

**BELDOCK LEVINE & HOFFMAN LLP**  
99 PARK AVENUE, PH/26<sup>TH</sup> FLOOR  
NEW YORK, N.Y. 10016

JONATHAN MOORE  
DAVID B. RANKIN  
LUNA DROUBI  
MARC A. CANNAN  
CYNTHIA ROLLINGS  
JONATHAN K. POLLACK  
HENRY A. DLUGACZ  
STEPHEN J. BLUMERT  
MYRON BELDOCK (1929-2016)  
LAWRENCE S. LEVINE (1934-2004)  
ELLIOT L. HOFFMAN (1929-2016)

TEL: (212) 490-0400  
FAX: (212) 277-5880  
WEBSITE: blhnj.com

COUNSEL  
BRUCE E. TRAUNER  
PETER S. MATORIN  
KAREN L. DIPPOLD  
MARJORY D. FIELDS  
EMILY JANE GOODMAN  
JUSTICE, NYS SUPREME COURT, RET.  
FRANK HANDELMAN

REF:

WRITER'S DIRECT DIAL:

May 17, 2024

Re: **Monitor's 20th Report**

*Floyd, et al. v. City of New York*, 08-CV-1034

Dear Judge Torres:

Pursuant to Fed.R.Civ.P. 53(f), we write on behalf of *Floyd* Plaintiffs ("Plaintiffs") in the above-entitled actions to object to several key parts of the statistical analyses used by the Independent Monitor in her April 11, 2024 20<sup>th</sup> Report regarding "Racial Disparities in NYPD Stop, Question, and Frisk Practices." (the Racial Disparities Report) *See* Dkt. 927-1. For the reasons set forth below, Plaintiffs, along with Plaintiffs' experts, Professors Jeffrey Fagan and Jack Glaser, object to several of the conclusions the report makes regarding "diminished" racial disparities in stops made by New York City Police Department officers. As the Monitor team rightly pointed out in the Racial Disparities Report itself, "the overall percentage of stops by race and ethnicity remained largely unchanged," especially when taking into account the large proportion of stops by the New York Police Department revealed to be undocumented. *Id.* at 53. However, there are other significant design and analysis decisions which undermine the validity and conclusions of the Report. Under the circumstances, many of the conclusions regarding the lack of racial disparities in police stops in New York City are misleading and simply not accurate.

We appreciate the efforts made by the Monitor's team to engage in robust discussions with the Plaintiffs and the City regarding issues surrounding this report. Nevertheless, we feel these issues are so important that they need to be brought to the Court's attention.

Plaintiffs emphasize that the disparities noted in this report remain virtually the same since these cases were tried in 2013, and in recent years, have grown more pronounced. *Compare* Table 1, Black, 2013 to 2017-2022 and Hispanic, 2013 to 2017, 2018, 2022. The issue of racial disparities in SQF encounters is one that must be addressed at *all* levels of NYPD leadership—through emphasis and discussion at trainings, supervisors that identify and correct police officer misconduct

and documentation issues, early intervention that not only identifies misconduct but prevents future misconduct through application of interventions, and meaningful discipline that holds officers accountable for their misconduct with more than instruction. We stand by the comments submitted to the Monitor team on January 12, 2024, by *Floyd* Plaintiffs' experts identifying the significant methodological concerns they raised with this Report. We briefly reiterate their primary concerns below. Should the Court wish for more detail, Plaintiffs would welcome the opportunity to provide the Court a copy of the experts' full analysis.

1. **The serious problem of under-reporting raises significant questions about the Report's findings regarding diminished racial disparities in stops.** In its executive summary, the Report notes that "[a]ny analysis of racial disparities and compliance using reported stops must acknowledge that the actual results might be different if all stops were documented." Page 4. Regardless of the conclusions that have been reached, this is a profound methodological problem that renders unreliable the Report's ultimate conclusion.

This problem was not present in the two reports<sup>1</sup> by Plaintiffs' expert, and comparisons of the scope of missing or undocumented observations in the present report seem to be inappropriate and uninformative. Plaintiffs' expert reports submitted in 2010 and 2012 presented evidence of racial and ethnic disparities in stops and stop outcomes based on study populations<sup>2,3</sup> that averaged 531,384 encounters per year from 2006-2012. These robust sample sizes across primary sampling units - precincts and census tracts - permitted estimations of disparities with minimal risk for sampling bias owing to small sampling averages or missing observations.

Compared to the 2010 and 2012 Plaintiffs' reports, the 20th Monitor report confronts both small numbers of observations across sampling units and undocumented stops of unknown size within units. From 2014-2022,<sup>4</sup> the report analyzed 140,899 stops, an annual average of 14,089 stops across the 75 NYPD precincts. Under these conditions, even the simplest estimate of a mean number of stops is confounded with the sample size itself within temporal (months, years) or spatial units (precincts, census tracts or block groups) and appears to be a random variable.<sup>5</sup> One cannot simply ignore these problems, including the fact that the sampling strata are not mean independent of the potential outcomes, even after conditioning on the observed covariates.<sup>6</sup> In other words, these twin problems of (a) small

---

<sup>1</sup> See Fagan, J. (2010), Expert Report in *Floyd v. City of New York et al.*, U.S. District Court for the Southern District of New York, 08 Civ. 1034 (SAS); Fagan, J. (2012), Second Supplemental Report, *Floyd v. City of New York*, 08 Civ 1034 (SAS); and Fagan, J. (2012).

<sup>2</sup> 2,095,281(2006-2009) encounters. Fagan, 2010, Table 1 at 19.

<sup>3</sup> 1,624,410 (2010-2012) encounters. Fagan, 2012, Report at 7.

<sup>4</sup> The first year after the *Floyd* order went into effect was 2014.

<sup>5</sup> Molenberghs G, Kenward MG, Aerts M, et al. On random sample size, ignorability, ancillarity, completeness, separability, and degeneracy: Sequential trials, random sample sizes, and missing data. *23 Statistical Methods in Medical Research*. 11 (2014), doi:[10.1177/0962280212445801](https://doi.org/10.1177/0962280212445801)

<sup>6</sup> Jiang J, Yang S, and Ding P. "Multiply robust estimation of causal effects under principal ignorability." *84 Journal of the Royal Statistical Society Series B: Statistical Methodology* 1423 (2022). See, also, Gaebler, J., Cai, W., Basse, G., Shroff, R., Goel, S. and Hill, J., 2022. A causal framework for observational studies of discrimination. *9 Statistics and Public Policy* 26 (2022).

samples with (b) unknown numbers of potentially biased unobserved cases undermine the reliability of the Monitor's conclusions on racial disparities.

Despite these problems, the report concludes that, based on an audit of a random sample of body-worn camera (BWC) video, 31.4% of stops are undocumented, and offers estimates of the impact of these undocumented (missing) cases on the estimates of racial disparity. Plaintiffs' and the City's experts have many questions about the bases for these conclusions. First and foremost, however, the documentation of a high rate of nonreporting is proof positive that many stops covered under the *Floyd* court order are not being documented. This is a first-order problem for two important reasons: 1) The recorded rates of police activities to be monitored are substantially below the actual police activity rates; and 2) There is good reason to surmise that the rates of underreporting are masking racial disparities.

The first point is self-evident – the presence of stops observable in BWC data that are not included in what is supposed to be a complete accounting of stops *proves* that there is noncompliance. Unless the sampled BWC data are by some stroke of terrible luck dramatically non-representative of police activities, the underreporting is substantial. The second point reflects that, in addition to issues around documentation, there may be issues of activation or de-activation of BWCs as well. At this point we simply do not know what the compliance rate on BWC activation is. If officers are less likely to activate (or perhaps more likely to deactivate) their BWC's when stops are constitutionally questionable, (e.g., improperly based in part on race), this could reflect a systematic bias in the data as well as in the actual conduct of the police that the data reflects. This has to be considered in the context that, historically, the data have consistently shown non-Whites to be stopped at disproportionately high rates, and, specifically, to be stopped at higher rates for more dubious reasons (e.g., furtive movements, high crime area, bulge), and to prove less justified (i.e., lower search yield rates). Therefore, if officers are more likely to stop non-Whites for dubious circumstances, and less likely to have their cameras activated for dubious stops (both highly plausible conditions), stops of non-Whites would be even more underreported than the audit finds. These reporting deficits compromise estimates of racial disparities that go to Fourteenth Amendment compliance as well as constitutional compliance with Fourth Amendment requirements.

The significant issues surrounding the lack of adequate documentation of stops, questions and frisks, which has now persisted for years, begs the question of how the resources in this case are being utilized. Rather than simply kicking the can down the road with regard to the lack of documentation problem, the resources of the City, with the assistance of the Monitor and the Parties, would be better spent attempting to solve the documentation problem rather than producing reports on racial disparity that are fundamentally flawed and, at the end of the day, do not advance the discussion about whether the City is in compliance with the remedial orders of this Court.

2. **The estimate that 31.4% of stops are not reported is troubling is itself compelling evidence of non-compliance.** However, more information is needed to explain the estimate. The estimates are based on coder ratings and agreements of a set of videos whose sampling parameters - due to non-reporting or non-activation - are unknown. But this raises a critical question: is that estimate global, applicable to all racial and ethnicity groups? Wouldn't a race- or ethnicity-specific estimate, with upper and lower bounds, yield more accurate

estimates of racial disparities adjusted for incomplete data? And can we assume that these undocumented stops are missing at random? It seems unlikely, given the explanation of the parameter. And even if you know, what race? stop rationale? And what locations? In addition to the potential bias from undocumented stops, we are concerned that these undocumented stops may not be randomly distributed by location, compounding the problem of race and undocumented stops.

We are not persuaded by the arguments offered for the estimate for unreported stops in the Report, given the widespread implementation problems the NYPD continues to have with their use of BWCs. The volume of missing observations severely undermines claims that there is no meaningful disparate impact. Given the sampling and measurement issues in the BWC study (late or no videos, no estimates by suspected crime, time of day, variability by locale, etc.), the 31.4% estimate of undocumented stops should be treated as a parameter, and models using this parameter should consider upper and lower bounds that capture the uncertainty in the BWC data. The adjustments in the Report's appendix are opaque on how the bounds were set and applied. At the least, this parameter should be estimated as race-specific rather than global measures.

In particular, the estimate for documentation rate for Whites is likely to be very unstable, given that there are only 10 total stops of Whites in the BWC data. This leads the Report to conclude that White stops are undocumented at a rate of 40% (literally, 4 out of 10), which is higher than the Black and Hispanic nonreporting rates, but this number of undocumented White stops could easily be 10 or 20% or 60 or 70%, for that matter.<sup>7</sup> Even if the racial distribution of those stopped (as indicated by BWC data) but not reported is similar to the racial distribution of those in reported stops, this is not sufficient information to conclude that there is no bias in the underreporting. Specifically, we would need to know the breakdown of stops with respect to both reasons and outcomes, similar to the Expert Reports at the time of trial in 2013.<sup>8</sup> Not all stops are the same; reasons and outcomes for undocumented stops may differ by race of subject. With such a small sample of undocumented BWC stops, this analysis cannot be done reliably.

3. **As presented, the 2013-2019 year-specific data analysis adds little to the consideration of constitutional compliance.** Testing for racial disparities for each year separately can hide the systemic factors that produce racial disparities over time, including precinct differences, bases of suspicion and suspected crime, and factors unique to specific years such as changes in NYPD leadership. Year to year differences can mask trends that span the longer time periods.<sup>9</sup> The original analyses cited by the District Court in the *Floyd* liability Opinion was a panel design with continuous time series by police precinct, with fixed effects for individual years to control for unique time events. *See* Dkt. 373 at 31.<sup>10</sup>

<sup>7</sup> Using simple binomial probabilities, if the real rate of underreporting of Whites is 31.4%, there's a 38.8% chance that four or more out of 10 stops would be undocumented at random.

<sup>8</sup> For example, suspected crime as well as the suspicion indicators, and the outcomes including arrest, summons, search, UOF, and contraband seized.

<sup>9</sup> Previous analyses of stop rationales by race suggests that officers invoke standardized race-specific scripts to explain decisions for stops and searches. See, Jeffrey Fagan and Amanda Geller, "Following the Script: Narratives of Suspicion in Terry Stops in Street Policing," 81 *University of Chicago Law Review* 51 (2015). Jeffrey Fagan, "No Runs, Few Hits, and Many Errors: Street Stops, Bias, and Proactive Policing," 68 *UCLA L. Rev.* 1584 (2021).

<sup>10</sup> Here, we are presented with summary data for each year and a claim that the race differences are approaching zero. At the least, a test for serial correlation of a large sample of units (tracts, precincts, or block group) over time, and the

4. **The report continues to rely on significance tests to draw conclusions about racial disparities.** Relying on statistical significance to draw conclusions risks assuming that small effects matter may not be robust when we consider the design and sample sizes. The data are a universe of stops, searches, arrests, etc. The significance tests suggest that small differences matter or don't matter. But there's a difference between statistical significance and practical significance. Some of these differences are just too small to imply a meaningful substantive difference between groups or other units of analysis.<sup>11</sup>

Even worse, inferring a trend based on changing significance levels is not an appropriate application of statistical significance testing. P-values can increase (i.e., go above the .05 conventional standard for “significant”) based on reductions in sample sizes and/or increases in variance (i.e., variability in rates). This is particularly relevant to the exercise described in Section V, Part D of the Report, which estimates the effect of correcting the “sample” size for the under-documentation rate derived from the BWC audit. The results (reported in Table 7) suggest that some of the disparities would be statistically significant if the sample size were increased to correct for the estimated under-documentation rate. This is well intentioned, no doubt, but it reinforces the inappropriate application of significance testing.

The differences are what they are, and the conclusions about whether they are problematic should be based on assessments of practical, not statistical significance. In other words, they should be based on how large the differences are.

5. **Year to year comparisons for 2021-22 are unhelpful.** Comparing only the recent years, even with omitting 2020, makes any conclusions temporary and limited to the unique contexts of those years. This isn't a panel design; any changes could be random or temporary in a short time span, with even short-term spikes in race-specific stop rates changing the results of this model. If there is policy or legal relevance to the more recent period, a more robust design would include a panel from 2017-22, with a sufficient pre- observation window to the 2020 or 2021 start date. But even this interval could pose validity threats due to period effects (i.e., specific, idiosyncratic changes, such as a pandemic). First, 2020 data are omitted, with little explanation.<sup>12</sup> Second, both 2020 and 2021 were pandemic years, with disruptions both in policing and public (social and economic) life. Given the artifact of the

---

identification of a trend over time, becomes relevant when data are reported with a unitary estimate. *See, e.g.,* J. Durbin and G.S. Watson, Testing for Serial Correlation in Least Squares Regression. I. In: Kotz, S., Johnson, N.L. (eds) Breakthroughs in Statistics. Springer Series in Statistics. Springer, New York, NY 9 (1992). Claims of diminishing disparities should be presented as a temporal trend over time showing sustained change, with varying time trend tests (i.e., linear vs. quadratic) and other standard adjustments such as seasonality. At a time when stops are increasing, , whatever year-end data are shown could well reverse in a longer window. A more robust panel analysis would test for trend. This becomes more important with the measure of interest being difference in group rate when the variances in those group rates may be changing over time.

<sup>11</sup> All this might be more acceptable if accompanied by a statistical power analysis (i.e., calculations of the likelihood of false negative results given the sample size). It is especially problematic when inferences about trends are made based on changes in significance level (e.g., from “significant” –  $p < .05$  to “nonsignificant” –  $p > .05$ ). P-values represent the probability that we are committing a false positive error (i.e., that there is a difference, when, in fact, there isn't) if we conclude that there is a difference in the population based on what is observed in a sample. To the extent that these data represent the population of stops that occurred during a given time period (not a random sample drawn from a population), underreporting notwithstanding, p-values are not applicable.

<sup>12</sup> How many stops were there in 2020? The N reported (8,947 for 2021) is also reported in fn. 2. for 2020. Which is it for 2020 or 2021?

exclusion of 2020 data, the conclusion that a small number of precincts are responsible for the majority of the stops in the subsequent two years seems speculative.

6. **There is no attention to the differences by race or ethnicity in the reasons for the search, and little attention to the reasons for stops.** Consent searches are not distinguished from probable cause searches or search incident to arrest. Probable cause can be teased out of the stop dataset, with attention to the specific search rationales. The conclusion about higher search thresholds in 2021 (p. 3) can be challenged with a more controlled and articulated model. Are those higher threshold rationales supposed to be proxies for the probable cause standard for search? Again, what conclusions can be made with the arbitrary exclusion of 2020 data? The absence of attention to stop reasons suggests that conclusions about stop rates by race may be overlooking salient race-correlated factors that explain differences by race in both stop events and search outcomes.
7. **It's not clear whether the statistical models used by the Monitor controlled for the suspected crime.** If so, those parameters were not reported. Past analyses, including the analyses proffered at trial, controlled for the suspected offense. When the report does control, it is for all crimes, undifferentiated by type of crime. In past analyses, the suspected offense was a consistent predictor of differences in stop rates by race. Given that police allocations are informed by crime severity, the absence of suspected crime suggests a possible omitted variable problem.

Because of these significant problems with the Monitor's 20<sup>th</sup> report, the *Floyd* Plaintiffs object to the filing of this report on the docket.

Thank you for your consideration.

Respectfully submitted,

/s/ Jonathan C. Moore

Jonathan C. Moore

Luna Droubi

Marc Arena

BELDOCK LEVINE & HOFFMAN, LLP

99 Park Avenue, Penthouse Suite

New York, NY 10016

(212) 490-0400

/s/ Baher Azmy

Baher Azmy

Samah Sisay

Remy Burton

CENTER FOR CONSTITUTIONAL RIGHTS

666 Broadway, 7th Floor

New York, NY 10007

(212) 614-6464

*Counsel for the Floyd Plaintiffs*