December 2007, include descriptions of a method for quantifying differences in the circumstances of two groups reflected by their rates of experiencing some outcome or its opposite that is unaffected by the frequency of an outcome.

Earlier extended treatments of the problems with standard measure of differences between outcome rates (with many examples of fundamentally unsound analyses in the law and the social and medical sciences), but without suggesting a solution to those problems beyond understanding the effects of the frequency of an outcome with respect to various measures, may be found in the 2006 *Chance* editorial mentioned above, as well as my “The Misinterpretation of Health Inequalities in the United Kingdom,” British Society for Populations Studies Conference 2006; “Race and Mortality,” *Society* (Jan./Feb. 2000); “Divining Difference,” *Chance* (Fall 1994); “Comment on “McLanahan, Sorensen, and Watson's 'Sex Differences in Poverty, 1950-1980'”,” *Signs* (Winter 1991); “The Perils of Provocative Statistics,” *Public Interest* (Winter 1991); and “The ‘Feminization of Poverty’ is Misunderstood,” *Plain Dealer* (Nov. 11, 1987).

Extensive graphical and tabular illustrations of the pertinent patterns may be found in methods workshops given at arms of American universities between 2012 and 2015,¹⁰ as well as

¹⁰ See “The Mismeasure of Health Disparities in Massachusetts and Less Affluent Places,” Department of Quantitative Health Sciences, University of Massachusetts Medical School (Nov. 18, 2015) “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

in more than a score of presentations given at statistics, epidemiology, demography, and public health conferences in North America and Europe between 2001 and 2011 (available [here](#)).

Treatments of the patterns with respect to particular outcomes, usually with tabular illustrations, may also be found on the pages and subpages of [jpscanlan.com](#) devoted to measurement issues.¹¹

Over 140 online comments (available [here](#)) address the problems in various studies in medical, epidemiological, and health policy journals that attempted to measure aspects of health or healthcare disparities – or to provide guidance on the interpretation of subgroup effects or the calculation of number-needed-to-treat – without consideration of the patterns by which measures tend to be affected by the frequency of an outcome.

Recent, fairly succinct treatments of the issues in the context of the government’s encouraging lenders and public schools to relax lending and discipline standards under the mistaken belief that doing so will tend to reduce relative racial differences in adverse borrower and discipline outcomes, as well as the Department of Justice’s actions regarding Ferguson, Missouri, may found in [“Things DoJ doesn’t know about racial disparities in Ferguson,”](#) *The Hill* (Feb. 22, 2016) (discussed above); [“Things government doesn’t know about racial disparities,”](#) *The Hill* (Jan. 28, 2014); [“The Paradox of Lowering Standards,”](#) *Baltimore Sun* (Aug. 5, 2013); [“Misunderstanding of Statistics Leads to Misguided Law Enforcement Policies,”](#) *Amstat News* (Dec. 2012); [“Disparate Impact’: Regulators Need a Lesson in Statistics,”](#) *American Banker* (June 5, 2012); [“The Lending Industry’s Conundrum,”](#) *National Law Journal* (Apr. 2, 2012).

[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).

¹¹ The principal measurement pages are: [Measuring Health Disparities](#), [Scanlan’s Rule](#), [Mortality and Survival](#), [Statistical Reasoning](#), [Immunization Disparities](#), [Immunization Disparities](#), [Educational Disparities](#), [Disparate Impact](#), [Discipline Disparities](#), [Lending Disparities](#), [Employment Discrimination](#), [Feminization of Poverty](#). The pages have close to 100 subpages.

Older, somewhat longer treatments of the misunderstanding of the effects of relaxing standards on measures of disproportionality, which involve a variety of settings and illustrate both how longstanding and how universal is that misunderstanding, may be found in my “Mired in Numbers,” *Legal Times* (Oct. 12, 1996); “When Statistics Lie,” *Legal Times* (Jan. 1, 1996); “Getting it Straight When Statistics Can Lie,” *Legal Times* (June 23, 1993); “Bias Data Can Make the Good Look Bad,” *American Banker* (Apr. 27, 1992); and “An Issue of Numbers,” *National Law Journal* (Mar. 5, 1990).

The last mentioned item discusses the fact that, at the time of the article, it was believed that the high African American representation among persons disqualified from competition in intercollegiate athletics by National Collegiate Athletic Association academic eligibility requirements was a function of the fact that standards were set too high, when in fact lower standards would have resulted in an even higher African American representation. That misunderstanding is as pervasive today as it was a quarter century ago, as reflected by, among other things, the Department of Justice’s Ferguson Report and the instant suit.

Recent, fairly succinct treatments of the misunderstanding of the implications of reducing the frequency of an adverse outcome on relative differences in experiencing the outcome that discuss the potential unconstitutionality of disparate impact claims under the Fair Housing Act may be found in my “Is HUD’s Disparate Impact Rule Unconstitutionally Vague?,” *American Banker* (Nov. 10, 2014) and “Case may reveal government’s perverse fair lending enforcement,” *The Hill* (Dec. 29, 2014), an issue also treated in the TDHCA brief (at 31-32). A key point in the analysis of that issue involves the fact that the Department of Housing and Urban Development’s regulations require that, even when practices with a disparate impact can be justified by sound business purposes, the covered entity must adopt the least discriminatory alternative that equally

serves those purposes (which commonly, in the view of the government and others, involves relaxing a standard). But covered entities are provided no guidance on whether appraisals of the comparative size of a disparate impact will be based on the relative difference in the favorable outcome or the relative differences in the adverse outcome (though probably they will be judged on the measure that relaxing a standard tends to increase).

The same points would apply to any disparate claim in circumstances where it is unknown whether reducing the frequency of an outcome will be regarded as reducing the disparate impact of the outcome on the basis of the relative difference in rates of avoiding the outcome or as increasing the disparate impact on the basis of the relative difference in experiencing the outcome. The Civil Rights Act of 1991 specifically requires that employers adopt the practices with the least disparate impact, which in the employment context is sometimes measured in terms of the relative difference in the favorable outcome and sometimes measured in terms of the relative difference in the adverse outcome. There was no Congressional intent with respect to which relative difference should be used to measure impact because then, as now, Congress was unaware that it was even possible for the two relative differences to change in opposite direction as a standard is altered, much less that this tends to occur systematically.

A particularly problematic feature of the proposed Consent Decree in this case may be found in its Paragraph 74 (at 18-19), which specifically requires that the Ferguson Police Department identify and implement any available alternative to a “program, initiative, activity, or service” having a disparate impact that has “less of a disparate impact.” All other issues aside, no Consent Decree should be entered unless it specifies how it will be determined whether one practice has a less of a disparate impact than another. See also the discussion in Section B.5

infra of the fact that the decree requires monitoring racial differences regarding some practices that are cast in terms of favorable outcomes and some that are cast in terms of adverse outcome.

The Consent Decree's Paragraph 285 (at 65) also requires that the Ferguson Police Department employ hiring practices with the least disparate impact. It is whether the requirement imposes any obligation beyond that imposed by Title VII of the Civil Rights Act. But if such a provision is to be incorporated into a decree, the decree should clarify the measurement issue.

Treatments of the above-described and related patterns may also be found in letters to governmental and nongovernmental entities addressing the ways that analyses of demographic or other differences in outcome rates that the entities conduct, fund, or provide guidance on – or that in some manner pertain to the entities' activities – are undermined by the failure to recognize patterns by which measures of differences between outcome rates tend to be affected by the frequency of an outcome. In addition to illustrating the patterns (and failures to understand them) in particular settings and with regard to particular activities, the letters reflect the pervasiveness of the misunderstanding of the patterns even among the nation's most prestigious institutions.¹² The misunderstanding is just as pervasive among like institutions in all nations.

¹² Such letters include the following (some of which have already been mentioned): Council of Economic Advisers (Mar. 16, 2016), City of Madison, Wisconsin (Mar. 12, 2016), Stanford Center on Poverty and Inequality (Mar. 8, 2016), City of Boulder, Colorado (Mar. 5, 2016), Houston Independent School District (Jan. 5, 2016), Boston Lawyers' Committee for Civil Rights and Economic Justice (Nov. 12, 2015), House Judiciary Committee (Oct. 19, 2015), Chief Data Scientist of White House OSTP (Sept. 8, 2015), McKinney, Texas Independent School District (Aug. 31, 2015), Department of Health and Human Services and Department of Education (Aug. 24, 2015), Agency for Healthcare Research and Quality (July 1, 2015), City of Minneapolis, Minnesota (June 8, 2015), Texas Appleseed (Apr. 7, 2015), Senate Committee on Health, Education, Labor and Pensions (Mar. 20, 2015), United States Department of Justice and City of Ferguson, Missouri (Mar. 9, 2015), Vermont Senate Committee on Education (Feb. 26, 2015), Portland, Oregon Board of Education (Feb. 25, 2015), Wisconsin Council on Families and Children's Race to Equity Project (Dec. 23, 2014), Financial Markets and Community Investment Program, Government Accountability Office (Sept. 9, 2014), Education Law Center (Aug. 14, 2014), IDEA Data Center (Aug. 11, 2014), Institute of Medicine II (May 28, 2014), Annie E. Casey Foundation (May 13, 2014), Education Trust (April 30, 2014), Investigations and Oversight Subcommittee of House Finance Committee (Dec. 4, 2013), Mailman School of Public Health of Columbia University (May 24, 2013), Senate Committee on Health, Education, Labor and Pensions (Apr. 1, 2013), Federal Reserve Board (March 4, 2013), Harvard University et al. (Oct. 26,

Recognition of the above-described pattern of relative differences by statisticians of the National Center for Health Statistics (NCHS) may be found in (a) Keppel KG, Percy JN. Healthy People 2010: Measuring Disparities in Health. *Chance* 2009;22(1):6-7; (b) Keppel KG, Pamuk E, Lynch J, *et al.* Methodological issues in measuring health disparities. *Vital Health Stat* 2005;2 (141); (c) Keppel KG, Percy JN. Measuring relative disparities in terms of adverse events. *J Public Health Manag Pract* 2005;11(6):479–483.; (d) Keppel KG, Percy JN, Klein RJ. Measuring progress in Healthy People 2010. Healthy People Statistical Notes. No. 25. Hyattsville, MD: National Center for Health Statistics. 2004. See also recent recognition that relative differences in experiencing an outcome tend to be large where the outcome is uncommon by scientists from the Centers for Disease Control and Prevention (apart from NCHS) in Moonesinghe R, Beckles GLA. Measuring health disparities: a comparison of absolute and relative disparities. *PeerJ.* 2015;3:e1438; DOI 10.7717/peerj.1438.

Recognition of the patterns by persons not part of the United States Government may be found in (a) 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); (b) Lambert PJ, Subramanian S. Disparities in Socio-Economic outcomes: Some positive propositions and their normative implications. *Soc Choice Welf* 2014;43:565-576); (c) Eikemo TA, Skalicka V, Avendano M. Variations in health inequalities: are they a mathematical artifact? *International Journal for Equity in Health* 2009;8:32; (d) Houweling TAJ, Kunst AE, Huisman M, Mackenbach JP. Using relative and absolute measures for monitoring health inequalities: experiences from cross-national analyses on maternal and child

2012), Harvard University (Oct. 9, 2012), United States Department of Justice (Apr. 23, 2012), United States Department of Education (Apr. 18, 2012), The Commonwealth Fund (June 1, 2010), Institute of Medicine (June 1, 2010), National Quality Forum (Oct. 22, 2009), Robert Wood Johnson Foundation (Apr. 8, 2009)

health. *International Journal for Equity in Health* 2007;6:15; and (e) Carr-Hill R, Chalmers-Dixon P. The Public Health Observatory Handbook of Health Inequalities Measurement. Oxford: SEPHO; 2005.

The NCHS recognition of the pattern whereby the two relative differences tend to change in opposite as the frequency of an outcome changes should not be read to indicate that NCHS has found a satisfactory means of dealing with issues the pattern raises. NCHS should have regarded the fact that the two relative differences commonly (or ever) yield opposite conclusions regarding whether disparities are increasing or decreasing as calling into question the value of either relative difference for quantifying differences in health- and healthcare-related circumstances of advantaged and disadvantaged groups. But the agency simply determined that in appraising progress on the disparities reduction goals in Health People 2010, both health and healthcare disparities would be measured in terms of relative differences in adverse outcomes (that is, in the case of healthcare, non-receipt rather than receipt of appropriate care). This approach, which was adopted by the Agency for Healthcare Research and Quality in the National Healthcare Disparities Reports, caused healthcare disparities that previously were usually regarded as decreasing instead to be usually regarded as increasing. See "Race and Mortality Revisited" (at 331-335) and the FCSM paper (at 15-20). In 2015, however, NCHS reversed the recommendation such that in appraising progress on the disparities reduction goals in Health People 2020, healthcare disparities will be measured in terms of relative differences in favorable outcomes. Yet few people in or out of government recognize that this reversal effectively repudiates a decade of National Healthcare Disparities Reports and a substantial body of other research that, in reliance on earlier NCHS guidance, measured healthcare disparities in terms of relative differences in adverse outcomes. See my forthcoming "The Mismeasure of Health

Disparities” mentioned above. See also the April 6, 2016 letter to the Consortium of Social Science Associations (at 8, 16-19, 21-26) and the March 8, 2016 letter to Stanford Center on Poverty and Inequality (at 3-4).

The discussion in the preceding paragraph may not be directly relevant to issues now before the Court. But it may serve to counter the seemingly reasonable belief that because the government has for many years been monitoring demographic differences, it must be doing so in an essentially sound manner. That is not actually the case even as to the more scientific arms of the government. And it is certainly not the case with regard to arms of the government that have never recognized that it is possible for the two relative differences to yield opposite conclusions or that mistakenly believe the reducing the frequency of an outcome will reduce relative differences in experiencing it or the proportion disadvantaged groups make up of persons experiencing it.

In any case, as suggested in the Introduction and Summary, while few people are familiar with the patterns described above, the existence of the patterns is not open to dispute.

This does not mean that there will be no departures from the patterns, as I have explained in numerous places for a great many years. Observed patterns are functions of (a) the described frequency-related patterns and (b) the strength of the forces causing the outcome rates to differ in the settings being compared. The strength of those forces may vary greatly from setting to setting and may change substantially over time (especially longer periods of time) particularly during periods of dramatic changes in the frequency of an outcome. Such factors may cause the frequency-related patterns not to be observed even though they are having a substantial effect. The crucial consideration is that so long as measures tend to be affected by the frequency of an

outcome in any manner, one cannot rely on the measures to appraise the strength of the forces causing rates to differ without consideration of the role of the frequency of the outcome.

I note, however, that in the instant situation, one reason that there might occur departures from the standard frequency-related patterns involves potential changes in the strength of the forces causing outcome rates of African Americans and other persons to differ. The Department of Justice's Ferguson Report and its complaint in this case indicate a belief that part of observed racial differences in outcome rates (or disproportionality in experiencing adverse outcomes) is a function of racially biased conduct on the part of police and other officials, and certain actions required by the Consent Decree are intended to eliminate such conduct. (See discussion in Section B.4 *infra* regarding the role of the magnitude of disparities in supporting inferences of bias.) To the extent that observed differences are in fact partly functions of biased conduct, and the Consent Decree achieves reductions in such conduct, such reductions should reduce all measures of differences between outcome rates and all measure of disproportionality.

The same may also hold for certain specific requirements of the Consent Decree that are aimed at neutral practices rather than biased conduct. For example, consideration of ability to pay a fine in the course of the varied procedures and events that might ultimately lead to a sanction for nonpayment is a factor that could be sufficiently more beneficial to African Americans than to whites that the result would be to reduce all measures of differences between outcome rates and disproportionately.

Further, reductions of arrests for certain types of offenses, where racial differences (properly measured) are particularly large, may reduce all measures of overall arrest differences even if the reductions in arrests for the particular types of offenses increase relative differences in arrests for those types of offenses.

Nevertheless, the frequency-related patterns I describe are substantial enough that they commonly will be observed even if the strength of the forces causing outcome rates to differ changes substantially. Recently, the government has strongly encouraged public schools to relax discipline standards in order to reduce relative differences in discipline rates and the proportion racial minorities make up of persons disciplined. In doing so the government commonly expresses a belief that racial bias plays a role in observed disparities. There is thus reason to believe that schools that are generally reducing discipline rates are also taking actions to reduce disparities by eliminating bias or by other actions of a type that would reduce all measures of racial differences. Yet, consistently across the country, recent reductions in discipline rates have been accompanied by increased relative differences in discipline rates (which, as noted, necessarily correspond with increases in the proportions groups with higher discipline rates make up of persons disciplined). See the following subpages of the Discipline Disparities page of jpscanlan.com discussing situations where, in the jurisdiction indicated in the subpage title, recent reductions in discipline rates were accompanied by increased relative differences in discipline rates: California Disparities, Colorado Disparities, Connecticut Disparities, Maryland Disparities, Minnesota Disparities, Beaverton, OR Disparities, Denver Disparities, Henrico County, VA Disparities, Los Angeles SWPBS, Minneapolis Disparities, Montgomery County, MD Disparities, Portland, OR Disparities, St. Paul Disparities.¹³

But, whatever the actual implications of the factors just described, the fact remains that it is impossible to soundly interpret data on group differences without understanding the way each

¹³ There have no doubt been some contrary patterns. But the situations treated in the press or that have otherwise caught my attention have invariably or almost invariably involved patterns of increasing relative differences. Discussions of reductions of discipline disparities in press accounts or studies will generally involve situations where disparities are measured in terms of absolute (percentage point) differences, which, in the rate ranges at issue, tend to change in the opposite direction of the relative difference in the adverse outcome (as discussed in "Race and Mortality Revisited" and many other places).

measure tends to be affected by the frequency of an outcome. And it is certainly impossible to do so while having an understanding of the way a measure tends to be affected by the frequency of an outcome that is the opposite of reality.

In light of the above, and irrespective of the way the points made above or in the material cited above bear on perceptions about disparities in Ferguson warranting remedy, I suggest that it would be manifestly unjust to impose on the City the many disparity monitoring obligations in the Consent Decree without addressing the Department of Justice's misunderstanding of pertinent statistical issue.

B. Implications of the Patterns by Which Measures Tend to Be Affected by the Frequency of an Outcome With Respect to Particular Requirements of the Consent Decree

The points in the subsections below involve matters that are, for the most part, implicit in the discussion in Section A or specifically addressed in materials referenced in Section A. Nevertheless, there are a number of implications of the patterns described earlier that may not be obvious to all readers and that it is crucial for officials attempting to comply with, or monitor compliance with, the proposed Consent Decree to understand. Some of these are addressed in the subsections below.

1. The Fundamental Unsoundness of Analyses of Discrimination Issues Based on the Proportion a Group Makes Up of Persons Potentially Experiencing an Outcome and the Proportion the Group Makes Up of Persons Actually Experiencing the Outcome

Most of my recent extended treatments of the patterns by which measures tend to be affected by the frequency of an outcome discuss a means of quantifying the difference between the circumstances of two group reflected by their favorable or adverse outcome rates. By way of example, the favorable or adverse outcome rates in each row of Table 1 would yield a difference between the underlying means of .5 standard deviations. The rows of Table 2 would yield

values of .55 and .54 standard deviations.¹⁴ Other useful illustrations may be found in the tables of "Race and Mortality Revisited" and the FCSM paper and the methods workshops listed in note 10 *supra*. Particularly instructive is Table 8 (at 342) of "Race and Mortality Revisited," which involved what the Departments of Justice and Education in March 2014 regarded as especially large racial disparities in preschool suspension rates without consideration of the role of the fact that suspensions are extremely rare in preschool.

But in order to employ this measure (or any other plausible measure) one must have the actual outcomes rates. Information solely on the proportion a group makes up of persons potentially experiencing the outcome and the proportion it makes up of persons actually experiencing the outcome, while enabling one to derive the relative difference between rates of experiencing that outcome, does not enable one to derive the rates themselves. See discussion of this issue in the Section B (at 23-26) of the Kansas Law paper and Section I.B (at 23-27) of the TDHCA brief.

There are additional problems with appraisals of the differences in the circumstances of advantaged and disadvantaged groups based on comparisons of such proportions. These are touched up in the TDHCA brief (at 26-27) and addressed at length in the IDEA Data Center Disproportionality Guide and the Ferguson, Missouri Arrest Disparities subpages of the Discipline Disparities page of jpscanlan.com. See also slides 98-108 of the October 2014 University of Maryland methods workshop referenced in note 10 *supra*.¹⁵ But since comparisons

¹⁴ The fact that the value is not the same for each row is an indication of the imperfection of the method in the particular circumstance, given that the posited scenario involves a situation there solely occurred a change in the frequency of the outcome.

¹⁵ One of those additional problems is that for any given pair of rates, the larger the proportion the subject group makes up of persons potentially experiencing the outcome, the smaller will be the relative difference between the proportion the group makes up of persons experiencing the outcome the proportion it makes up of persons actually experiencing the outcome (or the corresponding ratio of the former proportion to the latter proportion, often called the Disparity Index, as in the submission of commenter Keith Kallstrom). See Table 2 of IDEA Data Center

of the referenced proportions do not provide a sound method for appraising differences in the circumstances of two groups in any event (just as relative differences in either outcome do not¹⁶), it is unnecessary to belabor here the additional problems with such analyses.

One must be mindful, however, that in order to derive outcome rates for groups being compared, it is necessary to identify the appropriate numerators and denominators.¹⁷ In some cases, identification of those figures is straightforward. But it is not straightforward in the case of arrest data. See the Addendum to the above-mentioned Ferguson, Missouri Arrest Disparities subpage of the Discipline Disparities page of jpscanlan.com. As reflected in the Addendum, I have yet to satisfactorily resolve this issue.¹⁸

But, as difficult as it may be to resolve this issue in order to employ a sound measure, such difficulty does not provide a reasonable basis for reliance on a fundamentally unsound measure.

Disproportionality Guide subpage and Table 24 (slide 108) of the University of Maryland workshop. An illustration of this problem may be found in discussion in the Kallstrom submission (at 8) regarding the fact that the Disparity Index for Ferguson was only 1.37, which was much smaller than the average for the randomly selected jurisdictions whose data were studied. Given that the African American proportion of Ferguson's population is 67%, even if African Americans were subjects of 100% of traffic stops, the Disparity Index could not exceed 1.49. The largest Disparity Index shown in the table on page 2 of Mr. Kallstrom's submission is the 10.72 figure for Glendale, while the next largest is 5.02. The size of the Glendale Disparity Index is no doubt in some part a function of the fact that African Americans make up less than 1% of the population of Glendale.

¹⁶ See "Race and Mortality Revisited" (at 339-41) and the American Statistical Association letter (at 12-13) regarding the fact that, irrespective of the patterns I describe above, the rate ratio (or its associated relative difference) is an illogical measure of association., See also the Illogical Premises, Illogical Premises II, Subgroup Effects, Subgroup Effects – Nonclinical, and Inevitability of Interaction subpages of the Scanlan's Rule page of jpscanlan.com and the Comment on Hingorani BMJ 2013. For an illustration of this point with regard to the "four-fifths rule" of the Uniform Guidelines on Employee Selection Procedures, see the Four-Fifths Rule subpage of the Disparate Impact page of jpscanlan.com

¹⁷ One should not confuse issues concerning the appropriate numerators and denominators for the purpose of calculating each group's outcome rate with the issue concerning which group's outcome rate to use as the numerator in the rate ratio (which was discussed in note 5 *supra*).

¹⁸ See the Kansas Law paper (at 21) regarding a similar interpretive problem in the employment context. See also my "Measuring Hiring Discrimination," *Labor Law Journal* (July 1993), regarding this problem in the employment tester context.

2. The Common Inadequacy of Adjustment for Differences in Characteristics and the Implications of the Statistical Significance of Residual Disparities

The Consent Decree requires the City to conduct various regression analyses to identify racial disparities that exist after adjustment for certain characteristics that might provide a nondiscriminatory explanation for disparities. Regardless of the extent to which such analyses may be subject to the measurement problems described above, the same features of risk distributions that underlie the above-described pattern of relative differences undermine efforts adequately to adjust for characteristics. Specifically, the group that is disproportionately concentrated in a category that is associated with higher rates of experiencing some outcome will tend to be even more disproportionately concentrated in each increasingly more susceptible segment of the category. For example, as shown in the Table 2, African Americans, who are disproportionately represented among persons living on incomes below the poverty line, are even more disproportionately concentrated among persons living on incomes below 75% of the poverty. For the same reason, among persons living with incomes between the poverty line and 75% of the poverty line, African Americans will be concentrated in the lower ranges of the category (as well as in the lower ranges of persons living on incomes below 75% of the poverty line and above the poverty line itself). For a more complete illustration of this issue (as reflected in income data), see Table 1 and Figure 1 of the 2006 *Chance* [editorial](#).

In consequence of such pattern, analyses that attempt to adjust for characteristics by grouping people into even seemingly refined categories related to likelihood of experiencing an outcome virtually never fully adjust for differences in characteristics. Hence, there will almost

always be some residual disparity.¹⁹ There is a tendency to regard the residual disparity as meaningful (that is, demonstrative of bias) when it is found to be statistically significant.

But inasmuch as there is reason to expect a residual disparity whether or not bias is present, the fact that the residual disparity is statistically significant, or even highly statistically significant, does nothing to address the question of whether bias in fact is responsible for all or part of the residual disparity. Thus, the only sound, albeit imperfect, approach to determining the likelihood that bias is responsible for some part of the residual disparity is to make a judgment, based on all the circumstances, of whether the residual disparity is large enough to render it unlikely that it is solely a function of the failure of the analysis to adjust adequately characteristics.

For any such judgment to be sound, however, it must be informed both by an understanding of the measurement issues discussed in the references cited above and by recognition of the limited implications of the fact that the residual disparity is statistically significant.

3. The Drawing of Inferences on the Basis of the Comparative Size of Relative Differences

"Race and Mortality Revisited" (at 339-341) discusses ways that observers commonly draw inferences on the basis of the fact that a relative difference (or the proportionate effect of a factor) is larger in one setting than another, with regard to one type of outcome than another, or with regard to one subpopulation than another. The article also explains that invariably

¹⁹ In the case of *EEOC v. Sears, Roebuck and Co.* mentioned at the outset of this submission, the plaintiff EEOC attempted to prove that women were discriminatorily excluded from commission sales positions. In doing so, EEOC attempted to adjust, to the fullest extent that data would permit, for differences in selection-related characteristics of male and female sales applicants. It was made clear during the course of the ten-month trial, however, that no matter what the EEOC did, given the limitations of the data, it was never able to fully adjust for those differences. See the [Sears Case Illustration](#) subpage of the [Scanlan's Rule](#) page of [jpscanlan.com](#).

observers draw these inferences without understanding the effects of the frequency of the outcome, and without appreciating that the comparative size of relative differences (or proportionate effects) for the opposite outcome would commonly support a very different inference.

One example of statistically unsound inferences of this nature involves the pattern where relative demographic differences in selection rates tend to be smaller among highly-credentialed job applicants than among applicants without such credentials. Observers interpret this pattern to indicate that decision-makers exhibit less bias where objective indicators of qualifications are present than when such indicators are not present. But in drawing that inference, observers have failed to understand the implications of the differing frequencies of selection for applicants with and without objective indicators of qualifications and have overlooked that relative differences in rejection rates tend to be greater for highly-credentialed job applicants than applicants without such credentials. See discussion of this issue in "Race and Mortality Revisited" with respect to its Table 5 and in the Kansas Law paper with respect to its Table 1 (especially note 16 on page 14).

The point can also be illustrated by Table 1 *supra* if one simply regards the first row as reflecting the situation of less qualified job applicants and the second row as reflecting the situation of more qualified job applicants, and regards the test passage and failure rates as hire and rejection rates. Among the more qualified applicants, the relative difference in selection rates is smaller, but the relative difference in rejection rates is larger, than among the less qualified applicants.

See also the [Offense Type Issues](#) subpage of the [Discipline Disparities](#) page of jpscanlan.com regarding inferences based on patterns of racial differences in school discipline

rates depending on whether the offense tends to be objectively identified or subjectively identified.

The Department of Justice's Ferguson Report discussed the fact that African Americans made up a larger proportion of persons subject to traffic stops based on method such as an officer's visual assessment than stops based on radar or laser. At page 8, the report states that the disparity is 48% percent larger in the former case. At page 69 the report states that African Americans accounted for 72% of citations based on radar or laser and 80% of citations based on other methods, further stating that, "controlling for other factors," this disparity was 48% larger in the latter case. The complaint in this case (Paragraph 116, at 34) makes the same point, though without providing figures. Presumably, the Department cites these differences as an indicator of racial bias (or greater racial bias) in stops based on non-objective indicators.

I do not understand how the 48% larger difference was calculated or the implications of the control for other factors. But the fact that African Americans made up a larger proportion of persons stopped based on other methods than based on laser or radar would seem to indicate that the relative racial difference is larger for the former types of stops than the latter (as discussed in Section B.1 *supra*).

In interpreting the implication of this pattern, however, one must consider the following. Assuming that the fact that African Americans make up a larger proportion of persons stopped for speeding on the basis of radar or laser than they make of the pertinent driver population indicates that African Americans more commonly speed than other drivers, one would typically find that African Americans make up an even larger proportion of persons exceeding the speed limit by a substantial amount than they do of persons exceeding the speed by a modest amount. Suppose then that officers will more readily stop persons who are merely exceeding the speed

limit by modest amounts when they have radar or laser to indicate the level of excess speed than when they must make such assessment visually, but that there is little difference in stop rates (depending on method of assessment) when drivers are exceeding the speed limit by substantial amounts. The end result will be that African Americans will tend to make up a larger proportion of persons stopped on the basis of visual assessment than based on laser or radar.

I, of course, do not know whether the posited scenario is accurate. I have no basis even to be certain that stops based on laser or radar are unbiased and hence that the proportion African Americans make up of person stopped on such basis accurately indicates that they speed more often than other drivers. But the scenario is not implausible.

The key point, however, is that it is impossible to interpret data without understanding the patterns by which measures tend to be affected by the frequency of an outcome that I have briefly described above and that I have described at greater length in the references cited above. That includes understanding that circumspection of the part of law enforcement officers and other officials, unless focused solely on bias itself, will tend to increase relative differences in the outcome that circumspection tends to reduce, while tending to reduce relative differences in the corresponding outcomes that circumspection tends to increase.

The same points hold with regard to inferences based on increasing relative racial differences (or increasing disproportionality) at each deeper level of the criminal justice system, as I explained almost two decades ago in "Mired in Numbers," *Legal Times* (Oct. 12, 1996). Now, as then, one cannot sound draw inferences on the basis of increasing relative differences in experiencing adverse outcomes at each deeper level of the system – or decreasing relative differences in rates of avoiding those outcomes at each deeper level – without understanding the statistical implications of fact the criminal justice system increasingly restricts severer adverse

outcomes to the segments of the overall (relevant) population that are most susceptible to the outcome (meaning, in this context, more warranting of severer adverse outcome). The point also holds with respect to predicting the statistical consequence of further restricting any adverse outcome to the segments of the overall relevant population that are most susceptible to the outcome.

4. Inferring Bias on the Basis of the Magnitude of Disparities

Section G of the complaint (at 32-40) alleges that racial bias underlies the City's implementation of policies that disproportionately affect African Americans. Paragraph 122 (at 35) suggests that a basis for inferring racial motivations may be found in the magnitude of the disparities and the continuation of policies in light of the evident magnitude of the disparities.

In order to soundly appraise the magnitude of a disparity, one must understand that there is no rational basis for maintaining that the disparity is larger in one row of Table 1 than in the other (though in a testing case the Department of Justice would maintain the first row reflects the larger disparity and in a lending case it would maintain that the second row reflects the larger disparity). The same holds for the four rows of data in Table 5 (at 335) of "Race and Mortality Revisited," versions of which may also be found as Table 1 (at 14) of the Kansas Law paper (at 14) and Table 5 (at 22) of the American Statistical Association letter (and in many other places). But, like others analyzing discrimination issues (or health and healthcare disparities issues), the Department of Justice does not yet understand this issue.

This does not necessarily mean that a perception about magnitude could not underlie racially motivated decisions. Like the Department of Justice, City officials might fail to understand that increasing the number of missed court appearances to trigger issuance of an

arrest warrant would tend to increase, not decrease, the proportion African Americans make up of persons against whom such warrants are issued.

Nevertheless, any drawing of inferences about motivations based on either the actual magnitude of disparities, or decision-makers' mistaken perceptions about magnitude, must be informed by an understanding of the patterns described briefly here and treated at greater length in the references discussed above.

5. Disparities in Favorable Outcomes

I have not had time to fully familiarize myself with the consent decree or divine the background to, or intentions respecting, particular requirements. But I have noticed that, whereas the Department of Justice's allegations respecting disparities in Ferguson have generally focused on the adverse outcomes, the Consent Decree contains provisions for monitoring disparities with respect to outcomes that are cast in both favorable terms and adverse terms. Items 1 to 5 and 8 of the monitoring subjects identified in the Consent Decree's Paragraph 435.c (at 109-110) are cast in terms of adverse outcomes. Item 6, which pertains to issuance or recall of municipal arrest warrants, is cast in terms of both an adverse outcome and a favorable outcome. Item 7 (reducing or modifying charges) and Item 9 (reducing fines) are cast in terms of favorable outcomes.

Since the Consent Decree requires that these disparities be monitored by means of regression analyses I am uncertain of the effects of the frequencies of the outcomes, which may turn on the type of regression method and the frequency ranges at issue. See "Race and Mortality Revisited" (at 336) and the TDHCA brief (at 16) regarding logistic regression. See also note 20 (at 24) of the April 6, 2016 letter to the Consortium of Social Science Associations.

It nevertheless warrants mention that that the greater leniency in law enforcement that seems to be a general object of the degree might be expected to increase the favorable outcomes mentioned in Items 6, 7, and 9. That would tend to reduce relative differences in rates of experiencing those outcomes and increase the proportion the group with a lower rate for experiencing the outcome makes up of persons experiencing the outcome. That is, the result of increasing these favorable outcomes is as reflected in the fourth and second last columns of Table 1 *supra*.

Thus, those examining the changes in disparities according to the usual measures should not be surprised that patterns of changes in disparities for matters cast in favorable terms are the opposite of patterns of changes in disparities for matters things cast in adverse terms.

C. Suggested Disposition of This Matter

In light of the points made above, including those pertaining to the key premise of the Department of Justice's actions regarding the City of Ferguson, I respectfully suggest the Court defer decision on the approval of the Consent Decree and proceed as discussed below. I attempt to limit to the suggestions to matters as to which my knowledge of the issues addressed above may give me pertinent insight.

I assume that the City of Ferguson can withdraw its consent to the Consent Decree any time prior entry of the decree and it may be sensible to allow the City some time to make that determination in light of points made in this submission. Assuming that the City does wish to go forward with entry of a Consent Decree of the nature of that proposed, I suggest that the Court proceed as follows.

First, the Department of Justice should be given a reasonable period time to file a pleading explaining to the Court whether, and the extent to which, its suit and the proposed

Consent Decree are based on the premise that unwarrantedly aggressive policing and unjustifiably harsh court practices cause African Americans to make up a higher proportion of persons experiencing adverse outcome at the hands of the police and the courts than would be the case with less aggressive policing and less harsh court practices, and whether it continues to believe that the premise is sound. In light of my understanding of the issues, I suggest that 30 days would be a reasonable period of time for the Department to provide a response on this limited matter. But, given what persons with substantial mathematical or statistical backgrounds have advised me regarding the difficulty of grasping that the two relative differences tend to change in opposite directions as the frequency of an outcome changes, I cannot say that a request of the Department of Justice for significantly greater time than that would be unreasonable.

Second, following the Department's submission of its response on this limited matter, the City of Ferguson should be given a reasonable period of time to consider the response and to decide whether, in light of the response, it continues to wish to enter into a Consent Decree of a type similar to that proposed. My understanding of the statistical issues in this matter does not provide me a basis to suggest to the Court what such reasonable period of time should be.

Third, assuming the City of Ferguson does wish to go forward with a Consent Decree of this nature, the Department of Justice should be given a reasonable period of time to specify the methods by which the City of Ferguson will monitor racial differences in outcomes (including with respect to whether one practice has less of a disparate impact than another), and by which the Department will oversee such monitoring, that take into account, or are unaffected by, the frequency of the outcome examined. Based on my understanding of the statistical issues, I believe that a reasonable period of time would be no less than 90 days. And I note that, as suggested at the end of Section B.1 *supra*, there are issues that I remain uncertain about after

considering these issues for many years. So, in my view, a claim by the Department that it needs considerably more than 90 days would not be unreasonable.²⁰

Fourth, upon the Court's approval of the method proposed by the Department of Justice or some other method, the City of Ferguson should be given a reasonable period of time to determine whether it will consent to a decree into which such method is incorporated.

There may well be more sensible or efficient specifics for resolving the measurement issues and their associated uncertainty issues. But I suggest that those issues are important enough that resolution should be deemed a necessary predicate to entry of a Consent Decree of this nature.

My understanding of the issues does not provide me a basis, or justification, to suggest to the Court how it should proceed after those issues are resolved.

I acknowledge that the recommended approach is unprecedented. I assume that many hundreds or thousands of cases have been resolved, and like numbers of consent decrees have been entered, without consideration of the issues I maintain are so important. But the fact is that the government and the courts have generally failed to understand such simple facts as that the two relative differences tend to change in opposite directions as the frequency of an outcome changes or that there is no rational basis to maintain that strength of the forces causing the outcome rates of the advantaged group and disadvantaged group to differ varies between the two rows of Table 1.

A city should not be made subject to a court decree while not knowing whether its actions will be regarded as having increased racial disparities (according to one measure) or having

²⁰ I note that in requesting that the American Statistical Association and other entities form committees to fully address the measurement issues addressed in my letters to them I have assumed that, in the event such committees are formed, their work may take several years. But I do not expect a Court to consider such protracted time frames in a matter such as this.

decreased disparities (according to a different measure). That is especially so in a context where there is reason to believe that actions the decree requires the City to take will tend to increase disparities according to the way the government is likely to measure them.

Date: April 11, 2016

Respectfully submitted,



James P. Scanlan