Disparity Review – Part I

Using Propensity Score Matching to Analyze
Racial Disparity in Police Data

April 2019

EXECUTIVE SUMMARY

The Consent Decree included several requirements related to Bias-Free Policing, including the general mandate that the Seattle Police Department ("SPD") "deliver police services that are equitable, respectful, and free of unlawful bias, in a manner that promotes broad community engagement and confidence in the Department." To that end, it also required updates to SPD's Bias Free Policing policies and training. These requirements were satisfied during Phase I of the Consent Decree. [CITE]. However, the Consent Decree also stated that, in consultation with the Community Policy Commission ("CPC"), SPD should consider whether to revise SPD Manual 5.140. In [YEAR], SPD did amend Policy 5.140 (Bias-Free Policing), which now contains a "Disparate Impacts" provision.

Under that policy, the Seattle Police Department committed to eliminating policies and practices that have an unwarranted disparate impact on certain protected classes. While recognizing that it is possible that the long-term impacts of historical inequality and institutional bias could result in disproportionate enforcement, even in the absence of intentional bias, the Department's policy is to identify ways to protect public safety and public order without engaging in unwarranted or unnecessary disproportionate enforcement.

As the Monitor found in its Tenth Annual Systemic Assessment on Stops and Detentions, disparity in law enforcement activities exists in the City of Seattle with respect to Stops and Detentions. SPD's own analysis, including in preparation of this report, confirms this fact. The question, therefore, becomes whether the disparities are "unwarranted." A common example in thinking about warranted vs. unwarranted disparity is enforcement activities with respect to a criminal gang. A gang may be formed by members with the same race or national origin. If that gang engages in criminal activity and law enforcement addresses that activity, the statistical results may indicate a higher rate of enforcement against that particular race or ethnicity. Such disparity would not be considered "unwarranted."

The challenge, of course, is in discerning what observable disparities are warranted vs. unwarranted. As the Monitor recognized in its Tenth Annual Systemic Assessment on Stops and Detentions, quantifying unwarranted racial disparities in police contacts is difficult and there is no academic consensus about the best method to do so. In that assessment, using various approaches, the Monitor's team sought to examine, both quantitatively and qualitatively, the nature of the disparity observed and, broadly, the extent to which stops, and outcomes, were supported by policy and law.

Among the methodologies selected, the Monitor used the technique of Propensity Score Matching (PSM) – an approach that has gained popularity in social science fields, including criminal justice – that uses regression to "score" how similar events are to each other across a variety of factors and match them for comparison.   For example, PSM was used to match a *Terry*

stop where the stopped individual was Black to a stop where the subject was White, but all other known (available in fielded data) were as similar as possible.

In the current report, SPD builds upon and extends the Monitor's application of Propensity Score Matching.[1] However, SPD can do so in even more robust manner through its Consent Decree-driven investments in data systems and analytic tools and professionals. From these systems, SPD has created the capacity to routinely analyze its processes and actions in a way historically restricted to multi-year engagements with research partners. Accordingly, under this review and report, SPD was able to (1) match additional factors that now are available through SPD's data analytics platform but were not available to the Monitoring Team at the time of data production for the Tenth Systemic Assessment; and (2) apply Propensity Score Matching to examine the role of race in Use of Force data relating to the pointing of a firearm – an area that, like stops and frisks, is often a matter of substantial officer discretion.

As is shown in the published reports SPD continues to release, and directly analyzable in the public dashboards SPD has created, it is important to note that the behaviors under review here are relatively rare events. To put the numbers presented here in context, over the study period, SPD answered 978,608 calls for service, comprising 2,175,181 associated officer dispatches. Across these events, officers, engaged in 19,544 *Terry* stops and 4,371 uses of force – approximately three-quarters of which force involved no greater than low level, Type I force. The Department emphasizes this context not to minimize the occurrence of these events, nor the potential disparity that exists in their occurrence. The intent is quite the opposite: to highlight the need to examine these events with a precise instrument capable of isolating true differences between relatively rare events occurring in a sea of activity across an entire complex and ever-changing city.

As SPD has committed in its policy, if and when SPD identifies and verifies unwarranted disparate impacts, the Department will consult with neighborhoods, businesses, community groups,

---

[1] See O'Toole, K (2018, pp. 70-71) *The Garda Inspectorate: Driving Collaborative Reform Through a Model of Equilibrated Governance*. (Submitted and accepted in fulfillment of PhD requirements of the School of Business, Trinity University Dublin), describing the model for good research in non-traditional scientific method:

> The learner works through a cyclical process of consciously and deliberately (1) assessing a situation which is calling for change; (2) planning to take action; (3) taking action; and (4) evaluating the action, leading to further cycles of planning, taking action, and so on. Just as the empirical researcher may adjust a methodology to account for confounding variables that may surface, research cycles are repeated and modified to account for knowledge learned from the prior iteration, until such time the researcher determines the process is complete.

and/or the Community Police Commission, and the Office of the Inspector General for Public Safety, to explore equally effective alternative practices that would result in less disproportionate impact. Alternative enforcement practices may include addressing the targeted behavior in a different way, de-emphasizing the practice in question, or other measures.

This report moves SPD one step closer toward that end: it solidifies findings of disparity regarding frisk rates for weapons and the pointing of firearms. As discussed in further detail below, SPD will focus additional research and analytics on these issues in the next iteration of its study of disparity, which is due to be filed with the Court in December 2019.

Summary of Major Findings

The report examines racial and ethnic disparities in a variety of police actions where research suggests there is more of an opportunity for disparate effects. The data for this report cover two-and-a-half years (2016 through mid-2018) of interactions. The primary results of these analyses, which are based on a relatively low number of observations, along with initial thoughts and/or next steps for analyses and action, are summarized below:

1. Frisks
   a. Our overall findings demonstrate that minorities are frisked at a higher rate than non-minorities.
   b. There were 3,021 frisks[2] conducted during the 14,036 stops (21.5%) reported and matched in the study time period:

      i. Asians were frisked (frisks = 123; rate = 25.7%) more than whites (frisks = 92; rate = 19.2%) in matched stops
      ii. Blacks were frisked (frisks = 1,183; rate = 24.7%) more than whites (frisks = 1,002; rate= 20.9%) in matched stops
      iii. Hispanics were frisked (frisks = 229; rate = 27.2%) more than whites (frisks = 177; rate = 21%) in matched stops
      iv. American Indians/Alaska Natives were frisked (frisks = 127; rate = 10.6%) more than whites (frisks = 122; rate = 10.2%) in matched stops
      v. Non-white[3] individuals were frisked (frisks =1,636; rate = 23.3%) more than whites (frisks = 1,385; rate = 19.7%) in matched stops

2. "Hit" rate (finding a weapon during a frisk)
   a. We found that frisks of minorities produce weapons about 1/3 less frequently than frisks of similarly-situated whites.

---

[2] Note: these incident counts are different than total counts reported later in the report as these are the total events within each matched population.
[3] Subjects reported to be of "unknown," "missing" or "other" perceived race by officers.

    b. Of the 3,021 frisks in the study, a weapon was found in 627 incidents, an overall hit rate of 20.7%

        i. Asians had a weapon found (hits = 21; rate = 17.1%) in fewer similar frisks than whites (hits = 20; rate = 21.7%)

        ii. Blacks had a weapon found (hits = 191; rate = 16.1%) in fewer similar frisks than whites (hits = 247; rate = 24.7%)

        iii. Hispanics had a weapon found (hits = 42; rate = 18.3%) in fewer similar frisks than whites (hits = 40; rate = 22.6%)

        iv. American Indians/Alaska Natives had a weapon found (hits = 26; rate = 20.5%) in fewer similar frisks than whites (hits = 42; rate = 34.4%)

        v. Non-whites[4] had a weapon found (hits = 275; rate = 16.8%) in fewer similar frisks than whites (hits = 352; rate = 25.4%)

3. Stop Durations
    a. Of the 14,036 stops in the study, the average duration was 10.9 minutes.

        i. Asians were stopped (average time = 11.1 mins) for a shorter period than whites (average time = 11.9 mins) in similar stops

        ii. Blacks were stopped (average time = 11.0 mins) for a shorter period than whites (average time = 11.3 mins) in similar stops

        iii. Hispanics were stopped (average time = 15.4 mins) for a shorter period than whites (average time = 15.6 mins) in similar stops

        iv. American Indians/Alaska Natives were stopped (average time = 10.9 mins) for a longer period of time than whites (average time = 10.5 mins) in similar stops

        v. Non-whites[5] were stopped (average time = 11.1 mins) for a longer period of time than whites (average time = 10.8 mins) in similar stops

4. Use of Force – Pointing of Firearm
    a. Firearms were pointed at non-whites about 30% more often than at similarly situated whites.
    b. In the 3,138 incidents in this study where force was used, a firearm was pointed in 675 (21.5%)

        i. Asians had a firearm pointed at them (point = 47; rate = 14.2%) more frequently than whites (point = 33; rate = 10%) in similar incidents

---

[4] Subjects reported to be of "unknown," "missing" or "other" perceived race by officers.

[5] Ibid.

ii. Blacks had a firearm pointed at them (point = 290; rate = 13.8%) more frequently than whites (point = 203; rate = 9.7%) in similar incidents

iii. Hispanics had a firearm pointed at them (point = 54; rate = 19.4%) more frequently than whites (point = 29; rate = 10.4%) in similar incidents

iv. Those individuals who were not specified or not in a large enough population to separately examine had a firearm pointed at them (point = 149; rate = 12.4%) more frequently than whites (point =184; rate = 10.1%) in similar incidents

v. Non-whites[6] had a firearm pointed at them (point = 384; rate = 12.2%) more than whites (point = 291; rate = 9.3%) in similar incidents

Taken together, these results generally are in-line with those produced by the Monitoring Team in the Tenth Assessment. The current analysis did find a lower rate of disparity among the decision to frisk an individual during a stop; this difference likely is the result of the SPD model having additional data available to it to ensure the stops were more comparable.

It should also be noted that SPD recently conducted an audit of its stop and detention practices for the period from January 1 to June 30, 2018. It determined that in 93.5% of all stops during the study period, the stop template filled out by the officer contained a narrative that documented adequately reasonable suspicion for the stop;[7] in over half of a random sample of the remaining 6.5% of cases, adequate reasonable suspicion was identified in the accompanying police report. The Monitor and the U.S. Department of Justice each performed an independent validation of these findings and concluded that SPD continues to maintain compliance with the relevant Consent Decree requirements. This audit was submitted to the Court on March 7, 2019.

While SPD has limited its approach to Propensity Score Matching for purposes of this report, SPD does not intend to suggest that there are not other ways to examine these issues; certainly, the robust argument in academic circles as to how best to address the issue of disparity well illustrates the debate. Rather, given the nature of SPD's data, in the interest of focusing this report, and with a good foundation for future steps laid using this methodology in the Monitor's Tenth Assessment, SPD has selected Propensity Score Matching as the most appropriate approach for the purposes of this research cycle and this particular inquiry.

The PSM methodology allows SPD to know that these are real and true differences in how the department interacted with these populations while providing police services. They do not, however, explain what was occurring in each of these interactions, including a variety of very important factors not yet available to the model – officer prior knowledge of the individual, behaviors and words expressed during the incident, and recent crime trends in the specific area. SPD plans to include more of these factors in Phase II report work, as well as to undertake the qualitatively intensive analyses of "scoring" the actual interactions. SPD is working with experts in assessing police behavior and disparity to develop the tools for this analysis so the nature of

---

[6] Subjects reported to be of "unknown," "missing" or "other" perceived race by officers.

[7] As noted previously, the lack of clearly documented reasonable suspicion does not mean the remaining 6.5% of stops were not proper.

each interpersonal interaction can be understood beyond administrative data. SPD will leverage the results of these analyses to help prioritize the areas where it should begin additional work in understanding what may be leading to the outcomes of the interaction and how to potentially address those factors through policies, training, and operational guidelines.

While all the matches that showed signs of disparity will be further examined, SPD will prioritize its analytical work for those pairs that showed larger differences, namely: (i) the disparate rate of frisk for Asian, Black, and Hispanic individuals compared to Whites; (ii) the lower rate of finding a weapon during those frisks for Asians, Blacks. Hispanics, or American Indian/Alaska Natives compared to Whites; and (iii), the higher rate of pointing firearms at Asian, Black and Hispanic individuals compared to Whites. It is possible that the differences in these behaviors are attributable to circumstances not captured currently in fielded data (such as officers' prior knowledge of the suspect, etc.).

INTRODUCTION

Under Paragraph 223 of the Consent Decree, the Court retains jurisdiction over this matter "until such time as the City has achieved full and effective compliance and maintained such compliance for no less than two years."   On January 10, 2018, the Court entered an order finding the Department to be in "full and effective compliance" as of the date of the Order, thus commencing the two-year "sustainment period." Dkt. #439. The Court further ordered the parties and the monitor to "meet, confer, and prepare a plan for discharging their obligations under the Consent Decree" during this two-year period.

On March 13, 2018, the Court entered an order approving the Sustainment Plan developed pursuant to the Court's January 10th order. This plan, and an attached matrix of deadlines, became the governing documents for this Sustainment Period. To a large extent, the Sustainment Plan follows the scope of Consent Decree itself, as it requires SPD to self-assess those core topical areas that, collectively, cover the entirety of SPD's obligations as set forth in Section III of the Consent Decree ("Commitments") and assessed, comprehensively, in Phase I. The Sustainment Plan also includes commitments by SPD not specifically called for under the Consent Decree. SPD is committed to continuing the work it has done to further itself as a "learning" institution and driver, nationally, of innovative and data-informed policing. This analytic project is a part of that promise.

Included within the Sustainment Plan is SPD's commitment to continue its emphasis on impartial, bias-free policing, to regularly examine disparities across its data to ensure that this emphasis is carried out in practice, and to provide the Court with a fuller view of "disparities with respect to stops, searches, and seizure; use of force; and other law enforcement activity." SPD also demonstrates that commitment with its Bias-Free Policing policy (Seattle Police Manual, 5.140) and requiring minimizing bias training that is coordinated with the state's Criminal Justice Training Center.

The Monitor noted in his Tenth Systemic Assessment,

> Sorting out whether disparity on the basis of suspect classifications, like race, is the result of intentional discrimination, the result of unknowing or subconscious bias, or is the effect of one or many factors having nothing to do with race or that are tangled up with race is challenging. When there are reasonable and legitimate reasons for a practice that produces disparities with respect to whom the practice is applied, the courts have been historically reluctant to invalidate government actions as discriminatory and impermissible.

> Consequently, neither the Consent Decree nor the Court-approved policies on stops and bias-free policing demand that SPD immediately stop practices that it may determine are linked to disparate impacts. Instead, and importantly, it requires that SPD determine whether such disparities are warranted or unwarranted and, where "unwarranted disparate impacts are identified" with respect to a given SPD practice or policy, "the Department will consult as

appropriate with neighborhood, business and community groups, including the Community Police Commission, to explore equally effective alternative practices that would not result in disproportionate impact."

Tenth Systemic Assessment, Dkt. #394, pp. 40-41 (internal citations omitted).

Consistent with this understanding, SPD's work over the past six years to identify disparities, and its evaluation as to whether disparities are "unwarranted," has taken many forms. As routine practice, SPD includes in each of its annual reports on use of force, crisis response, and stops and detentions, a section that details the subject demographics in each area. To foster further transparency and encourage exploration of its data, SPD presents data in each area in public-facing, queryable dashboards;[8] to foster the analytic inquiry of the many researchers who regularly seek this data, SPD offers public links to the raw data underlying the same. In meeting a key deliverable under the Sustainment Plan, SPD presented in its Stops and Detentions Audit, filed with the Court on March 7, 2019, a qualitative analysis of an agreed-upon subset of its stops data, to which the Monitor and DOJ verified continuing compliance with Consent Decree obligations; a quantitative effort (Pearson's Chi-square) to test the relationship between the outcome of the audit and the perceived race and gender of the subject failed to achieve statistical significance.

In the Tenth Systemic Assessment, the Monitor undertook a two-pronged approach to examine SPD's stops and detentions: a quantitative approach, which included several independent statistical approaches to examining disparity, and a qualitative approach, which involved a review of a statistically valid subset of stops documentation to determine whether the stops, and actions taken thereafter (length of the stop and any subsequent weapons frisk) were supported by articulated reasonable suspicion. As to the latter, the Monitor found that while there was documented disparity proportional to population demographics as represented in tract-level census data, SPD was in compliance with relevant Consent Decree requirements relating to both the documentation and constitutionality of its stops over that study period. In that regard, the Monitor reported as follows:

- <u>The vast majority of stops were adequately justified</u>. SPD officers have reasonable, articulable suspicion that the involved subject had been, was, or would soon be engaged in criminal activity – which is the required legal standard for initiating a so-called *Terry* stop – in 99 percent of stops. (In the remaining one percent, the Monitoring Team was unable to reach a determination based on the level of articulation in the documentation it reviewed.)

---

[8] Accordingly, for those interested in examining the findings presented in this report in the context of SPD's overall stops and detentions, or use of force, crisis response activity, or overall crime stats, SPD refers those readers to the data and information pages of its website.

- <u>The vast majority (97 percent) of frisks were adequately justified</u>. In 97 percent of frisks conducted during *Terry* stops, officers had appropriate and separate grounds for conducting a minimally-invasive search for a weapon during a *Terry* stop and were not automatically conducting a frisk of subjects simply because they were stopped. (In the remaining three percent, the Monitoring Team was unable to reach a determination based on the level of articulation in the documentation it reviewed.)

- <u>Most stops were appropriately limited to a reasonable scope and reasonable duration, as required under law and SPD policy</u>. Race did not impact the odds of being subjected to a stop of an unreasonable scope or unreasonable duration.

- <u>Few additional policy issues with respect to initiating or conducting the stops were identified</u>.

Tenth Systemic Assessment, Dkt. #394, pp. 4-6 (emphasis in original).

As to the quantitative analysis, the Monitoring Team prefaced its approach:

Evaluating the extent to which an individual's race or ethnicity affects the likelihood that he or she will be stopped by police is a challenging analytical task. To do so, researchers must develop a benchmark against which to compare the racial distribution of actual stop data with an individual's risk of being stopped in the absence of bias. An appropriate benchmark must incorporate the various legal and non-legal factors that shape stop risk, including when, where, and how often a person is out in public, and the nature of their appearance, activity, and demeanor while engaging in public activity, among several other relevant variables.

As reliable records of these variables are difficult – if not impossible – to obtain, analysts are forced to develop statistical proxies based on rough assumptions about the profile and activity of a jurisdiction's residents and the priorities of the relevant law enforcement actors. This has led to a robust discussion in academic and social science literature about both the merits and disadvantages of a host of statistical tests and benchmarks.

Tenth Systemic Assessment, Dkt. #394, pp. 50-51 (internal citations omitted).

Explicitly avoiding the debate as to "what type of statistical analysis is best or most accurate, or what 'benchmark' is most appropriate or analytically powerful" (Id. at p. 51), and cautioning as to interpretation of the findings given the limitations of both the data and the analyses themselves, the Monitoring Team chose to examine the question of "whether stop activity affects some people more than others" through three distinct approaches, carried across 13,124 administrative records of stops (*Terry* templates) generated by Seattle Police Officers between July 1, 2015 and January 31, 2107, described as, and with findings, as follows:

1. Overall, population-based analysis that compared the racial distribution of stops to Seattle racial population data.

2. Statistical modeling of the effects of beat-level variation in resident demographics and crime incident on the distribution of *Terry* stops at the patrol beat level, to provide some control for crime factors that may influence the findings above, through a series of multi-variate regressions.

3. A series of statistical tests intended to analyze post-treatment stops and outcomes (stop disposition, duration, and rate of frisk), where the point of comparison is among the population of those who have been stopped, rather than the general population.

See, generally, Tenth Systemic Assessment, Dkt. #394, pp. 51-77.

SPD recognizes the Monitor's Tenth Systemic Assessment as a comprehensive foundation from which to launch further inquiries. SPD's purpose in presenting this report, pursuant to its general obligation under the Sustainment Plan to further examine disparity in its data, is not to revisit or attempt to replicate the many sophisticated analyses that, either, were done by experts on the Monitoring Team or that have formed the vast body of academic literature around disparity throughout the criminal justice system – efforts that notwithstanding the dedicated careers of many within the social sciences still, as the Monitoring Team acknowledged, have not settled on either a complete or agreed-upon approach to examining this complex subject.[9] SPD is proud of its committed partnerships, nationally and internationally, to advancing the study of the social science of policing, but SPD is not itself an academic institution with the expertise or resources to undertake analyses equivalent in scope or sophistication as those performed by professional researchers. Nor does SPD seek to wade into what the Monitor referred to as the "robust discussion in academic and social science literature" as to the "correct" approach.

Rather, SPD's purpose in presenting this report, and the report that will follow, is three-fold: (1) primarily, to critically examine the data currently available to the SPD to determine if its actions are applied in a disparate manner, ; (2) to build upon the Monitor's findings by leveraging additional data now available within SPD's Data Analytics Platform (DAP) to further refine, by accounting for additional potential confounding variables, analyses around the role of race in

---

[9] This report expressly does not seek to provide an aggregate overview of its stops data; for those interested in exploring that data specifically, SPD directs readers to its public-facing, navigable dashboards through which the user can explore demographic information relating to subjects across a broad range of police encounters, including stops and detentions, use of force, and crisis response. To the extent that the reader is interested in a more qualitative assessment as to the nature, and constitutionality, of SPD's stops and detentions, the Department refers the reader to its Stops and Detentions Audit, Dkt. No. 547-1, filed with the Court on March 7, 2019, in which the Monitor found that SPD was sustaining its compliance with Consent Decree requirements with respect to its stops and detentions.

post-stop outcomes using Propensity Score Matching (PSM) in order to identify whether identified disparities are "unwarranted"; and (3) to eventually use the findings of this iterative study to examine the implication of its findings on any SPD policies, training, and/or operations as an organization committed to data informed decision-making and continuous improvement.

Second, consistent with principles of critical, iterative review this report builds on the prior work of the Monitor and offers direction for further work, but, as is true of research across this field of study, it should not be understood to be exhaustive or conclusive as to any cause(s) of the disparity noted. As those who study this complex topic professionally will note, there are inevitably circumstantial factors present in these encounters that are not captured through fielded data that may be discerned through a qualitative review of the facts of these cases. For example, in their validation of SPD's recent Stops and Detentions Audit, both the Monitor and DOJ observed that a substantial number of SPD's reported stops appeared not to be investigative stops, but rather supported by the higher bar of probable cause. One question raised, that may bear on training, is whether some frisks documented on *Terry* templates are more appropriately categorized as non-discretionary searches incident to arrest.

In its follow-up report, SPD will analyze any descriptive information obtained from a structured review – including examinations of body worn video, report narratives, and audio from a weighted sample of cases that may yield additional insight into the numbers reported here and provide yet further direction for future research. Additionally, this work will be informed by structured discussions with community members regarding their perceptions of what occurred in the interactions and how they could have gone differently. These discussions will be grounded in the additional analyses in the Phase II report, as well as work with a national expert on evaluating disparity in police actions, as well as the CPC.

METHODOLOGY[10]

Propensity Score Matching

Propensity Score Matching (PSM) allows a researcher to estimate the effect of a condition, intervention, or policy, while accounting for factors that predict the likelihood of receiving that treatment– in this case, the treatment is the effect of race on (1) certain *Terry* stop outcomes and (2) the likelihood that an individual will be the subject of a Type I, firearm-pointing use of force. PSM accomplishes this by pairing treatment and control units that have similar propensity scores – based on the presence or absence of known variables – when random assignment is not feasible.

---

[10] This is not intended to be a technical report on the methodology itself. A comprehensive discussion of PSM theory, the mathematical underpinnings of determining statistically valid matches, and the application is contained in the Monitor's Tenth Systemic Assessment. For purposes of readability, SPD refers the reader interested in a fuller analytic discussion to that well-articulated explanation or to the many studies cited herein.

In the social sciences, involving dynamic human interactions in the real world where latent variability is inevitable, PSM helps to focus the analysis by holding constant factors that can be held constant – in other words, it "eliminates many things from being causes, and this is probably very good, since it gives more specificity to the meaning of the word *cause.*" Holland (1986, p. 959).[11] PSM has been applied consistently in medical and psychological/psychiatric settings since the 1980s, but has only relatively recently been applied to the criminal justice sciences. Slade et al., (2008)[12], for example, used PSM to explore the role of substance use disorders in adult incarceration in a study of 780 juveniles. Gibson et al., (2009)[13] examined gang membership and victimization. A 2010 chapter in Springer's *Handbook of Quantitative Criminology* cites myriad studies and a growing popularity of the approach in criminology and criminal justice research.[14]

While there are a number of statistical approaches that researchers have employed to examine disparate impact, an advantage of applying PSM in observational studies, at least one researcher has noted, is that "adjusting for confounding variables using the propensity score offers an alternative to multivariate regression that is more interpretable, less prone to errors in model assumptions, and ultimately easier to present to stakeholders[.]"[15]; see also the Monitor's Tenth

---

[11] Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association.*, 945-960.

[12] Slade, E., Stuart, E. A., Salkever, D. S., Krakus, M., Green, K. M., & Ialongo, N. (2008). Impacts of age of onset of substance use disorders on risk of adult incarceration among disadvantaged urban youth: A propensity score matching approach. *Drug and alcohol dependence*, 1-13.

[13] Gibson, C. L., Miller, J. M., Jennings, W. G., Swatt, M., & Gover, A. (2009). Using propensity score matching to understand the relationship between gang membership and violent victimization: A research note. *Justice Quarterly*, 625-643.

[14] Apel, R. J., & Sweeten, G. (2010). Propensity score matching in criminology and criminal justice. In *Handbook of quantitative criminology* (pp. 543-562). New York, NY: Springer.

[15] Ridgeway, G. (2006). Assessing the effects of race bias in post-traffic stop outcomes using propensity scores. *Journal of Quantitative Criminology*, 1-29.

System Assessment, discussing the advantages of PSM over linear regression.[16]   As another researcher argued,

> Propensity score (PS) methods offer certain advantages over more traditional regression methods to control for confounding [variables] by indication in observational studies.   Although multivariate regression models adjust for confounders by modeling the relationship between covariates and outcome, the PS methods estimate the treatment effect by modelling the relationship between confounders and treatment assignment. Therefore, methods based on PS are not limited by the number of events, and their use may be warranted when the number of confounders is large, or the number of outcomes is small.[17]

Within the intersection of policing and disparate impact research specifically, two publications have applied PSM to estimate the effects of "decisions to search."   Holding constant other variables, Higgins, Jennings, Jordan & Gabbidon (2011) found that subjects identified as Black were more likely to be frisked than White subjects, but found no differences between Hispanics and Whites, suggesting that "race, but not ethnicity, appears to be a causal factor in a police officer's decision to search."[18]   Applying PSM to examine the effect of race in traffic stop outcomes in Oakland, Ridgeway (2006) concluded, among other findings, that while non-white drivers were treated equitably in terms of traffic stop outcomes such as citation rates and consent search rates, white drivers were less likely to be frisked, and that race appeared to have the strongest influence on the duration of the stop, with stops of Black drivers less likely to last less than ten minutes. As to the methodology itself, Ridgeway noted that PSM provides "a transparent, intuitive, and easily implemented method for assessing race" in outcome determination.[19]

SPD also recognizes, however, that (as with other singular methods of data analysis), PSM has limitations. In particular, PSM may result in false negatives for disparity. Because this report found positive results for disparity, this concern is not currently as issue. However, in order to prevent the possibility of false negatives in the second phase of this analysis (due to be filed in December 2019), SPD commits to providing basic summary statistics without adjustment for comparison groups, as well as standard (probit/logit/logistic/linear probability model)

---

[16] As the Monitor explained, "The quantity of interest is the effect of a subject's race or ethnicity on the likelihood of experiencing certain post-stop outcomes, which can be estimated using marginal effects in the case of linear fixed effects model, and average treatment effects for the matching models [such as that used here]."

[17] Benedetto et al. (2018) Statistical primer: propensity score matching and its alternatives. *Eur J Cardiothoracic Surg.* 53(6) 1112-1117) (arguing for application of PSM in observational studies where the number of outcomes is low and the number of confounding variables high).

[18] Higgins, G. E., Jennings, W. G., Jordan, K. L., & Gabbidon, S. L. (2011). Racial profiling in decisions to search: A preliminary analysis using propensity-score matching. *International Journal of Police Science & Management*, 336-347.

[19] Ridgeway, G. (2006). Assessing the effects of race bias in post-traffic stop outcomes using propensity scores. *Journal of Quantitative Criminology*, 1-29.

regressions for comparative purposes, in order to help ensure against a misleading negative result.

I.   Investigative Stops – Data

The data population selected for purposes of examining outcomes relating to investigative stops comprised a total of **19,544** *Terry* stops reported over a two-and-a-half-year time (2016 through mid-2018). Controlled variables were selected based upon the information available in the Department's Data Analytics Platform (DAP), which includes not only the information available in the *Terry* templates but additional information relating to characteristics of the officer. Variables that were controlled in the match are presented in Table 1. These variables were selected for their theoretical effect on the outcome. Note: Again, a more complete discussion of aggregated data associated with many of these variables is presented in the Department's Annual Stops and Detentions Reports for the years in question; the data itself is available for download or exploration through the Department's Terry Stops Dashboard.

Table 1: Controlled Variables – Stops

Event (Stop)	Officer Demographics	Subject Demographics
Call Type	Gender	Gender
Priority	Race	Age
Sector	Age (at stop)	
Reported Year	Years of Experience	
Reported Month	CIT Certification	
Reported Weekday	Title	
Reported Hour	Squad Assignment	

A. Event Data

Call Type:[20]   All activities begin with either a call from the public (call for service, or CFS) or an officer-initiated action in response to activity or behavior they observe. Broadly, stops are characterized (call-typed) as either *dispatched* or *on-viewed*, depending on how they originated. In total, 68.5% of stops in this data set were associated with a CFS from the public; 27.5% were associated with an on-viewed incident. In either case, whether reasonable suspicion for a stop

---

[20] Approximately 4% (*n = 793*) of stops could not be related to the record used to identify call type. The SPD Computer Aided Dispatch (CAD) system logs and tracks all calls, whether officer initiated or the result of a CFS. The logistic regression underlying the PSM procedure requires either imputation of missing values or elimination of the observation from the analysis. There is no way to reconstruct or otherwise impute the call type. The 793 stops with missing CAD information (e.g. Call Type, Priority) were eliminated to control for missing data.

is relayed by a member of the public or a matter of officer observation, the decision of an officer to initiate the stop is ultimately subject to officer discretion. Generally, the distribution of dispatched and on-viewed stops remains stable in this data set when viewed across subject perceived race (even in demographics represented at or below 5%, where lensing, or exaggeration of a rate, is often a concern).

Priority: As part of the call taking and dispatch process, information available from the community member calling (dispatch) or the officer (on-view) is used to determine both *initial call type* and *priority*.   While a total of 205 distinct case types were observed during the study period, and while PSM can be a powerful tool for the balancing of many variables, for purposes of assuring adequate match this project utilized a higher level of aggregation – call priority – as a dimensionally scaled[21] representation of case type.

Call types are prioritized according to level of emergency. Priority 1 calls are incidents that require an immediate response, including incidents that involve obvious immediate danger to the life of a citizen or an officer. Priority 2 calls are noted as urgent, or incidents which if not policed quickly could develop into a more serious issue (such as a threat of violence, injury, or damage). Priority 3 calls are investigations or minor incidents where response time is not critical to public safety. Priority 4 calls involve nuisance complaints, such as fireworks or loud music. Priority 7 calls are officer-initiated events, such as traffic stops. Priority 9 is used to indicate administrative tasks or downtime.

The clear majority (more than 80%) of active calls (events requiring some level of officer response) were classified as either a Priority 1, 2 or 3.   Priority 7 calls comprised just over 12% of the sample. The remaining 7% of stops were observed across a variety of administrative types.

Sector: Sector is included as a unit of both patrol deployment and optimal[22] representation of geography. The SPD map is broken into five (5) Precincts, seventeen (17) Sectors and fifty-one (51) Beats. Squads are deployed to Sectors and individual officers or teams of officers (two person cars) are assigned to be primary in their "districts" or Beats. Sectors were found to represent the geographic variability of the city and aligns with units of management under the patrol deployment strategy.[23] In addition to controlling for neighborhood level characteristics, the use

---

[21] Caution must be exercised when matching too many variables. Risk of overfitting the model or regression to the mean applies. The Event Per Variable (EPV) ratio of 10 heuristic was used. See Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology*, *49*(12), 1373-1379 (finding that for values of EVP "10 or greater, no major problems occurred.").

[22] Beat, the most granular unit of patrol deployment, was piloted. R (the software package for running logistical regressions for purposes of achieving matched sets) performed poorly rendering balance tables when Beat was included. Sector, as the next highest level of aggregation was determined to be sufficient without sacrificing geographic variability.

[23] Similar to Ridgeway's (2006) findings in Oakland, this study represents citywide measures of disparity. Although Sectors represent squads, the PSM controls for squad level management variability. A study of disparity at various levels of management and supervision may yield additional insight into individuals or groups of individuals.

of Sector balances the effect of temporary assignment to a different span of control (e.g., an officer who is assigned to an administrative unit may work overtime in patrol; their administrative assignment may not be relevant but the geographic location of the stop and temporary span of control is.

Date/Time: Three date/time variables or combinations of variables can be related to stops. The CAD system applies date/time stamps at several points as an event is processed (e.g. queued, dispatched, first arrived, arrived, cleared, etc.). The officer indicates the "occurred" date and time on the report used to document the stop, and the RMS logs the date/time the report is submitted. Although the occurred date/time would ideally be split into its date parts for this analysis, implementation of the original report template[24] resulted in an unusable and inconsistent format for occurred date and time (free text). Further complicating this analysis is that "original time queued," which is the preferred CAD event date/time, is often reflective of the report of an offense or crime under investigation, but is not necessarily the time of the stop itself. For this reason, date and time were matched according to reported date (year, month, day of week) and Officer shift.

B. Officer Demographics and Assignment

Officer Ordinance Title:[25] During the study period, SPD employed 1,614 sworn officers; of those, 902 reported an investigative stop during this period. Over three-quarters (77.3%) of all stops were reported by officers holding the ordinance title of "Police Officer," followed by "Police Student Officers." Both Police Officers and Police Student Officers are most commonly assigned to the Operations Bureau, in Precincts, in Patrol or 911 response functions. Probationary officers - those within one (1) year of hire, between the end of Field Training (e.g. student officers) and permanent status, comprised another 8.7%.

Officer Gender: Approximately 85% of sworn officers were male; 15% were female. Male officers reported 88.9% of all stops; female officers reported 11.1% of stops. In eight (8) stops,

---

[24] Versaterm, SPD's current RMS, allows for customized "templates" to be created. Ad hoc implementation of these templates is often limited in their ability integrate with other RMS structures and utilities, the fields available (date/time stamps versus free text fields) and the format of the report once processed and stored (i.e. xml). The NRMS, scheduled to go-live on March 31st, 2019, natively integrates the investigative stop report as a "field contact" report.

[25] Ordinance titles are the City of Seattle permanent job titles for a position.

officer gender could not be accounted for, and those stops were accordingly omitted from the analysis.

Officer Race: Officer race is obtained at the time of hiring and reported under the standard used to transmit employment eligibility verification to the Department of Homeland Security (DHS), Form I-9. Overall, nearly 80% of all stops were conducted by officers identifying as "White." Officers identifying as "Hispanic or Latino" accounted for approximately 5% of all stops. Black officers accounted for almost 5% of stops. Asian officers accounted for almost 4% of stops. All other officer race demographics accounted for less than 5% of stops. The same eight (8) stops where the officer gender was missing were identified in the race demographic data and were thus controlled.

Officer Years of Experience: Officer Years of Experience (YOE) is calculated as the difference between the officer's hire date and the date of the stop. The average[26] YOE was an asymmetrical (SD = 7.19) 6.3 years. Overall, 16% of all stops were conducted by officers with between one and two YOE.

Officer Age: *Officer* age is calculated from officer date of birth (DOB) and the reported date of the stop. The average age of officers reporting stops was 35, along a symmetrical, approaching left-modal curve (SD = 8.5). Approximately one-third of all stops (30%) were reported by officer between 30 and 35 years of age. Officers between 25 and 30 accounted for 23.2% of stops; officers between 35 and 40 years of age accounted for 15.4% of stops. In all, officers between 25 and 40 reported nearly 70% of all stops.

Reporting Squad by Bureau: The functional organization of the SPD contains seven levels of management, each of which is defined by its relationship to or distance from the Chief of Police (COP). The broadest level of structured management[27] below the COP are the Bureaus, each of which is led by an Assistant Chief. Precincts/Sections roll up under Bureaus and report to either Captains or civilian managers. Within the Operations Bureau, out of which most stops are reported, geographic precinct areas are administered by Captains; the Patrol/911 response

---

[26] Average is reported here for information only. PSM is a non-parametric technique and is not dependent on satisfaction of distributional assumptions.

[27] Virtual structures exist immediately below the COP. These structures are led by the Deputy Chief and civilian Executive Directors.

function, which maintains 24-hour coverage, is separated into Watches, led by Lieutenants.[28] Squads are the most granular units of management and are administered by Sergeants.[29]

In total, 130 squads reported investigative stops during the study period. Of these, 78.3% reported to the Operations Bureau; 10.1% reported to the Investigations Bureau; and 9.3% reported to the Homeland Security and Special Operations Bureau.

With 130 squads, adding each squad as a discrete variable was impractical and tested the demonstrated validity of the PSM method. Squad groups were accordingly constructed by grouping squads that share similar functions to dimensionally scale the squad variable. Units assigned to 911 response assigned units reported 87.6% of stops. Officers assigned to special "beats" or emphasis patrols and bikes reported 8% of all stops. Tactical units, including the Anti-Crime Team (ACT) and SWAT, reported just 2.7% stops; another 1.7% of stops were reported by "Other" units, primarily units with an investigative function (Gang, Narcotics, etc.).

Officer Crisis Intervention Certification Status: In addition to descriptive characteristics of the officer, officer certification as a member of the Crisis Intervention Team (CIT) was included in the match. It was hypothesized that holding and maintaining certification as a member of the CIT would constitute both an officer attitude and advanced training which would affect an officers' decision-making and thus the outcome of the stop. In 9,116 stops (51.7% of the total), the reporting officer was CIT-certified.[30]

C. Subject Demographics

Under Washington State law, a community member who is subject to an investigative stop is not generally required to identify themselves. By SPD policy, "officers may request identification; however, subjects are not obligated to provide identification or information upon request." (Seattle Police Manual, 6.220).   When reporting a stop, officers are accordingly asked to

---

[28] Functional watches are differentiated from the temporal watch period. SPD maintains twenty-four-hour coverage of 911 response and some specialty unit functions. 911 response / patrol squads (the most granular unit of organization) are organized around six (6) nine-and-a-half (9.5) hour overlapping watches in three eight hour periods, roughly: 0300 to 1100, 1100 – 1900 and 1900 – 0300, the next day (1st, 2nd and 3rd Watch, respectively).

[29] Units are an intermediate level of management, between Precinct/Sector and Squad. Units are led by Lieutenants and, while all watches correspond to Units, not all Units occupy a standard watch, as described above.

[30] CIT certification is voluntary and requires 40 hours of specialized training and self-selection in the certification group. Some officers may have received the specialized training but are no longer part of the CIT certification group.

document their "perception" of the race, age and gender of the subject they contact. When the *Terry* template for reporting was constructed, it was believed that prompting the officers for their perception would render a data point closer to their decision-making and could aid in analysis, such as this. Confounding factors influence the consistency of this assumption, however; for example, although the officer is being asked for their perception, their response may be influenced by information they obtained during their investigation. An officer may have been exposed to statements by the subject or been provided voluntarily with identification indicating the subject's self-reported identifying characteristics. If an investigative stop produces probable cause leading to an arrest and booking, the officer may have confirmed identification of the subject from biometric scans conducted while processing into a Department of Juvenile and Adult Detention (DAJD) facility.

Overall, slightly more than 50% of all stops involved a subject the officer reported perceiving as "White." Subjects perceived to be "Black" comprised another 30.6%. All other perceived demographic groups, comprising in total 13.2% of stops, each represented less than 5% of the distribution. In 4.4% of stops, the officer reported the subject's perceived race as "Unknown." Stops involving subjects reported to be of "unknown," "missing" or "other" perceived race are included in the analysis, but only as part of the White/Non-White binary group. See Table 2. Blacks are vastly overrepresented in the stop data relative to the 2010 U.S. Census or the 2012-2016 American Community Survey where they are estimated to comprise 7.7% and 7.0% of Seattle's population, respectively.[31] While some of this 4-fold disparity in stop rates could be due to non-racial factors, given the magnitude of the disparity it will receive considerable focus in the next report.

Table 2: Stops by Perceived Subject Race

Race	Stop Count	Percent of Total
White	9,873	50.5%
Black	5,986	30.6%
Hispanic	937	4.8%
American Indian/Alaska Native	647	3.3%
Asian	548	2.8%
Multi-Racial	360	1.8%
Missing/Other/ Unknown	1,139	6.1%

Subject Age: Officers indicate perceived subject age by category. Overwhelmingly, the majority of subjects are perceived to be within the age brackets of 18 to 45 years of age, as shown in Table 3.

---

[31] https://www.seattle.gov/opcd/population-and-demographics/about-seattle#raceethnicity

Table 3: Stops by Perceived Subject Age[32]

Age	Stop Count	Percent of Total
1-17	892	5.06%
18-25	3,862	21.9%
26-35	5,921	33.59%
36-45	3,550	20.14%
46-55	2,126	12.06%
56 and above	759	4.31%

Subject Gender: Officers are provided the option to indicate perceived subject gender as male, female, and unable to determine. Overwhelmingly, the majority of subjects are perceived to be male. See Table 4.

Table 4: Stops by Perceived Subject Gender

Gender	Stop Count	Percent of Total
Male	13,678	77.59%
Female	3,679	20.87%
Unable to Determine	148	0.84%
Missing	123	0.70%

II.   Investigative Stops – Matching

Five separate matches were conducted: White/Non-White[33], White/Black (WB), White/Hispanic White/Asian, and White/American Indian Alaska Native. In each case, the match balanced the variables across the classification (treatment/control), where a minority race (Asian, American Indian Alaska Native, Black, Hispanic) is the treatment and majority race (White) is the control. After controlling for missing data, 17,628 observations (94% of the universe) were involved in the match. Valid, balanced matched populations were generated across all five treatment/control conditions. See Table 5.

---

[32] This table, and the analysis, exclude 518 stops where the subject age was not indicated.

[33] Subjects reported to be of "unknown," "missing" or "other" perceived race by officers.

Table 5: Match Summary – Stops

Treatment Group	Matched Pairs	Matched Population	Percent Matched
Non-White	7,081	14,036	80%
Black	4,790	9,580	88.8%
Hispanic	842	1,684	98.7%
American Indian/Alaska Native	598	1,196	99.5%
Asian	479	958	98.5%

Across the five observation groups, the match rate ranged between 80% and 99.5% of the treatment groups' representation in the population of complete observations (*n= 17,628*). Logically, a smaller number of observations resulted in a reduced number of permutable combinations of variables in smaller subsets (e.g. Asian, American Indian Alaska Native, Hispanic).

The availability of observed control permutations (e.g. stops involving control subjects) resulted in a high match rate. Unsurprisingly, the highest match rate was observed in the American Indian Alaska Native match, where only 601 control matches were possible. A nearly 15:1 control (*n = 8,861*) to treatment (*n = 601*) ratio resulted in 99.5% (*n = 598*) of stops involving a subject perceived to be Asian finding a control match. Predictably, the lowest match rate (80%) was observed in the Non-White treatment.

The goal of the PSM is to replicate, to the extent possible, the controlling effect of randomization by balancing the confounding effects of variables theoretically likely to affect an outcome. Several "model fit" statistics have been proposed for PSM; however, generally the standardized mean difference (SMD) is considered to be the most reliable diagnostic for measuring a treatment effect.

Table 6: Balance Summary - Stops

Treatment Group	Variables Balanced	Variables Not Balanced (>.1)	Maximum Balanced Variable	Balance (Difference Adjusted)
Non-White	90	0	Officer Age	-.017
Black	89	0	Officer Age	.012
Asian	87	0	Hour	-.066
American Indian/Alaska Native	87	0	Officer Age	.039
Hispanic	87	0	CIT Certification (No)	.032

Table 6 shows a summary, for each treatment group (subject race), of the number of balanced variables, the maximum balanced variable, and the adjusted difference. For each match, a balance table was generated to assess the quality of the resulting matched population.

III.   Use of Force - Data[34]

During the study period, 4,371 uses of force were reported. More than three-quarters (76.8%) of force was classified at the lowest level, Type 1. One-fifth of force (21.5%) was classified at the next highest level, Type 2. Cumulatively, 98.3% of all force was reported at these two levels, comprising complaints of "transitory pain" and firearms pointing (Type 1) and minor cuts, bruises and strains (Type 2). 1.7% of force was reported as Type 3 or Type 3 – OIS (Officer Involved Shooting). These highest classifications of force represent reports of serious injury and lethal or potentially lethal force. See Tables 7(a), (b);

Table 7a: Use of Force by Year – Percent of Total

Use of Force by Type	Percent of Total			
	2016	2017	2018	Total
Type 1	74.8%	76.5%	80.4%	76.8%
Type 2	23.7%	21.6%	18%	21.5%
Type 3	1.2%	0.7%	1.5%	0.6%
Type 3 - OIS	0.3%	1.2%	0.2%	0.6%

Table 7b: Use of Force by Year – Total Count

Use of Force by Type	Total Counts			
	2016	2017	2018	Total
Type 1	1,202	1,291	864	3,357
Type 2	381	364	193	938
Type 3	20	12	16	48
Type 3 - OIS	5	21	2	28

In contrast to the reporting of race in relation to investigative stops, the race of a subject in a use of force report is, generally, not limited to the officer's perception.[35]   Force often involves physical restraint and some limited custody of the subject, if not a full booking into a King County Department of Juvenile and Adult Detention (DJAD) facility. The process of booking a subject involves reporting, to DAJD, a subject identity (e.g. the detainee's self-report or government

---

[34] A The distribution of more complete discussion of SPD's use of force data is presented in the Department's Annual Use of Force Reports for the years in question, and the data are available for download or exploration through the Department's Use of Force Dashboard.

[35] Where force is used to disperse a crowd, for example, or apprehend a fleeing subject believed to pose a danger, there may not be an identifiable subject associated with the report.

identification). Additionally, DAJD may utilize biometric data (e.g. fingerprints) to confirm a subject's identity during intake.

Roughly 40% of all reported force involved a white subject; 32% of subjects were Black. Approximately one-fifth (20.1%) of subjects were listed as "Not Specified" (in contrast to the cumulative 5.7% of subjects of stops).[36] This analysis treats Not Specified as a discrete treatment group. Asian subjects (3.5%) were combined with Nat Hawaiian/Other Pac Islander (.8%) for cross-analysis consistency, for a cumulative 4.4%. Hispanic subjects were indicated in 3.4% of force.

Only demographics representing, cumulatively, more than 100 observations were handled as treatment (match) groups. American Indian / Alaska Native subjects (AIAN) were involved in 37 (.8%) uses of force and are represented in the White/Non-White match, only.

Of the 3,357 Type I uses of force reported over the study period, 1,049 involved the pointing of a firearm. A demographic breakdown of these uses of force is shown in Table 8.

Table 8a: Distribution of Force by Subject Race

	White	Black	Hispanic	Asian	American Indian/Alaska Native	Nat Hawaiian/Pac Islander	Not Spec.
Type I	1349	1045	118	119	24	29	637
Type II	337	326	29	31	12	2	201
Type III	20	18	1	1	1	2	6
Type III OIS	18	9	0	0	0	0	0
Total	1724	1398	148	151	37	33	880

Table 8b: Distribution of Type I, Firearm Pointing by Subject Race

Subject Race	Count	Percent of Total
White	312	29.7%
Black	402	38.3%
Hispanic	59	5.6%
Asian	49	4.7%
American Indian/Alaska Native	9	0.9%
Nat Hawaiian/Pac Islander	9	0.9%
Other/Missing/Unknown	209	19.9%

Variables controlled for in the analysis are presented in Table 9; descriptions provided above apply.

---

[36] Unknown and Missing subjects were assumed to be minority and included in the Non-White treatment group.

Table 9: Controlled Variables – Use of Force

Event	Officer Demographics	Subject Demographics
Call Type	Officer Gender	Subject Age
Priority	Officer Race	Subject Gender
Sector	Officer Age	
Occurred Year	Officer Years of Experience	
Occurred Month	Squad Type	
Occurred Weekday	Officer Title	
Occurred Hour	CIT Certification Status	

IV. Data Matching – Use of Force

Propensity scores were generated using logistic regression for each treatment group (black, Hispanic, Asian, and non-white and matched to the nearest propensity score in the control group (white).   Five matches were conducted: white/non-white, white/black, white/not specified, white/Hispanic, and white/Asian. Because of the extremely low number of available observations, Native American/Alaska Native was not included as a treatment group. In total, 14,371 observations were involved in the match. Valid, balanced matched populations were generated across all five treatment/control conditions. See Table 10.

Table 10: Match Summary – Use of Force

Treatment Group	Matched Pairs	Population Matched	Percent Matched
Black	1,049	2,098	60%
Hispanic	139	278	94%
Asian	165	330	90%
Not Specified	740	1,480	84%
Non-White	1,569	3,138	94%

Across the five treatment groups, the match rate ranged between 60% and 94% of the treatment group's representation in the population of complete observations (n= 14,371).

While not every factor was matched, every variable was balanced and no variables were found to be out of balance, post-match. See Table 11.

Table 11: Balance Summary – Use of Force

Treatment Group	Variables Balanced	Variables Not Balanced (>.1)	Maximum Balanced Variable	Balance (Difference Adjusted)
Non-White	82	0	Hour of Occurrence	-.019
Black	80	0	Subject Age	-.040
Asian	79	0	CIT Certification (No)	-.079
Not Specified	81	0	Hour of Occurrence	-.027
Hispanic	78	0	Subject Age	-.078

RESULTS

Using PSM, with subject race as the treatment variable, SPD examined disparity in four separate outcomes across two areas of police/community interactions.   Outcomes examined in the context of investigatory stops were (1) whether the subject was frisked (frisk rate); (2) whether a weapon was recovered pursuant to a frisk (hit rate); and (3) the duration of the stop. The outcomes examined in the context of use of force was whether a subject had a firearm pointed at them (Type I) during the encounter.

I.   Investigative Stops

A. Frisk Rate

An officer may decide to conduct a light search ("frisk") for weapons of a subject of an investigative stop whom they have reasonable suspicion to be "armed and presently dangerous." See Seattle Police Manual, 6.220.

During the study period, frisks were reported in **21.58% (4,220)** of the 19,554 stops reported. The distribution of frisks, across subject demographics, is presented in Table 12.

Table 12: Frisk Rate by Subject Race

Subject Race	Stops	Frisk Rate
White	9,873	18.21%
Black	5,986	25.99%
Hispanic	937	27.32%
American Indian/Alaska Native	647	21.02%
Asian	548	25.91%
Multiracial	360	21.11%
Other/Missing/Unknown	1,139	24.34%

A frisk rate comparison between control (white) and treatment (non-white) group subjects is presented in Table 13.

Table 13: Frisk Rate Comparison (Treatment/Control)

Treatment Group	Frisks – Treatment (Non-White)	Frisks – Control (White)	Frisk Rate – Treatment (Non-White)	Frisk Rate – Control (White)	Difference (Points)
Black	1,183	1,002	24.7%	20.9%	3.8
Hispanic	229	177	27.2%	21%	6.2
American Indian/Alaska Native	127	122	10.6%	10.2%	.4
Asian	123	92	25.7%	19.2%	6.5
Non-White	1,636	1,385	23.3%	19.7%	3.6

Within the comparison of matched sets as to the decision to frisk a subject, some disparity can be observed across all five (5) matched groups. Controlling for other variables, and although Asians account for only a small fraction of stops overall, subjects perceived by the officer as Asian were frisked approximately 6.5 percentage points more than control (white) subjects. Subjects perceived to be Hispanic were frisked 6.2 percentage points more frequently; and subjects perceived to be black were frisked 3.8 percentage points more frequently. The lowest frisk disparity was observed in subjects perceived to be American Indian/Alaska Native, at 0.4 percentage points.

B. Hit Rate (Weapon Recovery During a Frisk)

A hit rate comparison between control (White) and treatment (Non-White) group subjects is presented in Table 14.

Table 14: Hit Rate Comparison (Treatment/Control)

Treatment Group	Hits – Treatment (Non-White)	Hits – Control (White)	Hit Rate – Treatment (Non-White)	Hit Rate – Control (White)	Difference (Points)
Black	191	247	16.1%	24.7%	8.6
Hispanic	42	40	18.3%	22.6%	4.3
American Indian/Alaska Native	26	42	20.5%	34.4%	13.9
Asian	21	20	17.1%	21.7%	4.6
Non-White	275	352	16.8%	25.4%	8.6

When frisked, White subjects (control group) were found with weapons (hit) more frequently than minority subjects (treatment). Subjects perceived to be Asian were found with weapons 4.6 points less than white subjects in the same situations. The largest disparity in hit rate was found in subjects perceived to be American Indian/Alaska Native, who were found with weapons 13.9 points less than White subjects.

While some caution is necessary given the relatively low number of observations, given the fairly low frisk rate within each treatment group overall, these findings are generally consistent with those reported by the Monitor in the Tenth Systemic Assessment, who found that despite being frisked at higher rates, Non-White subjects are less likely to be found to be in possession of a weapon.

C. Stop Duration

Officers note the time a subject was detained in a dropdown box on the stop report. These data are not ideal for analysis; the bins are not equal (some range five (5) minutes, some range ten (10) minutes and some are indefinite), and an officers' perception of time is subject to many factors. SPD's new RMS system, which again comes online this spring, captures the beginning and end time of the "seizure," thus enabling more accurate future analyses.

For the purposes of this analysis, a weighted mean was calculated by finding the midpoint of the bin, multiplying the number of observations in that duration group and dividing by the total number of observations.

Table 15: Weighted Mean Duration Comparison

Treatment Group	Treatment	Control	Difference
	Average Time (minutes)	Average Time (minutes)	
Black	11.0	11.3	-0.1
Hispanic	15.4	15.6	-0.1
American Indian/Alaska Native	10.9	10.5	0.4
Asian	11.1	11.9	-0.8
Non-White	11.1	10.8	0.3

II.   Use of Force – Firearm Pointing (Type I)

The decision to point a firearm can depend on many things – and in some circumstances, as a matter of training, it is required. However, firearm pointing is also closely related to officer decision-making and is near the genesis of the force event. The officer will decide to point a firearm, for example, based on a perception of an imminent threat of danger from the subject, if they believe the subject to be armed, or in response to a set of conditions calling for the use of lethal cover (felony vehicle stops, for example). Subject race is not a conscious part of that

calculus; no training or policy allows an officer to judge the safety of a contact based on the race or perceived race of the subject. Subconsciously, however, race may influence an officer's perception of safety, and thus the decision to use lethal cover. For this reason, the pointing of a firearm was also selected as an outcome measure for analysis.

The PSM match procedure balances the observations under an area of common probabilistic support but the quality of the inference relies on the appropriate selection of variables used in the match. While this procedure attempted to match the type of call the officers were responding to (priority and call type), it was unable to match all 205 detailed call types, without exceeding the demonstrated performance of the PSM procedure. Priority and Call Type (on-view/dispatch) were thus selected as a dimensionally scaled representation of the nature of the call. Neither priority or type are exclusive to each other or Call Type Initial and so are partially independent.

The comparison between groups with respect to the pointing of a firearm is presented in Table 16.

Table 16: Comparison of Treatment Groups – Pointing of a Firearm

	Pointing – Treatment (Non-White)	Pointing – Control (White)	Pointing Rate – Treatment (Non-White)	Pointing Rate – Control (White)	Difference (Points)
Non-White	384	291	12.2%	9.3%	2.9
Black	290	203	13.8%	9.7%	4.1
Asian	47	33	14.2%	10%	4.2
Not Specified	149	184	12.4%	10.1%	2.3
Hispanic	54	29	19.4%	10.4%	9.0

There is notable disparity in these data of incidents in which a firearm is pointed with the caveat that these are overall low percentages of incidents. The greatest disparity in firearms pointing was observed with Hispanic subjects, who had a firearm pointed at them in 19.4% of Type I force cases in which they were involved compared to Whites who had a firearm pointed at them in 10.4% of similar incidents. Subjects Not-Specified were observed with the lowest disparity in firearms pointing, 2.3 points. However, a preliminary review of associated use of force reports reflects that a majority of these instances occur in circumstances (e.g. felony warrant arrests, stolen vehicle (high risk felony) stops, shootings) where the pointing of a firearm is less a matter of officer discretion than a trained tactical expectation. Potential related factors not accounted for in the model, such as known suspect criminal history, local crime trends, and statements/behaviors expressed during the incident will be examined in the Phase II report.

CONCLUSION

This report is the first in a two-part series called for under the Sustainment Plan for SPD to further review disparity in its stops and detentions. In this report, SPD built upon the good work done by the Monitoring Team in the Tenth Assessment and employed the technique of Propensity Score Matching to account for potential confounding variables inherent in investigative stops to better understand the disparity observed with respect to frisks, hit rate, and duration of stop.

This report also meets one of the Monitor's recommendations in the Ninth Systemic Assessment on Use of Force in that it applies a more sophisticated approach to examining disparity in certain force outcomes – the decision to point a firearm.

Our main findings are that there are racial disparities in all three of the outcomes examined. Minorities are frisked at a rate 15% higher than whites, even when using PSM to match on observable characteristics.   For Hispanics, the frisk rate is 30% higher than whites.   Although stopped and frisked at higher rates, firearms are found 1/3 less frequently in frisks of minorities, relative to the matched sample of white detainees.   Additionally, minorities have weapons pointed at them by police far more frequently (30%) than white detainees.

In the second disparity report, due to be filed in December 2019, SPD anticipates using additional enhancements to the data set to allow for an even more precise estimation of disparity in the areas identified by this report: namely, frisk hit rates and the pointing of a firearm. SPD is presently evaluating additional variables that it might factor into the next iteration of this review, to be presented in the second Phase II report later this year. In light of continued improvements in SPD's data collection and analytic capabilities, SPD anticipates potentially being able to include data regarding recent crime trends in a neighborhood, officer's knowledge of a subject's history and prior weapon-carrying behavior, and/or the behaviors/statements exhibited during the encounter.

SPD also plans to engage the services of a national expert on measuring, analyzing, and responding to disparity in law enforcement practices. This work will be informed by the findings of the Phase II report. This partnership will be used to create, test, and implement a structured instrument for turning such data sources as behaviors and statements captured in body-worn video, officer narratives, and victim statements, into quantitative and qualitative data that can be analyzed in aggregate. SPD invites the input of the Office of the Inspector General and CPC in this process.

As more information becomes available about the dynamics of these interactions, it is SPD's intent to be able to isolate the key situational differences and begin the difficult work of determining which law enforcement and other social policies and practices could be adjusted to reduce the disparity. Ultimately, the SPD views this report as another powerful statement that it is committed to being an innovative organization committed to self-evaluation, continuous improvement, data-driven decision-making, which has earned the recognition of being a national leader in policing.