Presumed Guilty Until Proven Innocent:

Benchmarking is Not a Valid Statistical Approach

Ulrich Schimmack

University of Toronto Mississauga

Author Note

Address: Ulrich Schimmack, University of Toronto, Mississauga, 3359 Mississauga Rd., L5L 1C6, email: ulrich.schimmack@utoronto.ca

Abstract

Cesario, Johnson, and Terrill (2019) repeatedly state that their data show no evidence of racial disparity in police shootings or even a bias to shoot more White citizens. I show that these conclusions rest on a novel and flawed statistical approach that the authors call benchmarking. Benchmarking makes strong assumptions rather than testing the causal influence of potential mediators (e.g., crime rates) on racial disparities in police shootings. Although crime is likely to be a contributing factor, the data do not test this. The authors claim that police has the opposite bias to shoot more White homicide suspects than Black homicide suspects is based on unproven assumptions that are likely to be false. The only empirical contribution of the article is to show that even among unarmed citizens who do not pose an immediate threat, Black citizens are 3 times more likely to be killed by police than White citizens. Thus, the actual conclusion that can be drawn from the data is that there is evidence of racial disparity in police shootings in the general population and among non-threatening civilians. This suggests that better training of police officers in use of force situations can reduce the number of people killed by police, who happen to be disproportionally Black.

Keywords: Benchmarking, Mediation, Causality, Correlation, Police Shootings, Race, Inequality, Racial Disparity

Presumed Innocent Until Proven Guilty: Why Benchmarking is Stereotyping

Cesario, Johnson, and Terrill (2019) published an article in this journal that alleged to examine racial disparities in police use of force. In this article, the authors make several unscientific claims that are based on a misunderstanding of statistics.

First, they claim that it is wrong to compute the compute the proportions of Black and White US citizens that are being killed by police relative to the proportion of Black and White US citizens in the population. I will show that doing so is perfectly reasonable to examine the question whether there are racial disparities in use of force. Second, the authors claim that they find no systematic evidence of anti-Black disparities in fatal shootings. This claim rests on the calculation of the wrong odds statistics that are only justified by strong and false assumptions. I show that their own data are entirely consistent with the hypothesis that police proportionally shot and kill more Black US citizens than White US citizens.

Based on their unscientific analyses of the data, make a number of claims about the causes of police shootings of US citizens that are based on their flawed statistical analyses or are not based on any scientific evidence at all. First, they authors suggest that different exposure rates to Black and White US citizens explain why police officers are more likely to shot and kill Black US citizens. They provide no data on actual exposure rates to Black or White citizens, and they provide no theoretical argument why high exposure to Black citizens or Black criminals causes police to shot and kill more Black civilians. Cesario et al.'s (2019) argument is that daily experiences of encountering Black citizens during investigations of criminal activities "reinforce associations between Black and criminals" (p. 592). This is pure speculation and ignores that police officers also encounter White criminals, which suggests that police officers are more likely to have an accurate representation of the association between race and crime than the

general population, which has no personal experiences to counteract stereotypes. In fact, the authors later suggest that more contact with Black citizens could reduce racial bias. Thus, their speculations about the influence of exposure on stereotypes is contradictory, which only shows that they do not know how exposure influences officers' stereotypes.

Cesario et al. (2019) then speculate that stereotypes may explain why "likely to misidentify a harmless object in the hands of a Black citizen due to stereotypes" (p. 592). This claim is also not supported by their article, but it also contradicts their main conclusion that there are no racial disparities in policing. How else would we characterize the fact that police officers are more likely to misidentify objects as weapons when they are held by a Black citizen, which may lead to a fatal shooting of the civilian. These kinds of mistakes are exactly the cases that are used in the media and by representatives for Black communities to argue that there is racial bias in policing. Thus Cesario et al. (2019) contradict themselves when they claim that there are no racial disparities and that police officers are more likely to misidentify harmless objects as weapons when they encounter a Black citizen.

Based on their unsubstantiated and inconsistent line of reasoning, Cesario et al. then suggest "that department-wide attempts at reform through programs such as implicit bias training will have little to no effect on racial disparities in deadly force" (p. 592). This claim is also not supported by relevant evidence and makes sweeping claims that undermine attempts to reduce police killings by reducing the influence of stereotypes that may lead to accidental killings of both Black and White citizens. Cesario et al. (2019) then cite a study with undergraduate students with short manipulations to claim that interventions do not reduce prejudice. However, only a real intervention study with police officers can examine whether programs to reduce unnecessary use of force are effective. Unfortunately, social psychologists do not conduct these kinds of real-

word studies, which are costly and time-consuming and do not help researchers to accumulate publications and citations. As a result, social psychologists have no empirical evidence to make scientific claims about the effectiveness of police training.

Cesario et al. (2019) then claim that “a more effective way of reducing racial bias in shootings would be to eliminate racial differences in crime rates” (p. 592). Once more, the authors contradict themselves. Here they claim that there is racial bias in shootings, while their main conclusion was that that “none of these tests provided evidence of systematic anti-Black disparity” (p. 591). The authors then make some unsubstantiated claims that it would be difficult to change racial differences in crime rates. However, decriminalizing minor drug possessions would not only reduce crime rates, it would also reduce racial disparities in crime rates. Thus, Cesario et al.’s (2019) final conclusion that changes will not be easy or fast is based entirely on their own speculations and not on any scientific analysis of the problem.

In the limitation section, the authors admit that racial disparities in police shootings could be the result of racial profiling. “If officers are more likely to stop and frisk a Black citizen, for example, then officers might be more likely to enter into a deadly force situation with Black citizens independent of any actual crime rate differences across races” (p. 593). In response they point out that “the number of police shootings that start with truly discretionary stops of citizens who have not violated the law is low (*5%) and probably do not meaningfully impact the analyses” (p. 593). However, this argument is incorrect. The authors already have the problem that they are trying to prove the null-hypothesis that there is no racial bias, which is logically impossible. It is only possible to show that racial bias is small or that racial bias leads to fewer killings of US citizens. Thus, it is important to consider effect sizes and to consider the statistical power of a study to test a hypothesis. The question is whether the death of 50 US citizens that

were killed after a discretionary stop is simply irrelevant because it was just 50 citizens out of a population of over 300 million. However, the public debate about these issues does not warrant the claim that 50 lives simply don't matter. By the same logic, one would have to argue that the 50 police officers that are shot and killed each year do not matter. This may be Cesario et al.'s (2019) opinion, but it is not the opinion of the general public or politicians who found social science research, who convened a hearing about these issues to address racial biases in policing. A proper scientific investigation could have examined individual cases carefully and separated them into justified use of force and accidental use of force to examine racial disparities in both groups. The authors attempted to do so by limiting the analysis to cases where suspects were either unarmed and not aggressive or were reaching for or holding harmless objects. These statistics show a highly significant racial disparity in situations that may not have required use of deadly force, (Black victims N = 59 (40%) vs. White victims N = 88 (60%), Population statistics 20% vs. 80%, chi2(df 1, N = 147) = 37.25, p = 1/ 963,391,137. Thus, even if the percentage of citizens shot and killed by police that were unarmed and not a threat is small, the numbers show a systematic racial disparity in the killing of these civilians. This analysis is consistent with the authors own observation that they reproduced "the standard finding that odds were 3.3 times higher for unarmed Black citizens to be killed by police gunfire given population proportions" (p. 588). Thus, one of the key findings of the authors analysis is that racial disparity does not disappear when the analysis is limited to incidences that resulted in the accidental shooting of civilians who did not pose an imminent danger. Despite this important finding, the authors main conclusion is that "At the national level, we find little evidence within these data for systematic anti-Black disparity in fatal police deadly force decisions" (p. 594). This claim is extremely misleading and does not adequately characterize their findings. Their statistics show that Black

citizens are shot and killed disproportionally more than White citizens, and that this bias persists even when the analysis is limited to cases that suggest victims were shot without a real threat to officers' lives.

Benchmarking

The authors conclusion rests on their criticism of the common approach to examine racial disparities. The common approach is to examine the proportions for the outcome relative to the percentages of the racial groups in the population. For example, if we wanted to examine whether police officers disproportionally shot brown eyed people, we would compare the proportion of brown-eyed vs. blue-eyed people among those shot by police relative to the proportion of brown eyed and blue-eyed people in the population. This makes sense if we want to test the hypothesis that there is a bias because the population proportions represent the null-hypothesis that eye color is not related to police killings. If we do not take population proportions into account, we would falsely conclude that there is a bias to shoot citizens with brown eyes simply because there are more brown eyed people in the population. It is not even clear what probabilities other than population proportions should be used for the null-hypothesis. Should we simply assign equal probabilities to brown-eyed and blue-eyed people? No we should not. The reason why we condition on population proportions is that population proportions represent the null-hypothesis that there is no racial disparity.

Cesario et al.'s wrong-headed conclusions rest entirely on their claim that the common practice is wrong and that it is a flaw to take population proportions into account. "We question the traditional benchmark used to study racial disparity and describe why claims of racial bias based on this benchmark can be misleading" (p. 586). There argument is that we should not expect equal proportions of police shootings for Black and White citizens. "Adjusting for overall

population values means that one has made the decision that the relevant pool of individuals against which occurrence of the event should be compared is the entire population. This implies that we expect groups to be shot in accordance with their overall population proportions" (p. 587). They fail to state clearly that this is the null-hypothesis and not an expectation of what we actually believe about the world. After all, we often postulate a null-hypothesis to refute it. Thus, it makes perfect sense to test the null-hypothesis that a person's skin, their eye-color, or the color of their underwear has nothing to do with being shot by police. We then examine actual data to see whether the data tell us that the null-hypothesis is false, and ideally also quantify the effect size of how wrong it is. When it comes to race, the null-hypothesis is false, even when we limit cases to cases where citizens did not pose an immediate threat to officers' lives. So, we might be justified to conclude from this finding that there is racial disparity in police shootings. However, Cesario et al.' (2019) claim that we are making a mistake. I quote this section in full because their final conclusion rests on the validity of this argument.

The problem with benchmarking an outcome against population proportions is that this carries with it a critical assumption. The opportunity for the event to occur is equally likely for every person within each group. In terms of understanding racial disparities in death by police gunfire, adjusting raw shooting values by population proportions necessarily requires that White and Black citizens are equally likely to occupy situations in which deadly force is used. If this assumption does not hold, then adjusting raw fatal shooting data for overall population values is in error. (p. 587).

The authors claim that the null-hypothesis makes only sense when we assume that Black and White citizens are equally likely to encounter police. This is true, but it is not a mistake to make this assumption when we are specifying the null-hypothesis. Instead different rates of encountering police officers would be one reason why the null-hypothesis is false. Thus, it is entirely possible that there is racial disparity in police shootings and that there are different rates of encountering police, which again should be conditioned on population proportions because in terms of absolute numbers we could still see that White civilians encounter police more often. The authors are simply wrong when they conclude "One cannot experience a policing outcome without exposure to police, and if exposure rates differ across groups, then the correct benchmark is on those exposure rates" (p. 597). What the authors could have said, but what they are not saying, is that we could examine exposure rates as a potential explanation, factor or cause of racial disparities in police shootings. That is, we might find that the racial disparity in police shootings accounts largely or entirely for the racial disparity in police shootings. Instead, the authors are claiming that we should use crime rates to create a new population of criminal citizens and examine the question whether there are racial disparities in police shootings of criminals. The problem with this approach is that the authors implicitly assume that we can use the number of people killed in the population as a measure of the number of criminals that are being killed because police only shot and kill criminals. After all, the statistic is a population statistic that records all civilians that were killed by police. It is not based on a count of death among homicide suspects or any other specified group of criminals.

Cesario et al. (2019) once more confuse the null-hypothesis with the alternative hypothesis when they claim that the use of population statistics requires the assumption that "Black and Whites are involved in crime to the same extent" (p. 587). It is entirely possible that racial disparity in crime partially or fully explains racial disparities in police shootings of citizens. With the proper data, it would be possible to test a new null-hypothesis; that is, there is no racial disparity in police shootings after statistically controlling for racial disparities in crime. The key point is that the ultimate outcome of interest is still the racial disparity in the population of US citizens, not racial disparities in police shootings of a specific group of suspects (e.g., homicide arrests).

Cesario et al.'s (2019) incorrect analyses of the statistical problem leads them to a statistically wrong approach to analyze the data that produces the false conclusion that there are no racial disparities in police shootings. Namely, they argue that we can use crime statistics and condition the population frequencies of police shootings on the frequencies of crime. This approach is so wrong-headed that the authors even find evidence for a racial bias to shot White citizens up to 5 times more often than Black citizens, using their approach. I illustrate their flawed statistical approach with the Summary Report System number of arrests for murder/non-negligent manslaughter. One advantage of this statistic is that it refers to an actual population of citizens that encountered police because they were arrested. In contrast, other crimes that are committed may not result in an encounter between the criminal and a police officer because no suspect was ever found. Table 1 in the article shows that there were 4,347 Black suspects and 3,908 White suspects who were arrested in 2015. In the same year, there were 261 Black and 526 White citizens killed by police use of force.

Table 1 shows that knowledge about racial disparities in homicide rates alone does not help us to compute odds/ratios of being killed or not being killed in the population of homicide suspects or the population of citizens who were not arrested for a homicide. We are simply missing the cell frequencies of how often homicide suspects were killed during an arrest and how often citizens who were not homicide suspects were killed.

Table 1. Insufficient information to examine racial disparity for homicide arrests.

			Black			White			
	Homicide Arrest	Not.Homicide	Total	Homicide	Not.Homicide	Total			
Killed			261			526			
Not Killed			42,975,698			198,076,639	Homicide	Not.Homicide	Population
Total	4,347	42,971,612	42,975,959	3,908	198,073,257	198,077,165	OR(B/W)	OR(B/W)	OR(B/W)
Odds (K/NK)	?	?	6.07E-06	?	?	2.66E-06	?	?	2.29

Cesario et al.'s ill-fated benchmarking approach sidesteps this problem by simply assuming that all of the police shootings occurred in the group of the homicide suspects. This assumption is sufficient to fill in all cell frequencies because a 2 x 2 table has only one degree of freedom.

Table 2. Cesario et al.'s hypothetical scenario where all shootings occur in the context of arresting homicide suspects.

			Black			White			
	Homicide Arrest	Not.Homicide	Total	Homicide	Not.Homicide	Total			
Killed	261	0	261	526	0	526			
Not Killed	4,086	42,971,612	42,975,698	3,821	198,072,818	198,076,639	Homicide	Not.Homicide	Population
Total	4,347	42,971,612	42,975,959	4,347	198,072,818	198,077,165	OR(B/W)	OR(B/W)	OR(B/W)
Odds (K/NK)	6.39E-02	NA	6.07E-06	1.38E-01	NA	2.66E-06	0.46	NA	2.29

Based on Cearios et al.'s (2019) benchmarking approach, we now conclude that the racial disparity in police shootings in the population is fully explained by racial disparities in the rate of

homicide arrests. Moreover, the results now even show the opposite racial bias. Controlling for homicide rates, Black homicide suspects are less likely to be killed by police than White homicide suspects. The same result is obtained with other crime statistics. Whenever crime statistics are used as benchmark, police officers seem to have a bias to kill disproportionally more White suspects. Before we even start to come up with explanation for this surprising finding, it is important to emphasize that this "finding" is entirely based on the assumptions. Once we change the assumptions, the conclusions can change and Cesario et al.'s data are entirely consistent with the hypothesis that police officers are more likely to use lethal force when they encounter a Black citizen. This is shown in Table 3, which assumes that some of the victims who were killed by police were not homicide suspects.

Table 3. Hypothetical scenario with racial disparity for homicide arrests and other encounters.

		Black			White				
	Homicide Arrest	Not.Homicide	Total	Homicide	Not.Homicide	Total			
Killed	151	110	261	126	400	526			
Not Killed	4,196	42,971,502	42,975,698	3,782	198,072,857	198,076,639	Homicide	Not.Homicide	Population
Total	4,347	42,971,612	42,975,959	3,908	198,073,257	198,077,165	OR(B/W)	OR(B/W)	OR(B/W)
Odds (K/NK)	3.60E-02	2.56E-06	6.07E-06	3.33E-02	2.02E-06	2.66E-06	1.08	1.27	2.29

I am not claiming that the scenario in Table 3 is realistic or true, which is not necessary to criticize Cesario et al.'s benchmarking approach. I am simply demonstrating that their benchmarking approach relies on an assumption and that their conclusions change when we change the assumptions. To make an empirical contribution, the authors would need to examine the actual circumstances that led to a police shooting.

In conclusion, Cesario et al. (2019) introduced a new statistical approach to analyzing racial disparities that they call benchmarking. Closer inspection of this approach reveals that it is

fundamentally flawed because it relies on assumptions rather than on empirical data. The benchmarking terminology makes it appear as if we no longer need to think about population proportions, but this is a mistake. Crime rates or other statistics might be potential mediators that explain racial disparities in the population, they do not make racial disparities in the population disappear or irrelevant. It is therefore misleading when the authors repeatedly claim that their analyses "reveal no evidence of systematic anti-Black disparity in police fatal shootings at the national level" (p. 586), that "Whites were either more likely to be fatally shot by police or police showed no significant disparity in either direction" (p. 588), "did not yield strong support for any substantial anti-Black disparity" (p. 589), "revealed little disparity" (p. 589), "provide little evidence of systematic anti-Black disparity in officers' decisions to shoot unarmed, nonaggressing citizens" (p. 589), "Officers either showed no meaningful disparity in either direction or, if anything, an overall pattern of anti-White disparity" (p. 589), "none of the benchmarks on crime rates revealed consistent anti-Black disparity in being fatally shot while reaching for/holding a harmless object" (p. 589), "Whites had higher odds of being killed by police gunfire than Blacks" (p. 589), "did not yield consistent anti-Black disparity" (p. 589), "no evidence of systematic anti-Black disparity in fatal police shootings when those decisions are in response to the misidentification of a harmless object or movement by the citizen" (p. 590), "only 1 of the 144 possible tests (0.7%) showed statistically significant anti-Black disparity" (p. 590), "none of these tests provided evidence of systematic anti-Black disparity" (p. 591), "it is clear that systematic anti-Black disparity at the national level is not observed" (p. 591), "we find little evidence within these data for systematic anti-Black disparity in fatal police deadly force decisions" (p. 594). All of these claims hinge on the problematic approach that changes the population from the population of citizens to a population of criminals and then assuming that all

shootings occurred within the context of this crime. The authors even apply this approach to shootings where individuals were not armed without justifying why crime statistics should be relevant here. Why should unarmed Black citizens be more likely to be shot, if there are proportionally more Black homicide suspects?

Thus, the main conclusion remains that when we take into account that the use of lethal force is often justified and proportionally kills more Black than White criminals, we are left with the observation that racial disparities in police shootings persist. “When looking at the raw data for shootings involving object misidentification, there were 26 Whites and 19 Blacks killed. Given population proportions, odds were 3.7 times higher for Blacks to be fatally shot while holding/reaching for a harmless object” (p. 592). This is the only empirical contribution in the article. While racial bias in police shootings may be explained to a considerable amount by racial disparities in crime rates, which the authors assume, but do not test, they do show that racial bias persists and is nominally stronger when unarmed citizens are shot. Thus, the answer to the authors own question in their title “Is There Evidence of Racial Disparity in Police Use of Deadly Force?” is “Yes, there is racial disparity and racial differences in crime and racial bias in policing are likely to contribute to the overall racial bias in police shootings at the national level. This conclusion is entirely consistent with the empirical data provided in the article. It is just not clear why the authors were unable to draw the right conclusion form their data, and what made them invent a flawed new statistical approach that led to the wrong conclusion that there are no racial disparities in police shootings. Politeness prevents me from speculating about the causes of the authors’ biased reasoning about police shootings.