

To: Michaela Martini
Executive Office of Public Safety and Security
Michaela.Martini1@mass.gov

From: Katy Naples-Mitchell
Charles Hamilton Houston Institute for Race & Justice
Harvard Law School
knaplesmitchell@law.harvard.edu

March 2, 2022

Public Comment in Response to the 2020 Massachusetts Uniform Citation Data Analysis Report

Background

The **Charles Hamilton Houston Institute for Race and Justice** (“Houston Institute”) at Harvard Law School was launched in 2005 by Charles J. Ogletree, Jr., Jesse Climenko Professor of Law. The Institute is presently led by Faculty Director Prof. Guy-Uriel Charles, the Charles J. Ogletree, Jr., Professor of Law. The Institute honors and continues the work of Charles Hamilton Houston, who engineered the multi-year legal strategy that led to the unanimous 1954 Supreme Court decision, *Brown v. Board of Education*. The Houston Institute’s long-term goal is to ensure that every member of our society enjoys equal access to the opportunities, responsibilities, and privileges of membership in the United States. The Houston Institute advocated for the inclusion of data collection provisions in the Hands-Free Driving Law in 2019,¹ and has raised concern about racially disparate traffic enforcement in amicus briefs to the Commonwealth’s Supreme Judicial Court in cases including *Commonwealth v. Buckley*, *Commonwealth v. Long*, *Commonwealth v. Garner*, *Commonwealth v. Daveiga*, and others. The availability of these data to the public is essential to develop public policy aimed at disrupting structural and systemic racism. In a public policy analysis of stop data, research focused narrowly on proving the existence of racial profiling with rigorous statistical methods is only one piece of what’s needed. The Commonwealth should engage in data analysis that raises questions about whether our law and its enforcement is creating inequity across racial groups in any number of ways—including racial profiling at the officer, unit, or department level *as well as* other possible vectors which embed racism.

Discussion

I. EOPSS Should Report on Overall Disparities in Traffic Stops.

The study should have included an answer to research question 2: a statewide and summary analysis of racial disparities in stops based on driving population benchmarks.

By its own terms, “The overall goal of this study is to learn more about potential patterns of racial disparities in traffic stops for the purposes of understanding the causes of these

¹ David J. Harris, Letter, *We need bias-free policing as much as hands-free driving*, BOS. GLOBE (Oct. 2, 2019), <https://www.bostonglobe.com/opinion/letters/2019/10/01/need-bias-free-policing-much-hands-free-driving/RJxXEOQDNARkwmPdRt614O/story.html>.

disparities.”² This asserted goal comports with the four research questions the study identifies on page 16:

1. What is the general pattern of traffic enforcement in Massachusetts?
2. Are Non-White motorists stopped more often than their representation in the driving population would predict?
3. Are Non-White motorists more likely to receive a warning, citation, or arrest than White motorists?
4. Are Non-White motorists more likely to be subject to a discretionary, non-inventory search than White motorists?³

Before reaching questions of statistical modeling of *racial profiling* (a possible *cause* of racial disparities in traffic stops), the study’s own goal and research inquiry is to document and learn more about patterns of *racial disparities* in traffic stops based on the “driving population.”

Unfortunately, the study does not answer that question (“Are Non-White motorists stopped more often than their representation in the driving population would predict?”), even though it could. A study attempting to learn more about potential patterns of racial disparities in traffic stops based on the driving population should begin by identifying those disparities. Here, that question could be answered by establishing a driving population estimate for the state or using as a benchmark the racial demographics of the statewide, driving-eligible population based on census data (given that 92% of stops were of Massachusetts drivers).⁴ The study’s findings section did

² MASS. EXECUTIVE OFF. OF PUB. SAFETY & SECURITY, 2020 MASSACHUSETTS UNIFORM CITATION DATA ANALYSIS REPORT at 7-8 (2022), <https://www.mass.gov/doc/2020-massachusetts-uniform-citation-data-analysis-report/download> [hereinafter “EOPSS Study”]; *see also id.* at 14, 41.

³ These questions differ from the 2004 Northeastern Farrell report, which also answered a series of four questions: (1) whether nonwhite drivers who were residents in a community were cited more often than their representation in the residential population would suggest; (2) whether nonwhite drivers over-all were cited more often than their representation in the population of people driving on the roadways would predict; (3) whether, once stopped, nonwhite drivers were more likely to be searched than white drivers; and (4) whether, once stopped, non-white drivers were more likely to receive a citation than white drivers. *Commonwealth v. Lora*, 451 Mass. 425, 448-449 (2008) (Ireland, J., concurring) (citing Farrell, Massachusetts Racial and Gender Profiling Study: Final Report, Inst. on Race and Justice of Northeastern University, at 2 (2004)).

⁴ More precise benchmarking might control for vehicle ownership rates or vehicle access. Not all people drive, so the percentage of a particular racial group among the resident population may be a poor proxy for the at-risk population of drivers from that racial group. Other researchers attempt to control for in-flow and out-flow to a particular jurisdiction by considering commuting populations and the demographics of jurisdictions in surrounding areas, or estimating how establishments and institutions like universities, museums, malls, movie theaters, and restaurants might affect the demographics of who is on the relevant roadways. Attempts at improving benchmark analysis focus on refining measures of the at-risk population of drivers.

not include any statewide driving population benchmark despite explicitly asking that research question, nor did it summarize the individual jurisdictions analyzed—unlike the Connecticut model (Barone 2020), which is cited approvingly.

Why does this matter? The findings section of this study reports on only one methodology for disparities in who is pulled over, aimed at determining one particular cause of disparity (direct, non-contextual racial profiling in the decision to stop, based on examining differences in rates of stops in daylight and darkness as a proxy for driver visibility). The flawed assumptions behind this methodology are discussed in greater detail *infra*. But within-racial-group comparisons of stops in daylight versus at night gloss over the underlying overrepresentation of people of color *across all traffic stops*, a figure omitted from discussion in the study’s findings section. “If ‘systemic racism’ is defined as a ‘system[or] institution[] that produce[s] racially disparate outcomes, regardless of the intentions of the people who work within [it],’ then our criminal justice system is rife with it.” *Commonwealth v. Long*, 485 Mass. 711, 740 (2020) (Budd, J., concurring). If we follow the definition of systemic racism used by the Chief Justice of our own Supreme Judicial Court, as a matter of public policy the Commonwealth should be concerned with documenting racial disparities overall.

Although this kind of benchmarking analysis has been critiqued by statisticians in the field as an outdated and imprecise method to explore racial profiling,⁵ a commitment to racial justice and to anti-racist public policy demands this kind of baseline analysis. As Chief Justice Budd has explained, “Years of data bear out what many have long known from experience: police stop drivers of color disproportionately more often than Caucasian drivers for insignificant violations (or provide no reason at all). In 2017, the Stanford Open Policing Project found that police stopped African-American drivers more than Caucasian drivers, controlling for population makeup, both nationally and in Massachusetts.” *Commonwealth v. Buckley*, 478 Mass. 861, 877-878 (2018) (Budd, J., concurring). A more recent study by the Vera Institute of Justice using Suffolk County data confirmed that 35% of traffic stops in Suffolk County between 2010 and 2019—some 200,000 stops—were for non-public-safety traffic violations, and Black people were pulled over by police for these low-level traffic violations at over twice the rate of their white counterparts.⁶

Acknowledging this over-representation is important because it gives voice, in data, to the lived experiences of people of color—and the dignitary harm and risks to physical safety these kinds of minor, often discretionary traffic stops can produce.⁷ It is certainly true that population

⁵ See, e.g., Jake Kara, *Does peer review cast doubt on traffic-stop analysis?*, CT. MIRROR (Nov. 20, 2017), <https://ctmirror.org/2017/11/20/does-peer-review-cast-doubt-on-traffic-stop-analysis>. The Peer Reviews discussed in this article are available at: <https://www.documentcloud.org/documents/4185596-Peer-Reviews-Copy.html>.

⁶ Selekee Flingai, Opinion, *Racism in traffic stops is real, and the data backs it up*, BOS. HERALD (Feb. 25, 2022, 12:10 AM), <https://www.bostonherald.com/2022/02/25/flingai-racism-in-traffic-stops-is-real-and-the-data-backs-it-up>.

⁷ David D. Kirkpatrick, Steve Eder, Kim Barker & Julie Tate, *Why Many Police Traffic Stops Turn Deadly*, N.Y. TIMES (Oct. 31, 2021), <https://www.nytimes.com/2021/10/31/us/police-traffic-stops-killings.html>.

benchmarking alone may “prove[] little, because other factors besides race could be in play. Because African-Americans are, for example, generally poorer than whites, they may have more expired vehicle registrations or other automotive lapses that attract officers’ attention.”⁸ But the authors could sensitively address these potential confounding variables in discussion (and, if data were available, in statistical modeling), instead of omitting this figure from analysis entirely. And this analysis could lead to important directions for public policy—including eliminating the financial penalties and socio-economic barriers that may enable apparently racially disparate enforcement of minor violations. Analyses like these may not prove racial profiling, but they can inspire valuable new directions for future research *and* for public policy. The study’s failure to answer its own research question using available data, or to provide a summary analysis in the findings section of the individual jurisdiction analyses based on the benchmarking methodologies, is an oversight that should be corrected in an amended version of this year’s study and in future versions for subsequent years of data.

Request for corrective action:

- Will the study authors update the findings section to answer research question 2 (“Are Non-White motorists stopped more often than their representation in the driving population would predict?”) with analysis of a statewide benchmark?
- Will the study authors update the findings section to summarize the individual jurisdiction analyses of the two population-benchmark-based methodologies?

II. Use of the racial categories “White” and “Non-White” fails to account for well-documented profiling of Black drivers and obscures more precise analyses.

For many of the methodologies applied, including the statewide Veil of Darkness (VoD) analyses, the study limited its analysis and discussion to differences between and among “White” and “Non-White” drivers. This is not sufficiently explained or justified in the study. The study merely says, “For some of the analyses, the race/ethnicity categories were further collapsed into Non-White and White.” (p. 17) Was this a function of the “n”—of small sample sizes? Was there some other reason for doing this? The study authors should have explained this decision for each analysis where they declined to apply a more precise racial breakdown.

Using the racial categories of “White” and “Non-White” is a significant deviation from other similar Veil of Darkness studies, including Grogger & Ridgeway (2006) and Barone (2020) which were cited as models by the EOPSS study authors. Those studies more precisely looked at racial profiling of Black drivers. This is appropriate: researchers have addressed the risks of “Driving While Black” in academic scholarship for decades across a variety of disciplines.⁹

⁸ Sharon LaFraniere & Andrew W. Lehren, *The Disproportionate Risks of Driving While Black*, N.Y. TIMES (Oct. 24, 2015), <https://www.nytimes.com/2015/10/25/us/racial-disparity-traffic-stops-driving-black.html>.

⁹ See, e.g., David A. Harris, *Driving While Black and All Other Traffic Offenses: The Supreme Court and Pretextual Traffic Stops*, 87 J. CRIM. L. & CRIMINOLOGY 544 (1997); Richard J. Lundman & Robert L. Kaufman, *Driving While Black: Effects of Race, Ethnicity, and Gender on Citizen Self-Reports of Traffic Stops and Police Actions*, 41 CRIMINOLOGY 195 (2003); Patricia

Patterns of policing are not consistent across all racial groups that fall under the umbrella of “Non-White,” and limiting racial categories to “White” and “Non-White” (1) implicitly defines whiteness as normative and (2) may obscure potential differences in how Black, Hispanic, Asian, South Asian, Middle Eastern, Native American, and Pacific Islander motorists are observed, treated, and policed.

In particular, in the findings for the Veil of Darkness analysis of the statewide inter-twilight period stops, the EOPSS study authors found that Non-White drivers were stopped at slightly higher rates in daylight than in darkness, but found no statistical significance using either regression or chi square tests. (EOPSS Study at 10, 31-32.) Readers are left to wonder whether that pattern would hold for Black drivers—perhaps Black motorists would have been found to be stopped at higher rates in daylight than darkness *with* statistical significance if considered in isolation.

Request for corrective action:

- Will the study authors produce statewide VoD analyses with greater specificity, considering at least Black drivers and Hispanic drivers in isolation instead of collapsing them, with other racial and ethnic groups, under the umbrella of “Non-White” drivers?
- Will the study authors explain why they used the chosen racial categories for each analysis?

III. The Veil of Darkness Methodology Overlooks Limitations that, When Controlled, Have Been Shown to Question or Alter its Findings.

Developed in 2006, the Veil of Darkness (VoD) methodology is considered to be a statistically robust method for examining whether racial differences in traffic stops are attributable to racial profiling by the police. This racial profiling test is based on differences in driver race visibility (using daylight as a proxy for visibility) and the within-race distribution of traffic stops across daylight and darkness. Prior to the development of this method, the field generally used benchmarking approaches, and some researchers continue to use benchmarking analyses in addition to applying the VoD methodology (for example, Barone’s work in Connecticut). Constructing an appropriate benchmark is difficult and may be imprecise. *See* note 4, *supra*.

To avoid the difficulties of benchmarking, the VoD methodology uses the presence of natural light as a proxy for whether officers can observe the race of the driver and groups stops

Warren et al., *Driving While Black: Bias Processes and Racial Disparity in Police Stops*, 44 CRIMINOLOGY 709 (2006); Timothy Bates, *Driving While Black in Suburban Detroit*, 7 DU BOIS REV. 133 (2010); CHARLES R. EPP ET AL., PULLED OVER: HOW POLICE STOPS DEFINE RACE AND CITIZENSHIP (2014); Stephanie Seguino & Nancy Brooks, *Driving While Black and Brown in Vermont: Can Race Data Analysis Contribute to Reform?*, 48 REV. BLACK POL. ECON. 42 (2021). *See also* David A. Harris, *The Stories, the Statistics, and the Law: Why "Driving While Black" Matters*, 84 MINN. L. REV. 265 (1999); DAVID A. HARRIS, AM. CIVIL LIBERTIES UNION, DRIVING WHILE BLACK: RACIAL PROFILING ON OUR NATION’S HIGHWAYS (1999), <https://www.aclu.org/report/driving-while-black-racial-profiling-our-nations-highways>.

that happened at the same time of day—in daylight during some periods of the year, and in darkness during other periods of the year. The methodology assumes that officers can see a driver’s race more easily in daylight than in darkness. Among the limited available data from self-reported findings, there is some support for that conclusion—but also evidence that in the vast majority of stops, officers self-report that they cannot observe driver race *at all* prior to effectuating stops. In a 2003 study of traffic stop data from Minnesota, researchers analyzed officer self-reported data required to be collected by law concerning whether the officer knew the driver’s race/ethnicity prior to the stop; officers had to check a box indicating ‘yes’ or ‘no.’ On average, officers reported higher “yes” rates (19.4% on average) during daylight than during darkness (9.7% on average). However, these data suggest that, overall, officers could not perceive driver race in the vast majority of stops.¹⁰

As other researchers have shown with more precise modeling, the assumption that natural light in daytime allows officers to perceive a driver’s race more readily than in darkness may be flawed—particularly in well-lit urban areas, and if the methodology fails to take into account weather patterns or other confounds that influence the presence of light, or the particularly pronounced within-season variation of light surrounding Daylight Savings Time.¹¹ While the EOPSS study acknowledges some of these limitations (*see* EOPSS Study at 38), it does not acknowledge that controlling for these variables can reverse researchers’ results.

A 2020 study by Vito et al. determined “a potential fallacy of the veil-of-darkness (VOD) hypothesis.”¹² The researchers noted that the VoD method “fails to account for other factors of officer decision making within racially motivated stops.” *Id.* These researchers concluded that, when a VoD analysis yields no differences in rates of drivers’ race across day and night, that is not necessarily conclusive of an absence of racial profiling. By contrast, “studies whose results lead to conclusions of racial profiling within the VOD hypothesis are of particular importance, as these studies are evident in brazen racial profiling within the agency(ies) sampled.” *Id.*

¹⁰ INSTITUTE ON RACE AND POVERTY & COUNCIL ON CRIMINAL JUSTICE, MINNESOTA STATEWIDE RACIAL PROFILING REPORT: ALL PARTICIPATING JURISDICTIONS at 37 (2003), https://scholarship.law.umn.edu/cgi/viewcontent.cgi?article=1113&context=imo_studies. In a lengthy discussion, this study acknowledges the limitations of self-reported data and the potential that officers who engage in profiling are particularly likely to underreport whether they could observe the race of the driver.

¹¹ *Cf.* Emma Pierson et al., *A large-scale analysis of racial disparities in police stops across the United States*, 4 NATURE HUM. BEHAVIOUR 736, 737 (July 2020), <https://www.nature.com/articles/s41562-020-0858-1.pdf> (“[D]arkness—after adjusting for time of day—is a function of the date. As such, to the extent that driver behaviour changes throughout the year, and that these changes are correlated with race, the test can suggest discrimination where there is none. To account for these potential seasonal effects, we applied a more robust variant of the veil-of-darkness test that restricts to two 60-day windows centred on the beginning and end of DST.”).

¹² Anthony G. Vito, Vanessa Woodward Griffin, Gennaro F. Vito & George E. Higgins, “*Does daylight matter*”? *An examination of racial bias in traffic stops by police*, 43 POLICING: AN INT’L J. 675, 683-684 (2020).

The Veil of Darkness methodology utilized by the EOPSS study does not take into account the presence of artificial lighting sources at night, nor does it measure in what percentage of stops an officer did *actually* observe a driver’s race – which would require self-reported data. Other studies have tried to adopt methodological approaches to adjust for these limitations, with results that raise important questions about the validity of the VoD methodology.

A. Accounting for artificial light sources in the Veil of Darkness may reverse the conclusion of no statistical support for racial profiling.

Horrace and Rohlin (2016) offered a useful refinement of VoD and critique of the assumptions underlying VoD.¹³ The authors hypothesized that urban environments may be well-lit at night, eroding the power of Grogger and Ridgeway’s VoD test. Horrace and Rohlin controlled for ambient light at night by mapping geographic data about streetlights in Syracuse onto their dataset of traffic stops in Syracuse between 2006 and 2009, limiting the sample to stop locations with lower levels of night-time ambient light, so that the darkness variable would be stronger. They found in this analysis of Syracuse stops that ignoring streetlights often leads to conclusions of no racial profiling, but their refined test leads to conclusions of racial profiling of Black drivers. The results changed in the direction of finding profiling of Black drivers when the presence of a streetlight near a stop was controlled.

This kind of analysis was not possible in the EOPSS study because the RMV dataset the study authors had access to only provided the zip code of where the stop occurred. However, it offers a particularly apt comparison given that the authors of the EOPSS study noted that the “statewide average is highly influenced by the stop data from the larger cities as larger cities make up the largest volume of stops overall.” (EOPSS Study at 39.) Where the Veil of Darkness methodology used in the EOPSS study does not account for artificial light sources—particularly in cities—its methodology may be flawed and may underestimate racial profiling.

B. The Veil of Darkness fails to account for whether officers can *actually* perceive a driver’s race.

Grogger and Ridgeway’s initial 2006 Veil of Darkness study was conducted with data from Oakland, California, and by its terms it found no evidence of racial profiling of Black drivers. A more recent study of Oakland police data from 2013-2014 suggests that significant disparities in overall stops persist using population benchmarking and that Grogger and Ridgeway’s conclusion of no racial profiling is undermined by officer self-reported data.¹⁴ In this 2016 Stanford study, in

¹³ William C. Horrace & Shawn M. Rohlin, *How Dark Is Dark? Bright Lights, Big City, Racial Profiling*, 98 REV. OF ECON. & STAT. 226 (2016).

¹⁴ REBECCA C. HETHEY, BENOÎT MONIN, AMRITA MAITREYI & JENNIFER L. EBERHARDT, DATA FOR CHANGE: A STATISTICAL ANALYSIS OF POLICE STOPS, SEARCHES, HANDCUFFINGS, AND ARRESTS IN OAKLAND, CALIF., 2013-2014 at 77 (2016), <https://stanford.app.box.com/v/Data-for-Change>. See also Clifton B. Parker, *Stanford big data study finds racial disparities in Oakland, Calif., police behavior, offers solutions*, STANFORD NEWS (June 15, 2016), <https://news.stanford.edu/2016/06/15/stanford-big-data-study-finds-racial-disparities-oakland-calif-police-behavior-offers-solutions>.

addition to population benchmarks showing evidence of pronounced racial disparities, the authors conducted select analyses comparing differences in stop rates by race as a function of whether or not the officer could determine the person's race prior to making the stop; officers were required to self-report on their stop forms whether or not they knew the race of the driver prior to the stop. The researchers found that, "among stops made in which the race of the person was unknown, 48% of those stopped were African American. In contrast, among stops made in which the race of the person was known, 62% of those stopped were African American."¹⁵ Using a simple paired t-test, the authors found a significant difference in the percentages of African Americans stopped among these two types of stops.

Grogger and Ridgeway critique benchmarking efforts, and thus may not be disturbed that more than a decade after their study, analyses of Oakland police data, including the 2016 Stanford study, continue to find that Black residents of Oakland are stopped in substantial disproportionality to their population representation—a finding that continues to alarm Black community members, who intimately understand the harms of living amidst this kind of policing.¹⁶ But the finding by Stanford researchers Hetey, Monin, Maitreyi & Eberhardt that when officers self-reported they knew the race of the driver, they were more likely to have pulled over African-Americans, does potentially suggest racial profiling and therefore casts doubt on whether the Veil of Darkness methodology's use of the presence of natural light is a satisfactory proxy for officer ability to detect a driver's race.

C. The Use of Mobile Data Terminals by all Massachusetts police departments undermines one of the central assumptions of the Veil of Darkness methodology—that the presence of natural light enables differential perception of driver race.

The Veil of Darkness methodology assumes that the amount of natural light in an environment translates to differentials in officers being able to visibly perceive a driver's race—and to be able to racially profile a driver on the basis of their perceived race. The methodology does not account for how officers may make assumptions about a driver's race based on external information unaffected by natural light.

For more than a decade, Massachusetts police departments have equipped officers with mobile data terminals—computers inside their vehicles that can access a database that includes information about the race of the vehicle owner and a photograph of them.¹⁷ See *Commonwealth v. Muckle*, 61 Mass. App. Ct. 678, 679 n.3 (2004) ("An officer inputs license plate numbers, and, within seconds, receives various information in reply, including the type of vehicle that is assigned to the plate, whether the vehicle is registered, whether the vehicle owner has an active license, and

¹⁵ HETHEY ET AL., *supra* note 14, at 77.

¹⁶ Stephanie Sierra, Lindsey Feingold & Yun Choi, *Black men are 8 times more likely to be stopped by Oakland police than White men, data shows*, ABC NEWS 7 (Sept. 9, 2020), <https://abc7news.com/oakland-police-opd-racial-profiling-traffic-stops/6414305>.

¹⁷ *Massachusetts Unveils New Crime Fighting Database*, GOV'T TECH. (July 27, 2010), <https://www.govtech.com/public-safety/massachusetts-unveils-new-crime-fighting-database.html> ("By the end of June, every police department and law enforcement agency in Massachusetts will be able to electronically access information on file at the RMV.").

whether any warrants are outstanding.”). This fundamentally disrupts a central assumption of the veil of darkness—that the ability to identify the race of the driver is contingent on natural light availability. Take, for example, the facts of *Commonwealth v. Long*. The decision opens:

At about eleven o’clock on a November morning, two members of the Boston police department’s youth violence strike force, who had been driving an unmarked vehicle, noticed a maroon Mercedes pass in front of them on a residential street. The driver was a Black man. The officers decided to query the vehicle’s license plate in their onboard computer. The results returned indicated that the vehicle was registered to a Black woman and that it lacked an inspection sticker. The officers stopped the vehicle.

Commonwealth v. Long, 485 Mass. 711, 712 (2020). The officers in *Commonwealth v. Long* testified that they could not see and therefore did not know the race of the driver before pulling him over. But the car was registered to a Black woman, and that was apparent to the officers before they pulled over the car because they had queried the vehicle’s plate. Two aspects of this anecdotal example are illustrative for understanding why the VoD methodology is at best incomplete and at worst unhelpful—obscuring the reality of policing:

- This stop happened at 11 AM—in broad daylight, on a residential street (not on an interstate, so the car’s speed was not an issue for visibility), and while the officers’ unmarked car was stopped and positioned perpendicular to the driver, and the car passed in front of them. In these circumstances, the Veil of Darkness presumes a driver’s race is most visible. But the officers testified it was not.
- The officers relied on database information, which they queried from their mobile data terminals, prior to effectuating the stop. The race of the registered owner, who U.S. Supreme Court precedent permits officers to assume is the driver, *Kansas v. Glover*, 140 S. Ct. 1183 (2020), was apparent to the officers—and would have been apparent to the officers regardless of the presence or absence of natural light.

The fact that the police run these queries before pulling people over means they may know someone’s race based on the photo or demographic information associated with the vehicle registration, regardless of whether they actually see the driver behind the wheel or any passengers. The VoD approach was developed before there was widespread use of mobile data terminals that display the car owner’s picture. Seeing a picture of the vehicle’s registered owner before deciding to stop the car provides an external information source unaffected by changes in natural lighting that may allow race to influence the decision to stop a car. In other words, the use of these mobile data terminals casts doubt on the validity of the VoD test by undermining its fundamental assumption: that police officers can perceive a motorist’s race at differential rates in daylight and in darkness. It is certainly possible that police officers are less able to read a license plate in darkness, and so their use of mobile data terminals prior to effectuating stops may be concentrated more in daylight hours, which would be consistent with the VoD’s assumptions. But without testing this assumption, and accounting for the role of other ways of perceiving a driver’s race, the VoD analysis is incomplete.

D. Driving habits and policing patterns may change due to the presence of daylight or darkness in a manner not accounted for by the standard VoD analysis.

Driving patterns may differ between light and dark, and might not vary in the same way for different racial and ethnic groups, but it is difficult to document these patterns. Similarly, different policing patterns during the night and day might mean different responses to these different driving patterns. And indeed, these patterns may influence one another. Drivers concerned about being racially profiled during daylight may adjust their behavior to prevent law enforcement contact—for example, speeding less in daylight hours, when they can be seen *and* when they might be able to see officers approaching.

One study of Massachusetts data tried to account for these factors and developed a follow-up test for identifying the direction of differential treatment by examining the speed distribution of motorists across daylight and darkness.¹⁸ The study used data on traffic stops in Massachusetts made by state and local police—but limited to stop populations containing a reasonable number of African Americans; accordingly, the authors restricted their analysis to 2001-2003 state police stops and local stops made by town police departments of the 10 largest jurisdictions (Boston, Worcester, Springfield, Lowell, Cambridge, Brockton, New Bedford, Quincy, Lynn, and Newton).¹⁹

First, they conducted a classic VoD analysis, following the model of Grogger and Ridgeway (2006). In Massachusetts, they found evidence suggesting that the odds that stops involve a Black motorist increases in daylight relative to darkness. A daylight stop in Massachusetts was approximately 4.5 percentage points more likely to involve an African-American motorist, indicating that these Massachusetts departments showed evidence of discriminatory policing. Then, they explored their hypothesis that the speed of stopped motorists of color would decrease in daylight in response to real or perceived discrimination at higher percentiles of the speed distribution. They calculated a relative speed, and found that African-American motorists in the upper half of the speed distribution travel more slowly in daylight, when presumably race is observed. In the Massachusetts sample (they also considered data from Tennessee separately), they observed “a large shift in the speed distribution of African-Americans between daylight and darkness near the top of the distribution for Massachusetts, 7 to 12 percent slower in daylight relative to the speed limit.” *Id.* at 25-26. The authors concluded, “Our empirical findings . . . provide strong evidence suggesting that minority motorists are adjusting their driving behavior during daylight, when their race is observable to police, in response to real or perceived police discrimination.” *Id.*

This study raises a number of questions about the EOPSS study. Although the data used for Massachusetts by these researchers was 20 years old, it again raises the need to conduct a more precise racial breakdown for the VoD analyses, looking at Black drivers in isolation. It also

¹⁸ Jesse Kalinowski, Matthew B. Ross & Stephen L. Ross, *Endogenous Driving Behavior in Tests of Racial Profiling* (Nat'l Bureau of Econ. Research, Working Paper No. 28789, May 2021), https://www.nber.org/system/files/working_papers/w28789/w28789.pdf.

¹⁹ The study also conducted a separate analysis of national crash data, not discussed here.

suggests that factors about how driving behavior may objectively differ in daylight and darkness should be considered and controlled in the VoD analysis.

IV. The EOPSS press release inaccurately summarized the study’s Veil of Darkness findings—and focused on them to the exclusion of other important findings about statistically significant differences in racial disparities in post-stop outcomes.

The EOPSS press release about the study said that “Researchers applied the ‘Veil of Darkness’ (VoD) analysis to study the racial distribution of traffic stops.”²⁰ But the Veil of Darkness is a within-race analysis—it is not a method of benchmarking that examines the racial distribution of traffic stops. The EOPSS press release also only summarized the statewide figures from *all stops*, not stops limited to the inter-twilight period. The all-stops figures were more favorable for the Commonwealth’s public image, but less related to the core of the VoD method, which by design focuses on the inter-twilight period. EOPSS did not mention in its press release that, although not statistically significant, “The results of this test for only the ITP stops show much more similar *within race* percentages of stops during daylight as compared to darkness, with Non-White motorists being slightly more likely to be stopped in daylight as compared to darkness and White motorists being slightly more likely to be stopped in darkness than in daylight.” (EOPSS Study at 10, 32.) This omission is material.

Further, and more importantly, the EOPSS press release did not summarize *any* of the findings about post-stop outcomes, which showed significant evidence of disparities in the treatment of Black and Hispanic drivers as compared to White drivers. These post-stop statistical analyses are more robust than the methodologies for assessing the role of racial profiling in the decision to stop because there is no benchmarking required; the researchers were able to draw conclusions based on the limited pool of traffic stops, without having to construct a benchmark for how many drivers *could have* been stopped or were on the road. The failure to summarize any of the post-stop outcomes is a significant omission in the EOPSS press release that has led to confusion in the public reporting about this study—and has muddied the report’s significant findings of racially disparate treatment for drivers of color.²¹

V. Other Questions and Concerns About Methodology, Data Coding, and Data Quality

- The study does not reproduce the regression formula used by the study authors. Other similar studies do. Can the regression formula be included in an amended appendix?

²⁰ Press Release, Exec. Office of Pub. Safety & Security, Executive Office of Public Safety and Security Releases Baseline Analysis of Police Traffic Stop Data (Feb. 7, 2022), <https://www.mass.gov/news/executive-office-of-public-safety-and-security-releases-baseline-analysis-of-police-traffic-stop-data>.

²¹ See, e.g., Mike Beaudet, *More evidence of racial bias seen in Massachusetts police stops than state news release suggests*, WCVB (Feb. 9, 2022), <https://www.wcvb.com/article/5-investigates-racial-bias-police-stops-massachusetts/39027992#>. Tonya Alanez & Danny McDonald, *For some, report on Mass. traffic stops shows stubborn racial biases persist in policing*, BOS. GLOBE (Feb. 8, 2022), <https://www.bostonglobe.com/2022/02/08/metro/some-report-mass-traffic-stops-shows-stubborn-racial-biases-persists-policing>.

- This would be important to assess to what extent the regression follows the model and includes all the variables used by Grogger and Ridgeway (2006) and other subsequent VoD studies that have more precise controls. For example, it is unclear whether the regression controls for clock time or darkness level; the time of the stop is listed as a variable available in the dataset on page 17 of the EOPSS Study, but not a variable explicitly considered in the regression on page 22.
- The Grogger and Ridgeway (2006) VoD study limited analysis to the dusk inter-twilight period (ITP) given the small number of stops in the morning ITP. Was this also an issue for the Massachusetts citation data? Did the EOPSS researchers consider limiting their ITP VoD analysis to the evening (dusk) ITP, instead of a combined dusk and dawn analysis? Could the researchers produce a version of the statewide VoD analysis limited to Black drivers in the evening (dusk) ITP, to be comparable to findings of the Grogger and Ridgeway study?
- For the individual jurisdiction comparative analyses, the study is inconsistent about whether it compares the demographics of *stopped motorists* by that jurisdiction’s police agency to the jurisdiction’s demographics (as described on p. 45), versus the demographics of *stopped *resident* motorists* by that jurisdiction’s police agency to the jurisdiction’s demographics (as described on p. 23). Based on the numbers in the tables in the appendix for individual jurisdictions, it appears the study authors compared the demographics of all stopped motorists within a jurisdiction (and not just intown motorists) to the jurisdiction’s population demographics. Is that what was done? Isn’t that an inappropriate benchmark? (By contrast, the Connecticut studies by Barone do a resident-only analysis and a driving population estimate analysis.)
- In the “state average” comparative analyses, which are loosely based on Barone’s work, the study says it’s “[f]ollowing the Connecticut model” (on p. 22), but it seems like Barone does a “net” analysis²² (he did this in the 2020 study of 2018 data as well)²³ that takes into account each jurisdiction’s underlying demographics compared to the state’s overall demographics, which was not replicated by the EOPSS study. (He also used a number of other different methods for evaluating stop disparities that this study did not use, including a synthetic control analysis using propensity scoring, an estimated driving population comparison, and a resident-only stop comparison.) Why did the EOPSS study not fully follow the model? Could the study be updated to replicate the full Barone analysis?

²² KEN BARONE ET AL., INST. FOR MUNICIPAL AND REGIONAL POL’Y, CENTRAL CONN. STATE UNIV., STATE OF CONNECTICUT: TRAFFIC STOP DATA ANALYSIS AND FINDINGS, 2019 at 33 (2021), https://assets.website-files.com/6076e3f57e39855392637f16/614b554b4f7e47dc24f9e895_2019%20CTRP3%20Traffic%20Stop%20Analysis%20and%20Findings%20Report-%20Draft.pdf#page=51.

²³ KEN BARONE ET AL., INST. FOR MUNICIPAL AND REGIONAL POL’Y, CENTRAL CONN. STATE UNIV., STATE OF CONNECTICUT: TRAFFIC STOP DATA ANALYSIS AND FINDINGS, 2018 at 24 (2020), https://assets.website-files.com/6076e3f57e39855392637f16/608969ac86055d0bd5d5e680_2018-Connecticut-Racial-Profiling-Report.pdf#page=43.

- Have any other communities raised questions about issues of underreporting of citations, as in Brookline?²⁴
- Will you make the underlying de-identified data public for statistical analysis by others?
- The scale of underreporting due to the fact that the RMV does not track stops that end only in a verbal warning is a significant limitation (*see* EOPSS Study at 36), and merits more discussion. Evidence from litigation suggests that certain specialized units, like gang units, engage in wholesale pretext stops that are often unrecorded. For example, one of the Boston Police Department Youth Violence Strike Force officers in the *Long* case testified that he pulled over at least a thousand people each year, but only wrote about 5 citations. (*See* Appellant’s Brief in *Commonwealth v. Long* at 16.)²⁵ This dramatic underreporting of stops that particularly target people and communities of color may dramatically distort the findings—at least for the affected units and departments, if not for the data from the entire state (given the predominance of cities in the data).
- The researchers note that race data being coded and reported by the police based on their perception is a limitation of their study. (EOPSS Study at 36.) This should not be framed as a limitation. In the same paragraph, the study authors accurately explain, “Although officer perception of race/ethnicity is what *should* be used in an analysis of law enforcement stops examining racial/ethnic disparities for potential bias, it is important to note this limitation because of the possibility that an officer’s perception of one’s race/ethnicity may or may not be correct.” *Id.* Noting the source of the data for clarity is appropriate, but framing this as a limitation of the study—or a “potential validity concern,” as Dr. Gustafson said during the March 2nd public hearing—is potentially misleading. Where the study is attempting to analyze how people were treated based on officer perception of their race, using the race noted by officers is not a limitation but rather an accurate method—indeed, a methodological requirement.
- The study cannot account for officers visibly perceiving the race of passengers (there are no data about the presence of passengers in the vehicle, and what their race might be), but the Supreme Judicial Court’s precedent acknowledges that racial profiling can also be motivated by the race of passengers.

²⁴ While the EOPSS Study analyzed 307 stops in Brookline for the relevant period (EOPSS Study at 102), the Brookline Police Department’s year-end report analyzed 3,400 moving violations for the full 2020 calendar year. *See* BROOKLINE POLICE DEP’T, 2020 YEAR END REVIEW at 19 (2021), <https://www.brooklinepolice.com/DocumentCenter/View/1363/2020-Year-End-Report#page=19>. It is unclear whether the “n” for moving violations should equal the “n” for total number of traffic stops, or the nature of the relationship, but the overall discrepancy raises questions. Did Brookline Police tally each moving violation separately, even if a driver was issued multiple violations at a single stop and on a single citation? Did Brookline Police omit traffic stops that resulted in a warning? The report also notes that 2020 saw a *decrease of 75.26%* from the 13,745 moving violations issued in 2019, *id.* at 5—emphasizing that 2020 was an anomalous year due to the onset of the COVID-19 pandemic, *see* EOPSS Study at 39.

²⁵ *See* App. Br. at 16, *Commonwealth v. Long*, 485 Mass. 711 (2020), http://www.ma-appellatecourts.org/pdf/SJC-12868/SJC-12868_01_Appellant_Long_Brief.pdf#page=16.