

Compliance Nightmare Looms for Baltimore Police Department

James P. Scanlan

The Federalist Society Blog (fedsoc.org/blog)

(Feb. 8, 2017)

In my January 4, 2017 post titled “[Will Trump Have the First Numerate Administration?](#),” I discussed Department of Justice (DOJ) actions regarding police practices in Baltimore, Maryland in the context of the longstanding situation where federal civil rights law enforcement policies have been based on an understanding of statistics that is the opposite of reality. Specifically, with regard to matters including lending, school discipline, employment, criminal justice, and voting, many government policies have been premised on the belief that relaxing standards or otherwise reducing the frequency of adverse outcomes tends to reduce (a) relative (percentage) racial and other demographic differences in rates of experiencing those outcomes and (b) the proportions more susceptible groups make up of persons experiencing the outcomes. In fact, generally reducing any outcome tends to increase, not decrease, (a) and (b).

Believing the opposite, however, the government has encouraged entities subject to the constitution or civil rights laws to take actions that generally reduce adverse outcomes. And, because the government continues to monitor the fairness of practices based on the size of (a) or (b) as to the particular adverse outcomes at issue, entities that accede to encouragements to reduce adverse outcomes increase the chances that the government will sue them for discrimination.

In the post, I discussed prospects that the new administration might recognize the above and other fundamental problems in analyses of demographic differences by governmental and nongovernmental entities.

At the time, DOJ and the City of Baltimore, pursuant to an agreement reached in August 2016, were negotiating a consent decree to remedy constitutional and statutory violations by the Baltimore Police Department (BPD) identified in a DOJ report issued the day before the agreement. In the post, I discussed pressures on the parties to submit a decree to the court before the change in administrations.

On January 10, 2017, DOJ filed a complaint in federal court against BPD and Baltimore’s Mayor and City Council alleging, among other things, that BPD systematically used unnecessary force and employed varied policing tactics that disproportionately affected African Americans. Contemporaneously, the parties submitted a joint motion requesting court approval of a proposed consent decree.

At a court hearing on February 1, DOJ counsel indicated that the agency intended to follow through with the consent decree notwithstanding the change in administrations. And Baltimore’s Mayor assured the court that the city would be able to bear the expense of implementing the decree, which one account estimated to run between nine and thirty million dollars. The term of the decree will turn on BPD’s fully complying with it, so total cost of the decree will depend on when compliance is achieved.

One part of this expense will be up to \$1,475,000 per year (though subject to increase in particular years where deemed necessary by the court) for an entity to serve as independent monitor of BPD's compliance with the decree. Tasks of the monitor include analyzing data on racial and other demographic differences in policing using "statistical techniques that are accepted in the relevant field."

The court indicated that, in accordance with the parties' request, it would accept written submissions from the public regarding the proposed decree and hold a public hearing on it. The court has yet to specify procedures for public comment.

Effects of Reducing Adverse Outcomes on Measures of Demographic Differences

The proposed decree, like the report and complaint underlying it and actions the Departments of Justice and Education have taken all across the country, is premised on the belief that reducing the frequency of adverse outcomes will tend to reduce relative racial differences in rates of experiencing those outcomes and the proportions African Americans make up of persons experiencing the outcomes. Despite an effort on my part to advise counsel for both parties that the belief is mistaken, it is doubtful that either party yet understands the matter and virtually certain that the court does not.

Beyond the general monitoring of demographic differences that is an essential aspect of the decree, among its features that are of particular pertinence to statistical issues addressed here are requirements that, both with respect to policing and hiring, BPD shall implement less discriminatory alternatives to practices having a disparate impact on particular demographic categories (as defined by "race, ethnicity, color, national origin, age, gender, gender expression or identity, sexual orientation, disability status, religion, or language ability"). The decree also requires that in evaluating officer performance and making promotion decisions, BPD shall consider whether officers have engaged in discriminatory policing or discriminated against other BPD employees.

To facilitate readers' understanding of the points that follow I set out in Table 1 the data underlying an example I have verbally expressed in a number of prior posts here and varied other places. The table shows (in numbered columns 1 through 4) 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). It also shows (in columns 5 through 8) measure that might be used to appraise differences in test outcomes of AG and DG.

Column 5 shows that at the higher cutoff, where pass rates are 80 percent for AG and 63 percent for DG, AG's pass rate is 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.

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) (a) DG/AG Fail Ratio	(7) DG Prop of Pass	(8) (b) DG Prop of Fail
1	80%	63%	20%	37%	1.27	1.85	44%	65%
2	95%	87%	5%	13%	1.09	2.60	48%	72%

That lowering a cutoff tends to reduce relative differences in pass rates is well understood in civil rights circles and underlies the widespread view that lowering a cutoff tends to reduce the disparate impact of tests where 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 (which is also designated (a) to correspond with the usage in the first paragraph), 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 to 2.6 times (160 percent greater than) AG’s failure rate.

Columns 7 and 8 show the proportions DG makes up of persons who pass and fail the test at each cutoff in a situation where DG makes up 50 percent of persons taking the test. Column 7 shows that lowering the cutoff increased the proportion DG makes up of persons who passed from 44 percent to 48 percent (hence, *reducing* all measures of difference between the proportions DG makes up of persons who took the test and persons who passed the test). And Column 8 (also designated (b) to correspond with usage in the first paragraph) shows that lowering the cutoff increased the proportion DG makes up persons who failed the test from 65 percent to 72 percent (hence, *increasing* all measures of difference between the proportions DG makes up of persons who took the test and persons who failed the test).

These patterns are not peculiar to test score data or the numbers I used to illustrate them. Rather, they exist to a degree in essentially all situations where groups differ in their susceptibility to some outcome (and its opposite), as illustrated, for example, in my “[Race and Mortality Revisited](#),” *Society* (July/Aug. 2014), my [comments](#) for the Commission on Evidence-Based Policymaking (CEBP) (Nov. 14, 2016), and my October 2014 University of Maryland [workshop](#), as well as scores of other places. That is, the less frequent the outcome, the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it (*i.e.*, experiencing the opposite outcome); correspondingly, the less frequent an outcome, the greater tends to be the proportion the more susceptible group makes up of persons experiencing it and avoiding it. But very few people analyzing demographic differences understand these patterns and a great many of them, like the government, believe that reducing the frequency of an outcome tends to reduce relative differences in experiencing it and the proportion the more susceptible group makes up of persons experiencing it.

In consequence of these patterns, whenever there occur increases in one outcome with corresponding decreases in the opposite outcome, observers who analyze disparate impact issues

in terms of relative differences in the increasing outcome (or the corresponding proportion the disadvantaged group makes up of persons experiencing that outcome) will commonly find a reduction in disparate impact. In the same circumstances, observers relying on relative differences in the decreasing outcome (or the corresponding proportion the disadvantaged group makes up of persons experiencing that outcome) will commonly find that the disparate impact has increased. But whether changes in policies or other factors that affect the frequency of an outcome in fact reduce or increase a disparate impact involves a more complicated inquiry in which these measures do not play a role. See pages 27-32 of my "[The Mismeasure of Discrimination](#)," Faculty Workshop, University of Kansas School of Law (Sept. 20, 2013) (at 27-32) and my "[Is the Disparate Impact Doctrine Unconstitutionally Vague?](#)," Federalist Society Blog (May 6, 2016) (at 7-8 of the PDF version available [here](#))

The described frequency-related patterns are also pertinent to interpretations regarding the likelihood of bias on the part of particular decision-makers. Consider the two rows of Table 1 as pertaining to a process where two decision-makers allocate favorable and corresponding adverse outcomes between members of two groups. Observers examining favorable outcomes would say the decision-maker in row 1 is more likely to be biased (or exhibits greater bias) than the decision-maker in row 2. Observers examining adverse outcomes would reach an opposite conclusion. Either approach, however, would be an accepted technique in the analyses of demographic differences.

Which is the more defensible position? Or does the answer depend on the nature of the favorable and corresponding adverse outcomes?

Neither position is defensible at all. Regardless of the types of outcomes at issue, there is no rational basis for distinguishing between the two rows as to the likelihood that bias influenced decisions. For the same reason, it is impossible on the basis any of the measures in Table 1 to determine whether the likelihood of bias in a particular case should be deemed great or small without information as to the frequency of an outcome, and, of course, knowledge on how to use that information.

See "Race and Mortality Revisited" at 335-336 for discussion of the same point with regard to a broader range of values and other measures that tend to be affected by the frequency of an outcome and at 339-342 regarding the varied flawed inferences observers draw on the basis of the comparative size of the relative differences (in the favorable or the adverse outcomes) that, as a result of convention or chance, they happen to be looking at in the particular situation. See my "[The Mismeasure of Health Disparities](#)," *Journal of Public Health Management and Practice* (July/Aug. 2016), regarding the way the National Center for Health Statistics, on recognizing that relative differences in receipt and non-receipt of appropriate care tend to change in opposite directions as healthcare generally improves, arbitrarily decided to measure healthcare disparities in terms of relative differences in non-receipt of care, and, a decade later, just as arbitrarily reversed itself.

Appraising BPD Compliance With the Consent Decree

The Baltimore consent decree would require a range of actions aimed at reducing things like police stops, searches, arrests, and the use of force, all premised on the belief that doing so should reduce relative demographic differences in rates of experiencing those outcomes and the proportions more susceptible groups make up of persons experiencing the outcomes. But the extensive analysis of the data contemplated by the decree commonly will show the opposite.

Thus, the more BPD reduces these outcomes, the more the measures typically employed by the government will tend to show increasing disparate impacts of policies and increasing evidence of biased policing.

To be sure, increases in these measures are not inevitable. Given the way in which certain outcomes disproportionately occur in Baltimore neighborhoods where African Americans comprise a very high proportion of residents, it is hard to predict effects on overall measures of difference of general reductions in aggressiveness of enforcement (or police presence) in particular neighborhoods. A similar issue exists with respect to changes in approaches to different types of crimes. And to the extent that any observed differences in outcome rates are functions of biased policing, and aspects of the decree reduce that bias, all measures of difference will tend to decrease.

But, by and large, the described patterns will be observed in situations where there occur substantial changes in the overall frequency an outcome even in the presence of countervailing factors. See the CEBP comments at 27 regarding the varied jurisdictions around the country, including the State of Maryland and Montgomery County, Maryland, where recent reductions in school discipline rates have been accompanied by increased relative racial/ethnic differences in discipline rates notwithstanding that teachers and administrators are presumably taking a range of actions aimed at reducing racial/ethnic differences beyond simply relaxing standards.

It is thus important to recognize that general reductions in an outcome may increase relative differences in rates of experiencing the outcome even when any bias contributing to those differences has decreased. Further, even when the described patterns are not specifically observed, they will almost invariably have an influence (driving one relative difference in one direction and the other in the opposite direction). In any case, it is not possible to analyze demographic differences or changes therein in any useful way without understanding the described patterns (and other patterns described in the references mentioned above).

The same issues exist with regard to patterns of conduct by particular officers. For example, assuming that the failure rates in Table 1 reflect the proportion of arrest or other confrontational situations where an officer uses force, officers who are most circumspect about the use of force or best master de-escalation techniques will tend show results more akin to those in row 2 than row 1 (*i.e.*, higher values for (a) and (b)), while other officers will tend to show results more akin to row 1 than row 2. Without understanding these patterns it is not possible to appraise, for purposes either of officer evaluation or officer discipline, the likelihood that particular officers or groups of officers have engaged in discriminatory policing.

The issues also apply to appraisals of the fairness of supervisor treatment of subordinates. If officers of different demographic groups differ in average level of job performance or rates of infractions, supervisors who have the most lenient standards or who best train or motivate their subordinates to meet standards will tend to show the largest values for (a) and (b) in the imposition of discipline for failure to meet standards. Even actions aimed specifically at ensuring the fairness of discipline, by generally reducing discipline rates, can tend to increase values for (a) and (b). See my "[Getting it Straight When Statistics Can Lie](#)," *Legal Times* (June 23, 1993).

The decree requires that BPD, with the aid of the monitor, conduct an in-depth review of hiring procedures to identify practices with a disparate impact on any demographic category and to implement less discriminatory alternatives to such practices. In the case tests that requirement may lead to recommendations for lower cutoffs, where at least the effects on relative differences in pass rates (the likely measure of impact) will tend to be as expected. But in the case of background factors that might be regarded as disqualifying criteria, it is doubtful that it will be recognized that larger relative differences in disqualification rates are functions of the leniency, rather the stringency, of the criteria or that further relaxing of the criteria (a commonly proposed means of reducing the impact) will tend to increase relative differences in disqualification rates. See my web page on the case of [Jones v. City of Boston](#).

Effectively Appraising Demographic Differences in Criminal Justice Outcomes

More so than in most other areas where disparate impact or biased decision-making is of concern, demographic differences in criminal justice outcomes are commonly analyzed in terms of a comparison of the proportion a group makes up of persons experiencing an outcome and the proportion it makes up of a population rather than relative differences in rates of experiencing the outcome. Such analyses implicate anomalies beyond the fact that the former proportion tends to be affected by the frequency of an outcome, as discussed, for example, in my [IDEA Data Center Disproportionality Guide](#) web page and slides 98 to 108 of the Maryland workshop. But a fundamental problem with such analyses is that they are not based on the rates at which two groups experience an outcome, and one must have those rates for any sound appraisal of differences in the circumstances of two groups regarding the outcome and its opposite. See Section B (at 23-26) of the Kansas Law paper and Section I.C (at 39 to 41) of the CEBP comments.

To identify each group's rate, however, one needs both a numerator and a denominator. Generally the numerator is evident enough. But the appropriate denominator regarding things like arrests or stops are commonly very difficult to identify, both conceptually and practically. As indicated in the Addendum to my [Ferguson, Missouri Arrest Disparities](#) web page, I have not satisfactorily resolved the conceptual issue in my own mind.

Resolving these issues would be a useful undertaking for the DOJ, but only after it has come to understand certain more fundamental matters concerning the relationship between the frequency of an outcome and measures of differences in rates of experiencing the outcome.

Selecting a Monitor and Otherwise Reconsidering Actions Toward BPD and Other Entities Covered by Similar Decrees or Agreements

As discussed in the January 4 post, the government's misunderstanding of the effects of reducing an outcome on relative differences in rates of experiencing an outcome and the proportions more susceptible groups make up of persons experiencing the outcome is but an element of the larger failure to understand ways standard measures of differences between outcome rates tend to be affected by the frequency of an outcome. And that failure has managed to persist for so long because it is also pervasive within the scientific and statistical communities that specialize in the analysis of demographic difference.

An important part of the functions of the monitor of the decree will be the analysis of the demographic differences that are a particular focus of the decree. If a monitor were chosen tomorrow, the likelihood that the monitor would understand the issues addressed here is negligible.

But in addition to the parties' seeking public comment on the decree, the decree itself envisions a thorough process for selection of the monitor. Whatever the efficacy of the process with regard to choosing a capable monitor, it may provide an opportunity for the DOJ to educate itself on the analysis of demographic differences. If the agency accomplishes that, it ought to recognize an obligation to examine all the varied agreements it and other agencies have with local authorities that are based on misunderstandings of statistics, especially agreements requiring actions that tend to increase demographic differences according to the government's method of measuring them. See my letter to [Oklahoma City School District](#) (Sept. 20, 2016) regarding a recent Department of Education agreement that envisions both a general reduction in suspensions and a reduction in the proportion African Americans make up of students suspended. See also my April 11, 2016 [submission](#) to the court regarding the Ferguson, Missouri consent decree.

Assuming public comment or anything else causes the court in the Baltimore case to grasp the key statistical issues, the following would be a sensible approach for it to deal with the matter: Require counsel for both parties to read this post and advise the court whether they understand the statistical issues it raises and whether they agree that the proposed decree contemplates changes in policies that will tend to increase demographic differences according to the way the government measures them. Assuming they do agree, the parties then should propose modifications to the decree that not only avoid this anomaly, but provide sound approaches to measuring demographic differences in ways that are unaffected by, or give appropriate consideration to the effects of, the frequency of the outcomes examined. If the parties do not agree with the points raised here, they can explain why and the court can then evaluate those explanations.

And, of course, the parties can withdraw the decree and decide whether and how to go forward after careful consideration of whether and how the government can effectively analyze demographic differences in criminal justice outcomes.