

RE: ACLU Matter vs. - REF# 1340012232

Response to Period 2 Comments

Ralph B. Taylor

Lallen T. Johnson

February 28, 2018

1 REMOVE CANCELLED ISRS – EXPLAIN HOW INCLUSION AFFECTED PREVIOUS

Our current understanding is that ISRs that were canceled (status code CLD) went to archived status (status code ARC).¹ Since no records were excluded based on any status code, the question arises how (if at all) did this affect the reporting for period 2?

The short answer is that we believe it did not affect:

- the legal narratives analysis
- the post stop outcomes analysis
- the ecological analysis
- the versioning analysis

Here is why. The identification of eligible single version ISRs, and multiple version ISRs in both cases relied upon a master list provided by the CPD. We are presuming – **and seek clarification from CPD certainly IF we are incorrect here** – that any canceled ISRs would not have been included in that master list. Because we relied on the master list **provided by CPD**, we are presuming that they had removed the cancelled records from this master list.

Nevertheless, going forward we will take the additional step of verifying that no included records had, at any point in time, a contact card status of CLD. We also will be sure this is the case before we send along any total numbers of records for the ecological analysis. We also will need to gain a better understanding of the difference between the two status codes.

¹ When we pulled in the “changes” file for ISRs for 2016, and selected only records dated July 1 and later, and cross-tabulated the report status variable (columns in the table below) with ufe status (rows in the table below), results were as shown. Of the initial 3,436 ISR versions labeled canceled (CLD) in ufe_status, 3,424 of them had a report status of archived (ARC). We still need to learn more about the operational uses of these two codes ufe_status and report_status.

```
. tab ufe_status report_status, missing
```

UFE_STATUS	REPORT_STATUS									Total
	APR	ARC	DEF	FIN	PRE	REJ	REV	SCN	SUB	
APR	51,349	0	0	376	0	0	0	0	0	51,725
CLD	0	3,424	0	0	0	0	0	2	0	3,426
PRE	0	0	0	0	39,008	0	0	0	0	39,008
REJ	0	0	395	0	0	2,533	118	0	0	3,046
SUB	0	0	0	0	0	0	0	0	54,393	54,393
Total	51,349	3,424	395	376	39,008	2,533	118	2	54,393	151,598

2 FIND AND DOCUMENT PRE STATUS

Going forward, and doing analyses for periods 3 and 4, we will be alert to records where the contact card status is preliminary (PRE). We did not attend to this in Period 2, after the Versioning System was put into place and the “contact_card_status” field became important re: ISR status codes.

3 SEARCH AND PAT DOWN HIT RATES INCLUDING WEAPONS OR DRUGS WERE INCLUDED IN TABLE – ARE THEY SAYING THEY WANT AN ANALYSIS OF *THESE* HITS?

In the City and the CPD (“City”)'s responsive comments to the Consultant’s Draft of the Second Report and the Technical Reports by us, the City indicates that their retained experts (“reviewers”) raised concerns about the weapons search hit rates and pat down weapons hit rates, and how they were characterized. The City’s reviewers also pointed out that including other types of things found following a search or pat down does generate higher “hit rates.” We think these concerns, as they might apply specifically to our statistical reports, are somewhat misplaced and may result in part from the reviewers overlooking important information provided in our earlier statistical reports. That aside, however, their points do raise important questions for the parties to decide: *what* hit rates are of primary interest?

3.1 ALL POSSIBLE SEARCH AND PAT DOWN HIT OUTCOMES WERE REPORTED DESCRIPTIVELY

Several different varieties of patdown hit rates and search hit rates **were reported descriptively** in our statistical reports. We include in appendix A Table 3 from page 18 in the post stop outcomes report. It shows we reported six types of hit rates:

- patdown leading to a hit on weapons,
- patdown leading to a hit on drugs,
- patdown leading to a hit on anything,
- search leading to a hit on weapons,
- search leading to a hit on drugs,
- search leading to a hit on anything.

Those hit rates range anywhere from 1.9 percent (patdown resulting in drugs) to 18.8 percent (search leading to anything).

Although the only two hit rates analyzed were weapons/firearms produced by pat downs, and weapons/firearms produced by searches, the rates for all these different types of "hits" were provided.

3.2 SELECTING “HIT” OUTCOMES GOING FORWARD

What are the parties suggesting? Would they like to see analyses of more search hit rates? More patdown hit rates? If so, which specific ones? We anticipate there will be discussion

among all parties about which specific hit rates to investigate going forward. As input to that discussion, we propose the following recommendations be given consideration.

- 1) For the sake of the public accounting function of these reports, we recommend continuing to include descriptive information for the six outcomes mentioned in table 3 and included as an appendix here. That way, the different prevalence rates can be tracked descriptively over time.
- 2) If parties recommend, starting with Period 3 / Period 4, we can report those six prevalence rates separately for each of the three ethnoracial groups, along with the standard errors associated with those rates. Doing so might enhance the public accounting function going forward.
- 3) When thinking about expanding the outcomes subjected to statistical analysis, it would be unwise analytically to conduct a separate analysis for each of the six outcomes without making statistical adjustments. Those adjustments are required given the statistical interdependence of the different outcome. Such adjustments would amount to making it "harder" to find statistically significant differences between different ethnoracial groups.² One party might like this, but another might not. Our recommendation would be something different. If possible, parties would agree on one search "hit" outcome and one pat down "hit" outcome ***in which they were most interested***.
- 4) If 3) is not feasible, we recommend parties consider two conceptually independent search/pat down "hit" outcomes going forward: weapons ***or*** drugs. That allows us to continue the weapons-hit statistical analyses going forward, which is helpful for public accounting. And, the added drug hit variable is not actuarially confounded with the weapons hit outcome, although empirically there may be correlations between weapons hit and drugs hit outcomes which require adjustments.

We look forward to receiving focusing input on this question.

4 WEAPONS THROWN AWAY

Weapons thrown away were not analyzed in the period 2 reports as a "hit." Officers had to check either the pat down weapon hit checkbox, or the search weapon checkbox for us to count a weapon as a "hit."

5 DISCRIMINATION AND DIFFERENT RACIAL LINKS WITH PAT DOWN HIT VS. SEARCH HIT

The City's experts, starting on p.8 of their comments, and as mentioned by the City on p.10, raise essentially two problems regarding interpretations of discrimination in post stop outcomes where

² Technically, what would happen would be that the significance level (e.g., $p < .05$) would be adjusted downward using some type of adjustment protocol (Aickin & Gensler, 1996). There are different ways to make this adjustment. Using the simplest type of adjustment (Bonferroni) one would proceed as follows. If one had 6 outcomes, and an original significance level of $p < .05$, the new significance level is $.05/6$ which is $p < .0083$.

selection is involved; more specifically, pat down hit rates and search hit rates. Multiple issues may be conflated in the comments. We attempt to disentangle.

5.1 NON ETHNORACIAL FACTORS DO SIGNIFICANTLY PREDICT PAT DOWN

The first point is that the City's reviewers suggest (p. 8) "if the data cannot tell us much about why police officers chose to patdown (say) one Hispanic citizen and not another, the data will also by definition have difficulty telling us much about why police decided to pat down a Hispanic citizen rather than a white citizen (or vice versa)."

In contrast to their suggestion, the non-ethnoracial predictors *can* tell us why officers choose to pat down one Hispanic citizen and not another. *Multiple non-ethnoracial factors, after controlling for race and ethnicity, are important.*

See, for example, Table 26, p.58, in our post stop outcomes report. Numerous non-ethnoracial predictors **did** significantly predict whether a detainee would be selected for a pat down. In the interest of economy we did not show those here, nor did we show them in the prediction of the pat down itself (Table 18, p. 43).

The reviewers seem to be inferring that our space saving tactic in table construction meant non ethnoracial factors could not predict whether a pat down occurred. This inference is not correct. Footnotes to both these tables do mention the additional detainee and context predictors.

See, for example, Appendix B. That lists all the variables used in the Heckman probit selection model to predict whether a pat down took place when only detainee race and ethnicity were used to predict whether the pat down led to a hit.

All significant predictors unrelated to race or ethnicity of detainee, and unrelated race or ethnicity of the context, which have a significant net impact on whether a pat down took place are highlighted in yellow. Over a dozen detainee or context factors unrelated to race and ethnicity or race and ethnic composition do significantly predict whether a pat down took place. Therefore, the inference that the reviewers are making is not correct. Non-ethnoracial factors do significantly predict whether a pat down took place, after controlling for ethnoraciality of context and detainee.

Perhaps the city's reviewers would feel more comfortable going forward if, when we do use selection models for the outcomes dependent on selection – pat down hit rates and search hit rates – we report fully on the selection portion of the model as well as the main portion of the model and list all predictors? This will require additional table preparation effort.

Nevertheless, please advise if this is requested. Such a request would significantly expand the size of the included tables for these analysis. Nevertheless, if it helps avoid misunderstandings like this perhaps it is worth it.

The assumption we are making – and we need to hear from the reviewers if they differ and if so on what grounds – is the following: because these impacts of non-ethnoracial predictors are significant *while controlling for ethnoracial features of the detainee and the context, they are telling us about “the difference within race / ethnic groups.”*

5.2 THE DIFFERENT RACE LINKS BETWEEN PAT DOWN HIT AND SEARCH HIT MEAN DISCRIMINATION CANNOT PLAY A ROLE

City's reviewers (p. 9) label as "internally inconsistent" the descriptive finding that ethnoracial differences on pat down weapons hit rates line up differently than ethnoracial differences on search hit rates.

Please note, however, we do not make the case that race or ethnicity of detainee plays ANY roles for this outcome. See Table 26, p. 58 in the post stop outcome report. There are no significant impacts of detainee race or ethnicity on pat down hit rates.

In other words, neither race nor ethnicity of detainee, once selection has been taken into account, plays any significant role in shaping the pat down hit outcome. We say this quite specifically on p. 57:

Results tell a simple story. In each sample, in every model variety, neither detainee race or ethnicity, nor racial or ethnic composition of the beat, significantly affects whether the pat down generated a weapon.

So we are **not** suggesting that, once selection and other factors are taken into account, there are net ethnic or racial impacts. Therefore, the different ethnoracial orderings on search hit rates and pat down hit rates are not necessarily problematic or inconsistent.

5.3 JUST DESCRIPTIVE

Finally, the point in the report the City reviewers highlight (on p. 9) is on our p. 26. A descriptive table with the ethnoracial differences they cite appears on p. 25. We emphasize on the following portion (page 26) of our report:

Speaking ***just descriptively***, pat downs of White non-Hispanic detainees were most likely to produce weapons/firearms (4.2 percent) whereas pat downs of Hispanic detainees were least likely (3 percent). The pat down weapon hit rate was intermediate for Black non-Hispanic detainees patted down.

This ordering of the three ethnoracial groups on pat down weapons hit rates has shifted from the first half of the year (Taylor & Johnson, 2017: Table 12) where the ordering was White non-Hispanic (4.1 percent), Hispanic (2.9 percent) and Black non-Hispanic (2.3 percent). This is just noted *descriptively*.

In other words, we are not conducting a hit rate analysis here. We are not attempting to document that these differences are significant. We are "speaking just descriptively."

5.4 THE CITY'S TAKEAWAY

The City's takeaway is that "the fact that nearly 10% of all ISRs created during the second reporting period have an inventory number associated with them provides yet another reason to discount the results of the 'hit rate analysis' prepared by the consultant's experts." (p. 11 city response). Two comments in reply.

First, the only analysis that we could even begin to defend as a hit rate analysis would be table 26 on page 58, where we control for selection, and control for a wide range of non-ethnoracial factors. **Note as discussed above this generates no significant ethnic or racial differences.**

Second, it is an open empirical question whether the factors, which, after controlling for selection, predict a weapons hit from a pat down, would be the same factors which predict hits defined more broadly; that is, hits defined as firearms *or* drugs *or* paraphernalia. **These are just two different – albeit related – outcomes. Because we did not analyze the broader range of possible patdown hits does not mean that the results shown here based on patdown hits of weapons should be "discounted." The reviewers are just asking about different outcomes.** A different outcome raises a question of external validity to a different outcome which is, as it must be, always an empirical question awaiting investigation (Taylor 1994: 164).

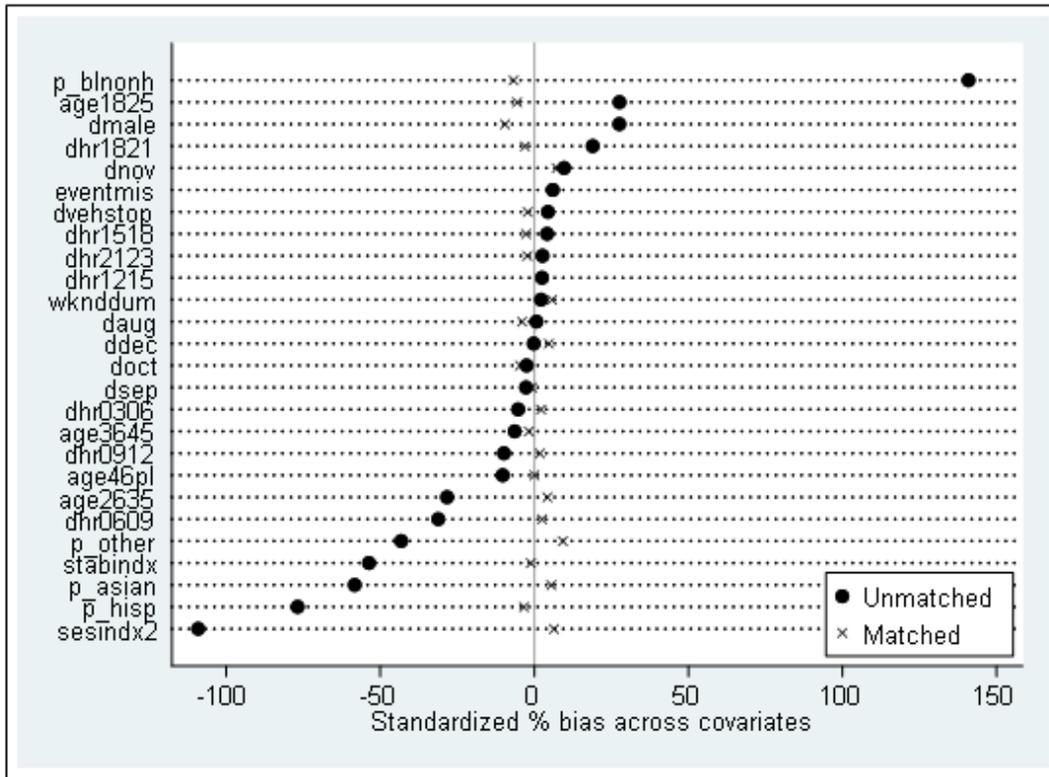
6 DISPARATE TREATMENT AND “MYRIAD FACTORS”

Although we do not presume to weigh-in on the legal discussion around the disparate impact question (City response, P.18, IX), in light of the "myriad factors" comment (p. 19), we want to point out how some of our analyses were specifically geared **to addressing this exact concern.**

6.1 OBSERVED FACTORS

We include immediately below figure 3 from the propensity score models for the patdown outcome on p.46 of the post stop outcomes report. Please bear with us a moment as we seek to explain what this shows about the "myriad factors" concern.

Figure 3 Covariate bias before and after matching: Black non-Hispanic and White non-Hispanic Sample 1 detainees



Note. Source: CPD ISR data. Period 2, July-December 2016. Two groups are Black-non-Hispanic detainees and White non-Hispanic detainees. Matching used caliper matching within .06 of a standard deviation. Results from Sample 1 only. Sample 2 results closely comparable.

Here is how this chart is to be read. Each row indicates a feature of the stop, or the detainee, or the stop context. The rows represent, reading down:

- the percent Black non-Hispanic residing in the beat where the stop to place
- whether the detainee was between the ages of 18 and 25
- whether the detainee was male
- whether the stop took place between 6 PM and 9 PM
- whether the stop took place in the month of November
- whether the event variable was missing for the stop
- whether it was the vehicle stop
- whether it occurred during specific time blocks (1500 to 1800, 20 100-2300 and so on)
- whether it took place on the weekend
- whether it took place in August
- whether it took place in December
- whether it took place in September

- whether the detainees have seen the ages of 36 and 45, or 46 and older, or 26 to 35
- percent other ethnoracial living in the beat
- residential stability index in the beat
- percent Asian living in the beat
- percent Hispanic living in the beat
- socioeconomic index in the beat

For each row, the location of the black circle on the horizontal axis indicates the standardized bias when the Black non-Hispanic detainees are compared with *all* the White non-Hispanic detainees in the first random sample. Some of the standardized biases are quite sizable. For example, the standardized bias associated with the residential stability index is -50 percent. This means that black non-Hispanic detainees as compared to all White non-Hispanic detainees had standardized scores that were quite lower.

Clearly these two ethnoracial groups score quite differently on numerous factors. If the standardized bias is greater than |10| percent, it is sizable.

The corresponding "X" on each line shows how much standardized bias there is when the Black non-Hispanic detainees are compared only to **matched** White non-Hispanic detainees. As one can see, all of the standardized biases are now less than |10| percent. **This means that, with regard to all of the factors listed here, the Black non-Hispanic detainees and the white non-Hispanic detainees are similarly situated.**

In other words, the “myriad factors” listed here have been taken care of and can no longer serve as potential alternate explanations for any observed significant race differences on the outcome.

6.2 UNOBSERVED FACTORS

How about possible differences between the two groups of detainees for which we do not have indicators, and thus are not measured? Could the Black non-Hispanic detainees in the sample and the matched White non-Hispanic detainees be different **on** unobserved factors?

Certainly. The question becomes: how different would they have to be on these other unknown factors to invalidate whatever significant race differences appeared here? The answer to that question appears in table 20 in the post stop outcomes report. It is reproduced below.

Table 20 Sensitivity analysis, propensity score models, pat down outcome, Black vs. White non-Hispanic civilians

		Gamma (Γ) value where race impact becomes non-significant
Caliper match = .06		
	Sample 1	1.15
	Sample 2	1.30

We also repeat immediately below our explanation of this from P.47 of the same report. We now place in bold the important conclusion.

In the first random sample if two individuals who are similar on the observed covariates differ in their odds of being Black and non-Hispanic versus White and non-Hispanic by only about 15 percent, then there is no significant impact of race on the pat down outcome. Given that this value of gamma (Γ) is relatively close to 1.0, the significant race impact seen is "sensitive to unobserved selection bias" (Aakvik, 2001: 30). Results from the second random sample are less sensitive to unobserved selection bias because the two individuals similar on observed covariates would have to differ in their odds of being Black vs. White by at least 30 percent.

In sum, especially in the first random sample, unobserved selection bias is a potential alternate explanation of the differences observed. Stated differently, there was something else going on, some feature of the detainee, the stop itself, the stop context, or all three, that is not captured by the variables used here, that might be "behind" the significant race differences observed.

In sum, we have explicitly acknowledged and examined ways that "myriad factors" could be operating. We have controlled for myriad factors that we could observe. We have also found that *unobserved* myriad factors could alter our conclusions and clearly stated the likelihood of that.

6.3 CLOSING IN ON SIMILARLY SITUATED STOP EVENTS?

Pulling the lens back and thinking more broadly, what we're attempting to do here is to examine more and more closely the disparate impact question. **Controlling for various factors does not imply intentionality.** It merely pursues the following question sequence.

- Is a raw ethnoracial difference observed?
- And, if so, does that ethnoracial difference remain statistically significant in multiple independent samples after focusing only on **similarly situated** events and detainees and contexts from two different ethnoracial groups?
- And, if so, how much influence would unobserved factors have to exert to make an observed statistically significant difference between similarly situated samples go away?

In short, we are **progressively refining – applying different levels of stringency to the comparisons – to take away more and more possible alternate explanations for any observed racial differences** which may or may not qualify as evidence of disparate impact.

Just because we are controlling for additional factors does not mean we are arguing along the lines of disparate treatment. Rather, we are constructing in a more and more restrictive fashion the conditions under which disparate impacts might be observed. We are defining "similarly situated" more and more narrowly.

7 BEAT RATIONALE

7.1 ADDRESSING ZEROS

The Ecological Analysis of the second reporting period took place at the district level. We addressed the problem of zero-inflated ethnoracial-specific violent arrests in two ways in that analysis. First, we transformed the violent arrest variable by increasing all district monthly ethnoracial-specific violent arrest counts by 1. That approach did not change the distribution of the data, but allowed us to overcome technical modeling issues. Second, we performed a spatial empirical Bayesian smoothing technique to pull extreme values closer to each district's neighboring group average.

It is the City's argument that drilling down to the beat level will exacerbate the zeros problem. We disagree. We believe that by 1) drilling down to the beat level and 2) expanding the temporal window from months to quarters that we will be able to effectively manage the occurrence of zeros in the data.

7.2 ANOTHER FORM OF ROBUSTNESS TESTING: CONSIDER DYNAMICS AT BOTH BEAT AND DISTRICT LEVELS

Neither we nor the City, however, can speak with certainty about the proposed beat-level approach until the data are descriptively analyzed. We are currently preparing the demographic data such that we can conduct analyses at *both* the district and beat level. Key differences in beat and district level analyses could be described in the report, if Parties wish. **Reporting results at both the beat and the district level will mean more effort in analytic time and report preparation. On the other hand, it has the potential to increase the robustness of findings if we find the patterns prove similar regardless of whether the beat or the district is used as the spatial grouping unit. Alternatively, it might show that patterns of impacts are *specific* to a one level of spatial aggregation. Spatial scaling concerns are an ongoing analytic and conceptual concern in community criminology (Taylor, 2015).** Note further that, in effect, Professor Fagan, in his *Floyd* reports, initially reported using the precinct, and then later reported using the census tract. So, the idea of testing models looking at police stop rates using multiple spatial scales is established.

We think it unwise to stick with just districts as the spatial unit because **the analysis using districts masks substantial heterogeneity in crime patterns, police activity, and demographics.** Simply put, Chicago police districts are large ecological units. Using a geographic information system, we were able to calculate the square mileage of each police district in the city of Chicago. The average district in the city is about **10 square miles**. Seven districts in the city are over 10 square miles. The smallest district (15th) is just under 4 square miles, while the largest is 31 square miles (16th district airport). The second largest district is 27 square miles, which is roughly half the square mileage of Boston. **We doubt that a reasonable person would assume that crime and policing patterns in an area that is half the size of Boston are uniform.** But this assumption can only be examined by looking at results at multiple spatial scales. In a host of areas within community criminology, leading scholars have moved to smaller spatial units such as streetblocks (Weisburd, Groff, & Yang, 2012). In the *Bailey* stop

and frisk ongoing monitoring, benchmark models have moved down to the police service area level (Rudovsky, Messing, & Lin, 2017), which is roughly parallel to beats in Chicago.

7.3 WE ARE NOT PRIORITIZING CONTEXT RACE OVER DETAINEE RACE

Additionally, the City and Chicago Police Department question whether “the proposal to transition from district-level to beat-level analysis is driven by a desire to investigate the impact of the race and ethnicity of the surrounding neighborhood (over and above the impact of the race and ethnicity of the detainee)” (p. 16).

It is not our intention to prioritize neighborhood-level effects over individual-level detainee effects. But, it is worth noting that all ecological analyses, whether at the beat or district level, remain at **an aggregate level** of analysis. In other words, such analyses attempt to predict the total count of Hispanic to White or Black to White stops within an administrative unit, within a particular time frame, *as opposed to* predicting a unique person’s likelihood of being stopped given certain circumstances.

Our main intention of beat-level analyses is to incorporate structural effects that are more specific to the actual location of police contact. This allows us to more effectively control for the structural context, whether those features of context are related to race or ethnicity, or something else like residential stability or residential socioeconomic status.

8 CONTINUING CONCERNS ABOUT VIOLENT ARREST DENOMINATOR MISPLACED

The City continues to express concerns about the use of violent arrest counts as the denominator used to transform counts into rates. During the 2nd reporting period we took care to address their most critical concerns about the violent arrest count denominator, specifically (see item 7.1). Although our preference is for the aforementioned denominator, it is worth noting that we conducted two sets of additional analyses to allay concerns. These included analyses using total arrests and age then gender-weighted population denominators.

Regardless of the denominator chosen, however, our findings show that investigatory stop counts of non-Hispanic Blacks significantly exceed those of non-Hispanic Whites in the average district, in the average month. **This significance pattern surfaced regardless of the denominator used.** In other words, we have demonstrated that the significant ethnracial disparities are **consistent across alternate denominators selected.** This would seem to be worth knowing.

In short, we argue that the Black-White stop count differential is **robust** across different kinds of measurements (denominators). Stated differently, regardless of how it is measured there is evidence of racial disparity in stop counts. *The robustness or consistency of this disparity across alternate denominators would not have been discovered had we not used the alternate denominators in the first place. The robustness itself seems to be important additional information.*

9 PETERSON & KRIVO VARIABLE (70 PERCENT OR MORE BLACK NON-HISPANIC) RATIONALE

We believe that subsequent ecological research needs to include a proxy measure of racial segregation, which has proven increasingly important in communities and crime research (Pratt & Cullen, 2005). Research indicates that communities of color are significantly more likely to concentrate poverty, instability, and disinvestment—structural conditions that generally predict higher rates of violent crime for several reasons. Poverty increases the likelihood that some individuals may resort to instrumental crimes for sustenance. Additionally, unemployment facilitates an unstructured lifestyle that may over-expose individuals to places conducive to violent conflicts. Moreover, residents of disenfranchised neighborhoods may lack the political clout necessary to organize for better police protection, in part, because the community is in a constant state of transition as residents constantly move in and out. Disinvestment contributes physical disorder as homes and facilities spiral into decay, which also lead to criminal opportunities. In sum, poor minority communities in a hyper segregated locale are at risk of becoming extremely socioeconomically disadvantaged, of producing activity patterns conducive to victimization, of developing a socially disorganized social climate, and of witnessing higher crime rates (Anderson, 2000; Wilson, 1996).

Existing research tells us that significant crime effects emerge when we compare and contrast predominantly Black, White, and Hispanic neighborhoods using a 70% or greater population threshold (Krivo, Peterson, & Kuhl, 2009; Peterson & Krivo, 2005, 2009, 2010; Peterson, Krivo, & Harris, 2000). For example, Peterson and Krivo (2010) used crime data from over 8,000 census tracts and 87 cities to empirically demonstrate how racial segregation concentrates structural conditions favorable for violence in minority communities, as opposed to predominantly white communities. Comparatively speaking, only 7% of White neighborhoods were in high or extreme poverty, compared to 78% of Black and Latino neighborhoods. In turn, violent crime rates in predominantly Black non-Hispanic neighborhoods were five times greater than in White non-Hispanic neighborhoods. Predominantly Latino neighborhood violent crime rates were over two times those of predominantly White non-Hispanic neighborhoods. Differences across racially distinct communities in residential instability, community investment, and disadvantage partially but markedly reduce these discrepancies. Without accounting for the above variables neighborhood crime violent crime rates are 4x greater in Black than White areas, and 2.5x greater in Latino than white areas. Accounting for those variables however, reduces Black-White and Latino-White disparities to 78% and 39%, respectively.

Therefore, considering the ways racial composition and in turn segregation concentrates structural conditions, it is necessary to capture the influence of segregated living environments. In line with Peterson and Krivo's research, we propose the use of a dummy variable, which would code neighborhoods that are at least 70% non-Hispanic Black as 1, and all others as zero. In sum, this indicator allows us to take community-level segregation into account, a crucial structural factor for understanding both racial impacts and community violence patterns. As noted below, we will be able to separate these threads analytically starting in Period 3 / Period 4 analyses.

10 THE 70 PERCENT BLACK NON HISPANIC BEAT RESIDENTIAL COMPOSITION VARIABLE: A CLARIFICATION GOING FORWARD FOR ECOLOGICAL, POST STOP OUTCOMES, AND LEGAL NARRATIVE REPORTS

10.1 THE MOTIVATION QUESTION

The City's experts infer that our inclusion of the racial composition variable at the beat level "presumably... is motivated by the desire to assess whether decisions made at CPD headquarters about how to allocate police resources across different districts might be the source of differences in policing outcomes, as distinct from any discriminatory behaviors by individual CPD officers."

This was not the reason. Rather the reasons were threefold.

1) We sought to align analyses more closely with those in the case of *Bailey et al. v. City of Philadelphia*. Individual level stop outcomes in those reports control for both racial and ethnic composition of the police service area (comparable to beats in Chicago), as well as ethnicity and race of detainee.

2) Structural ecology at the neighborhood level has confirmed for four decades three underlying fundamental and independent dimensions of neighborhood demographic structure: socioeconomic status, residential stability, and ethnic and racial composition (Golledge & Stimson, 1997; Hunter, 1971, 1974). Any analysis taking into account neighborhood demographic structure and **failing** to incorporate racial or ethnic composition would be incomplete and leave us open to the charge of conducting *mis-specified models*. This is a bad thing and to be avoided.

3) As explained above, racial segregation proves a key structural element for understanding community violence differences.

That said, the City experts' investigation removing violent crime from percent black population (their figure 6) leads us to the following proposal going forward:

In future analyses we separate this fundamental feature of beat ethnoracial composition (70% or higher Black non-Hispanic in residential composition) into two independent portions: the portion correlating perfectly with and predicted by violent crime; and the independent portion of ethnoracial composition. This will require an additional data file from CPD.

More specifically, we will need quarterly violent crime counts, at the beat level, starting with the last quarter of 2016. Because stop actions can shape crime levels, the violent crime levels will need to come from the quarter preceding the stop quarter investigated. For the district level analyses, we will need monthly violent crime counts starting in December, 2016.

Clarifying further, here are the additional data needed if we are to follow up on the City experts' suggestion:

- Quarterly violent crime counts, beat level, 4th quarter 2016 through 3rd quarter 2017
- Monthly violent crime counts, district level, December 2016 through November 2017.

We could accomplish this separation of community racial composition from community violent crime by regressing (logit model) the beat ethnoracial composition dummy variable on the temporally-lagged beat-level violent crime rates. (A parallel operation happens at the district level.) That regression generates two variables:

- The predicted probability that a beat will be 70 percent or more black and non-Hispanic based on its violent crime rate. **This variable correlates PERFECTLY with the violent crime rate from the preceding quarter.**
- The portion of the 70% or more black and non-Hispanic variable that is completely ***unrelated to the violent crime rate*** in the preceding quarter.

This addresses the City experts' concern in the following ways. They note that CPD is likely to deploy more officers to beats with higher violent crime rates. More deployment creates more resources for more stops. And if these higher violent crime rate areas are more predominantly Black and non-Hispanic, the demographic variable misleads us about what is essentially a deployment decision.

Our proposal here solves this problem. If we assume that CPD officers are deployed to the higher violent crime areas, then the portion of the Black non-Hispanic variable that correlates perfectly with crime (the predicted score) will capture that deployment dynamic. Consequently, if this portion of the racial residential composition variable proves important in predicting something like a pat down occurrence, a deployment/violent crime rationale is quite plausible.

As importantly, the same explanation **cannot** apply if the portion of the Black non-Hispanic residential composition variable that is **unrelated** to violent crime links to an outcome like pat downs. In this case, results would be saying there is something about residential racial composition when an area becomes predominantly Black and non-Hispanic, an aspect that is unrelated to violent crime, that links to an outcome like pat downs.

Of course, if the racial composition variable unrelated to violent crime continues to predict an outcome like pat downs, there are **still** ambiguities in interpretation. But at least one major plausible alternate explanation will have been removed.

In short, we are separating the beat ethnoracial composition variable to address the city experts' concerns while at the same time continuing to control in a comprehensive way for the three fundamental demographic aspects of neighborhood structure so that we do not have a mis-identified statistical model.

Please note that ***the predicted portion of ethnoracial composition at the beat or district level will correlate perfectly with the temporally lagged violent crime rate and thus is an absolutely perfect proxy for the previously occurring violent crime rate.*** Note further that the unpredicted portion of ethnoracial residential composition **cannot** be explained by any dynamic driven by the previously occurring violent crime rate.

11 WHY IS PAT DOWN + NEA RELEVANT, AND HOW IT DOES NOT IGNORE ARRESTS

Two questions arose regarding this outcome: experiencing a patdown combined with no enforcement action. The first asks how is this relevant as an outcome. The second raises a concern reviewers voiced about arrest being ignored.

11.1 PAT DOWN ONLY AS AN OUTCOME OF CONCERN FROM A PROCEDURAL JUSTICE PERSPECTIVE

We reproduce here text from the first post stop outcome report (Taylor & Johnson, 2017) that appeared on pages 26-27 of that report:

First, situated accounts of police civilian interactions highlight that pat down and release does occur and that it does bother civilians (McArdle & Erzin, 2001; Simon & Burns, 1997). Such interactions contribute to tension between inner city Black residents and police (Brunson, 2006, 2007a; Brunson & Gau, 2011; Gau & Brunson, 2010). To be patted down and released may strike many residents of color as simply being hassled by police (McArdle & Erzin, 2001).

Further, this outcome seems particularly relevant given a procedural justice perspective (Sunshine & Tyler, 2003; Tyler, 1988, 1997, 2001, 2003; Tyler & Huo, 2002; Tyler & Lind, 2001). The outcome reflects a component of the construct "degree of police intrusion during... stops" (Tyler, Fagan, & Geller, 2014: abstract), an outcome recently introduced by procedural justice scholars.

Tyler, Fagan, and Geller (2014: 763) used telephone survey data of young men living in New York City to learn about impacts of their contacts with police on both their views of police legitimacy and their willingness to cooperate with police and courts. In describing "general neighborhood experiences with police" participants reported on the "degree of intrusion during those stops" happening near where they lived. Several survey items contributed to a broader index reflecting intrusion. One of the items in this index was "did the police... 'Frisk or pat you down'" (Tyler et al., 2014: 784).

Would most agree that a stop ending with a pat down and release is more intrusive than a stop and no pat down and release? This certainly seems to be the implication of the work by Tyler, Fagan, and Geller (2014). Those authors observed significant impacts of police intrusiveness on respondents' willingness to cooperate with police (Table 6). This aligns with much of the ethnographic work on urban Black residents and police; this work agrees that unwarranted frisks are intrusive and affects residents' views of police (Brunson, 2005, 2006, 2007b).

That said, no inferences are drawn about the fraction of frisk-and-release stops where police had grounds for a much more intrusive stop such as for example a frisk-and-cite or frisk-and-search stop, or a frisk-search-and-arrest stop. Nor are any inferences made about the fraction of no-frisk-and-release stops where police similarly might have had grounds for more intrusive actions.

This text did not appear in the Period 2 report on post stop outcomes.

In short, the current scholarship explicitly links pat downs with intrusive policing, and reports of intrusive policing with decreased willingness to cooperate.³ In our view, a strong case can be made that in those instances where the stop was followed with a pat down, and no other features of the stop for the detainee would have justified more serious enforcement action by the officer, such encounters have the potential to adversely affect citizens' views about and their willingness to cooperate with the police.

How to move forward?

- Would the City's experts like us to include the above text in every report that includes this outcome?
- Or, are they saying that they do not accept the argument about the relevance of the outcome to broader views about police?
- Or, are they saying that the acknowledgment about the residual ambiguity of the indicator is insufficient?

Please advise.

11.2 ARRESTS ARE NOT IGNORED, RATHER THEY APPEAR IN A DIFFERENT CONTRAST AND ARE CONTROLLED

Arrests were not ignored or excluded during the analysis of stops with pat downs but no enforcement action. See footnote 31 on p. 54. This is also explained in the footnotes to table 24 on p.55.

A brief clarification follows. The four groups analyzed include the following:

- (1) no enforcement+no pat down vs.
- (2) no enforcement+pat down vs.
- (3) enforcement+no pat down vs.
- (4) enforcement+pat down.

These four groups also appear in table 4 on p. 20 in the post stop outcomes report.

³ Fagan (2017: 696) states this more succinctly: "A robust body of evidence shows that aggressive and unproductive stops produce a litany of petty and not-so-petty indignities that translate into cynicism toward police."

The way the multinomial analysis works is that it examines the following three contrasts *simultaneously*:

(1) vs. (2)

(1) vs. (3)

(1) vs. (4)

While the analysis is considering the differences between group (1) and group (2), *it is simultaneously taking into account the other contrasts. In other words, it is controlling for what was happening with arrests. Those are located in separate contrasts. All of these contrasts are estimated simultaneously (Long & Freese, 2006: 224).*

In short, arrests **were** controlled and are **not** problematic.

12 MORE TRANSPARENCY ON WHAT SPATIAL SMOOTHING IS DOING

The City's experts argue that use of the smoothing technique "has the drawback of mixing together data that can come from very heterogeneous types of police districts and neighborhoods" (p. 5). While each district's surrounding districts are used to perform the smoothing technique, adjustments are based on how far a particular case is from the group mean. In other words, cases that are farthest from the mean would receive greater adjustment than those that are closer. This is because Empirical Bayes adjustments are made.

That aside, the experts' points are well taken. Considering the sheer size of the ecological units used – districts -- it is possible that our smoothing approach is limited by substantial differences in ethnoracial arrest counts even among neighboring districts.

Such a critique, however, lends itself well to a beat-level analysis solution as long as we expand the temporal frame from months to quarters. Smaller spatial units of analyses (beats within districts) are likely to demonstrate less heterogeneity in arrest counts and, if spatially smoothed, yield more sensible arrest rates.

13 ACLU— WE DO NOT SAY WHAT QUESTIONS WE ARE ADDRESSING

The ACLU response on P.17, starting under subhead D says as follows: "the ACLU continues to have concerns about the statistical reports. Chief among them is that none of the reports identify what factual questions they are answering, how that factual question relates to disparate treatment, disparate impact, and fourth amendment questions raised on the agreement."

Please see section 6.2 titled "questions addressed" on p. 9 in the post stop outcomes report. Specific questions to be addressed are listed there.

Please also see section 4, titled "purpose", p.6-7, in the legal narratives report. This section lists four specific questions addressed in this report.

Finally, in the ecological report, the key question appears in its own paragraph on p.6:

“At the district level, are arrests earlier producing more stops later for non-Hispanic Blacks as compared to Non-Hispanic Whites, and for Hispanic Whites as compared to non-Hispanic Whites?”

In short, we think that each statistical report clearly states what factual questions are being addressed.

14 ALIGNMENT WITH FLOYD

On p.19 of the ACLU response, those authors raise concerns about how current models do not align with the models used in *Floyd*.

We would like to draw parties’ attention to section 8.3.3 on p. 15 of the draft post stop outcomes report. That section specifies changes made to the Period 2 post stop outcomes models. The changes align those models more closely with some of the *Floyd* models. More specifically, this was one reason for moving to beats instead of districts so that in Chicago, like in New York, there would be "several dozen spatial units" within which to organize the data, and to which we can attach demographic predictors as was done in the *Floyd* models. If we stick with Chicago police districts, numbering only a couple dozen, concerns arise about interpreting impacts of district level demographic predictors, including racial composition, with this small number of special units. Moving to beats helps solve this problem.

ACLU referenced ⁴ the revised Fagan models in the Floyd report. The models reported there are all ecological in nature, so the comparison is with the ecological models we report. Here are some specific comparisons and contrasts.

Fagan’s Table 5

- This model does not separate either complaints or stops by ethnoracial group. We differentiate both of these ethnoracially.
- By controlling for boroughs as fixed effects he is in effect making his entire model about *intra*-borough comparisons. To do something similar in Chicago would mean dividing the city up into zones and controlling for these zonal differences. We do not recommend this because Chicago is smaller than New York.
- This model includes a spatially lagged measure of complaints (crime). We have proposed something analogous for P3/P4 by introducing spatial lag variables for the stop counts themselves.
- Fagan uses logged population as an exposure variable. We do one series of ecological models with this outcome.
- Fagan’s census tract demographics fail to control for residential stability. This leaves his model open to a charge of mis-specification. We do control for stability.
- Like Fagan, we also control for context SES and context racial and ethnic composition. Although the specific variables differ, we would argue that our operationalization of ethnic and racial composition aligns more closely with current scholarship on understanding the impacts of racialized structure.

⁴ <https://ccrjustice.org/sites/default/files/assets/files/FaganSecondSupplementalReport.pdf>

- Fagan has patrol strength indicators. We have not requested those.
- **One big difference** is that Fagan has stops organized by relevant suspected crime category. **We have not been classifying stops** so can only look at total stops, the first column in his table.
- **Another big difference: he includes census tract and nearby census tract crime rates.** We do not include crime rates. *Whether and if so how to include these is an important question to seek clarity on.* We believe that our proposal above for separating racial composition and violent crime **may address this concern.**

Fagan's table 7

- Fagan uses count models. So do we.
- Fagan uses ethnoracial specific stop counts (under his case characteristics). So do we.
- Fagan includes ethnoracial “other.” We do not. We drop these small number of stops. Our focus is exclusively on the three central ethnoracial groups in Chicago.
- Fagan controls for demographics of surround (census tract in his case, police beats we are proposing in our case) in terms of SES and racial and ethnic composition. We do too. (But, as pointed out above, we also control for residential stability.)
- Again, Fagan has crime specific stop types and we do not.
- Fagan includes tract level crime (total complaints) from the previous quarter. **This is a big difference.** We do not include crime rates. *Whether and if so how to include these is an important question to seek clarity on, especially because of potential multicollinearity concerns when used in conjunction with our violent arrest exposure variable.* Again, we believe that our proposal above for separating racial composition and violent crime **may address this concern.**
- *A second important difference: his exposure variable (population logged) because it is listed as a tract characteristic is not ethnoracial specific. In other words, his model has an ethnoracial specific numerator BUT NOT an ethnoracial specific denominator. Therefore, these models, IF our surmise is correct, cannot be informing us about ethnoracial specific stop rate differentials.*⁵ We have worked extremely hard in our current ecological models to construct **both** numerators (the count outcome variable) **and** denominators (the exposure variables) that are ethnoracial specific. Of course, it is because we have sought this alignment that we have created the “zeroes” problem for ourselves that gets referenced elsewhere.

We are happy to engage in further discussion with parties about these points of correspondence and non-correspondence. *The most critical divergence has to do with context crime rates, and we think the City's experts' suggestion about residualizing is extremely helpful here and addresses this matter.* But, of course, let's discuss.

⁵ Fagan, personal communication, February 22, 2018 seems to confirm our surmise.

15 ACLU – P 21 – BOTTOM – INTERNAL REPLICATION – SAMPLE SIZE

On p. 21 in the ACLU response, they note

" Dr. Taylor's analyses ... divided the sample in two and ran his test on each half. This practice exacerbates the problem of small samples, making it far less likely to find statistical significance even when evidence of disparity is present. We also not aware of a methodological justification for this practice. Additionally, when one half the data shows the statistical significance and the other does not, that may be result of the halved data set... We suggest that he end this practice, or provide citations."

There are two parts of our response: the first is whether the practice adversely affects statistical power. It does not. The second part documents this common practice, called split-sample cross-validation. A few citations are provided. Many more could be provided were more time available.

15.1 IS STATISTICAL POWER COMPROMISED?

Certainly, ACLU's consultant is *generically* correct that cutting a sample in half has the potential to decrease statistical power below an acceptable level. As is well known by all statisticians, *a priori* statistical power is a function of

- sample size,
- the alpha level (what level of statistical significance is required for a difference to be significant?), and
- the “actual” size of the difference being investigated (Cohen, 1988; Faul, Erdfelder, Buchner, & Lang, 2009). Is the effect in question small, medium, or large?

So at the broad theoretical level, ACLU’s consultant is completely correct.

There is also the question of how much statistical power is needed for an analysis to be acceptable? Although norms may differ across disciplines, in many social sciences a level of statistical power of 80 percent or higher is considered acceptable (Cohen, 1992). This means the researcher’s chances of actually discovering a difference that, through some powerful all-seeing eye we know actually exists, are about 4 out of 5.

The discussion of statistical power in this context, however, must recognize that mixed models are employed. There are power issues at two different levels (Browne & Golalizadeh, 2008; Raudenbush & Liu, 2000). Both the number of groups (J) – here the number of districts or beats – and the average number of cases with in each group (n), affect statistical power. In the data setup such as we have here the more significant constraint on statistical power is the number of groups. This is because the F test of the average treatment effect relies on (J - 1) for the degrees of freedom in its denominator (Raudenbush & Liu, 2000: 202, Eq. 14). And J is a number driven by the number of groups. Further, in situations where variations in treatment effect might be in question "increasing J, the number of sites, is more crucial than increasing n, the number of participants per site" (Raudenbush & Liu, 2000: 202).

It is in part because of statistical power concerns at level 2, that we conducted analyses based on ISRs using beats or beats within districts as the analytical frame. The potential power concern also can be addressed in ecological models by shifting to beats rather than districts.

Statistical power at level 2 is not affected by internal sample splitting because the number of beats is the same in both samples, around 270. In short, one of the most crucial determinants of statistical power in multilevel models is not affected by sample splitting at all.

Further, note two important additional factors. First, in the post stop outcomes analysis, and in most outcomes in the legal narrative report, extremely large numbers of records were available **even after sample splitting**. For example, in tables 18, 23, and 24 in the post stop outcomes report, available cases number roughly 25,000. (We attempted in each table to include information about the number of cases in the analyses shown.) Even for propensity score matching models, we had more than 3000 cases in each split sample; see for example tables 19 and 21.

Given these numbers of cases in the thousands or tens of thousands we believe we can make the case that statistical power was not substantively affected by the sample splitting procedure. If the consultants would like to see a priori power estimations based on level I, case level counts, those can be provided in future reports. Please advise. Given time constraints, we have not constructed those power curves here.

In addition, when we do run out of cases it is not because of the split sample design, as happened in some of the legal narratives analysis. Rather, that problem arose because of our sampling strategy for selecting cases submitted to narratives analysis. We are in the process of modifying that sampling strategy going forward so that we do not "run out of cases."

In short, statistical power was not compromised in any substantive way by the split sample design. If folks wish, a priori statistical power calculation can be supplied in future. With the mixed effects models SPECIFICALLY the power calculations will require noticeable ADDITIONAL effort on our part, effort that is not currently budgeted.⁶

15.2 SPLIT SAMPLE CROSS-VALIDATION

Even more importantly, the benefits gained by our split sample approach are well-established in the literature, and highly recommended. This technique is called split sample cross validation and is routinely used in a wide variety of natural science and social science investigations (Janeksela & Miller, 1985; Tchakoute Tchuigoua, 2014; Wolins, 1967). For example, Tchakoute

⁶ The single level power calculations are relatively straightforward. The multilevel ones, not so. Specialized programs must be used, like MLWiN. Further, the power estimation itself requires gathering numerous parameters about the data themselves, and inputting those parameters into the run files. This can not be done by graduate students, and would require substantial *additional* amounts of Dr. Taylor's time.

Tchuigoua (2014: 2194) “randomly split the initial sample into two sub-groups” to “assess the external validity of results.”

Split sample cross validation is widely recommended for advanced multivariate modeling where model modification is done on one half of the data, and the modified model is tested on the second random half of the data (Bollen, 1989; Hayduk, 1987). But it is also recommended in the context of regression models if substantial numbers of cases are available (Wolins, 1967).

In short, split sample cross validation is a widely accepted technique for a range of models. The results can provide, as they do here, a robustness check.

16 APPENDIX A

Table 1 Descriptive statistics, binary outcome variables

Variable	Variable name	N	Min.	Period 2			Period 1		
				Max.	Mean	SD	Sum	Mean	Sum
Pat down conducted	dpat	50,715	0	1	0.295	0.456	14,945	0.339	18,364
Pat down → weapon (*)	pathit_w2	14,945	0	1	0.035	0.183	517	0.025	465
Pat down → drugs (*)	pathit_d2	14,945	0	1	0.019	0.136	280		
Pat down → any (*) (a)	pathit_2	14,945	0	1	0.063	0.243	940		
Search conducted	dsearch	50,715	0	1	0.138	0.345	7,002	0.177	9,595
Search → weapon (*)	se_hit_w2	7,002	0	1	0.050	0.217	348		
Search → drugs (*)	se_hit_d2	7,002	0	1	0.146	0.353	1,022		
Search → any (*)	se_hit_2	7,002	0	1	0.188	0.390	1,313		
Any enforcement action taken (b)	denforce_2	50,642	0	1	0.278	0.448	14,066	0.322	17,425

Note. For all binary outcomes, 1 = outcome occurred, 0 = did not occur

Note. “Drugs” means officer found either drugs or contraband or both.

Note. “any” means officer found either a weapon or (drugs or contraband).

Source: Period 2: July-December 2016 ISRs, CPD. Period 1: January-June 2016 ISR reports, CPD.

(*) = dependent variable depends on selection through another dependent variable. More specifically, the pat down “hit” variables were set to missing if no pat down took place; and, the search “hit” variables were set to missing if no search took place.

(a) The numbers for this variable were taken from the ISR form checkbox “was a weapon or contraband discovered as a result of the pat down?” Items included, in addition to weapons, firearms, and specific drugs such as cannabis, heroin, cocaine, other items: drug paraphernalia, other, other controlled substance, alcohol, and stolen property. So, the total for this variable will exceed the total of firearms plus weapons plus drugs found.

(b) This indicator is based on the ISR check box “any enforcement action taken.” This check box, however, was not checked 73 times when a box for a specific enforcement action was checked. More specifically, 70 times when “other” was checked as the enforcement type code, one time when arrest was checked, one time when violation was checked, and one time when personal citation was checked the overall enforcement check box indicated no rather than yes for any enforcement action taken. Those 73 cases are *not* included here because priority is given to the overall check box, and those 73 cases are set to missing.

APPENDIX B

Results from selection portion of model predicting pat down hit rates

Note: Significant non-ethnoracial predictors highlighted in yellow

dpat							
p_k_blanonh		.0482295	.0693964	0.69	0.487	-.087785	.1842439
p_k_hisp		.170922	.0816676	2.09	0.036	.0108564	.3309876
p_other		.0418304	.0253403	1.65	0.099	-.0078357	.0914964
p_asian		.0009772	.0032175	0.30	0.761	-.0053291	.0072834
sesindx2		-.1616728	.0600412	-2.69	0.007	-.2793515	-.0439942
stabindx		.1675399	.0372568	4.50	0.000	.0945179	.2405619
dblack		.1990808	.051692	3.85	0.000	.0977664	.3003953
dhisp		.1384236	.0542691	2.55	0.011	.0320581	.2447892
dmale		.8345581	.0388609	21.48	0.000	.7583921	.9107242
age1825		-.0318359	.034335	-0.93	0.354	-.0991313	.0354595
age2635		-.2366233	.0402701	-5.88	0.000	-.3155512	-.1576953
age3645		-.4277853	.048242	-8.87	0.000	-.5223379	-.3332328
age46p1		-.8574048	.049229	-17.42	0.000	-.9538919	-.7609177
daug		-.0513889	.0391419	-1.31	0.189	-.1281057	.0253279
dsep		.0501391	.0413914	1.21	0.226	-.0309864	.1312647
doct		.1230889	.0403596	3.05	0.002	.0439855	.2021922
dnov		.130569	.0471859	2.77	0.006	.0380864	.2230515
ddec		.0558841	.0493138	1.13	0.257	-.040769	.1525373
wknddum		.1015537	.0248402	4.09	0.000	.0528678	.1502396
dhr0306		.10716	.0720832	1.49	0.137	-.0341205	.2484405
dhr0609		-.3693059	.117967	-3.13	0.002	-.600517	-.1380948
dhr0912		-.3325161	.0543375	-6.12	0.000	-.4390156	-.2260166
dhr1215		-.3391108	.04756	-7.13	0.000	-.4323267	-.245895
dhr1518		-.1540019	.0529349	-2.91	0.004	-.2577524	-.0502514
dhr1821		-.2734616	.0452246	-6.05	0.000	-.3621003	-.184823
dhr2123		-.1584182	.0410711	-3.86	0.000	-.2389161	-.0779204
dvehstop		.4083845	.029968	13.63	0.000	.3496482	.4671208
eventmis		.350563	.0450235	7.79	0.000	.2623186	.4388075
_cons		-1.345741	.0982807	-13.69	0.000	-1.538367	-1.153114

17 REFERENCES

- Aakvik, A. (2001). Bounding a Matching Estimator: The Case of a Norwegian Training Program. *Oxford Bulletin of Economics and Statistics*, 63(1), 115-143. doi:10.1111/1468-0084.00211
- Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *American Journal of Public Health*, 86(5), 726-728.
- Anderson, E. (2000). *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. New York: W.W. Norton.
- Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley Interscience.
- Browne, W. J., & Golalizadeh, M. (Producer). (2008). Sample size calculations for multilevel models (Part II). PowerPoint presentation. ESRC Research Methods 2008 [ONLINE: <http://www.bristol.ac.uk/cmm/software/mlpowsim/> ; accessed 19 February, 2014].
- Brunson, R. K. (2005). Young Black Men and Urban Policing in the United States. *British Journal of Criminology*, 46(4), 613-640. doi:10.1093/bjc/azi093
- Brunson, R. K. (2006). Gender, Race, and Urban Policing: The Experience of African American Youths. *Gender & Society*, 20(4), 531-552. doi:10.1177/0891243206287727
- Brunson, R. K. (2007a). "Police don't like black people:" African-American young men's accumulated police experiences. *Criminology & Public Policy*, 6(1), 71-101.
- Brunson, R. K. (2007b). "Police don't like black people": African-american young men's accumulated police experiences*. *Criminology & Public Policy*, 6(1), 71-101.
- Brunson, R. K., & Gau, J. M. (2011). Officer Race Versus Macro-Level Context: A Test of Competing Hypotheses About Black Citizens' Experiences With and Perceptions of Black Police Officers. *Crime & Delinquency*. doi:10.1177/0011128711398027
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Fagan, J. (2017). Recent evidence and controversies in "the new policing". *Journal of Policy Analysis and Management*, 36(3), 690-700. doi:10.1002/pam.21995
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*, 41(4), 1149-1160. doi:10.3758/BRM.41.4.1149
- Gau, J. M., & Brunson, R. K. (2010). Procedural Justice and Order Maintenance Policing: A Study of Inner-City Young Men's Perceptions of Police Legitimacy. *Justice Quarterly*, 27(2), 255-279. doi:10.1080/07418820902763889
- Golledge, R. G., & Stimson, R. J. (1997). *Spatial Behavior: A Geographic Perspective*. New York: Guilford Press.
- Hayduk, L. A. (1987). *Structural equation modeling with LISREL*. Baltimore: Johns Hopkins University Press.
- Hunter, A. (1971). The ecology of Chicago: Persistence and change, 1930 - 1960. *American Journal of Sociology*, 77, 425-443.
- Hunter, A. (1974). *Symbolic communities*. Chicago: University of Chicago Press.
- Janeksela, G. M., & Miller, M. G. (1985). An Exploratory study of delinquency, criminal offenses, and juvenile status offenses via the cross-validation design. *Adolescence*, 20(77), 161-170.

- Krivo, L. J., Peterson, R. D., & Kuhl, D. C. (2009). Segregation, Racial Structure, and Neighborhood Violent Crime. *American Journal of Sociology*, 114(6), 1765-1802. doi:10.1086/597285
- Long, J. S., & Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata* (Second ed.). College Station, Texas: Stata Press.
- McArdle, A., & Erzin, T. (Eds.). (2001). *Zero Tolerance: Quality of Life and the New Police Brutality in New York City*. New York: New York University Press.
- Peterson, R. D., & Krivo, L. J. (2005). Macrostructural analyses of race, ethnicity, and violent crime: Recent lessons and new directions for research. *Annual Review of Sociology*, 31, 331-356.
- Peterson, R. D., & Krivo, L. J. (2009). Segregated Spatial Locations, Race-Ethnic Composition, and Neighborhood Violent Crime. *The ANNALS of the American Academy of Political and Social Science*, 623(1), 93-107. doi:10.1177/0002716208330490
- Peterson, R. D., & Krivo, L. J. (2010). *Divergent Social Worlds: Neighborhood Crime and the Racial-Spatial Divide*. New York: Russell Sage
- Peterson, R. D., Krivo, L. J., & Harris, M. A. (2000). Disadvantage and neighborhood violent crime: Do local institutions matter? *Journal of Research in Crime and Delinquency*, 37(1), 31-63.
- Pratt, T. C., & Cullen, F. T. (2005). Assessing macro-level predictors and theories of crime: A meta-analysis *Crime and Justice: A Review of Research* (Vol. 32, pp. 373-450).
- Raudenbush, S., & Liu, X. (2000). Statistical Power and Optimal Design for Multisite Randomized Trials. *Psychological Methods*, 5, 199-213.
- Rudovsky, D., Messing, P., & Lin, S. (2017). *Plaintiffs' seventh report to court and monitor on stop and frisk practices: Fourteenth Amendment racial disparity issues. Mahari Bailey et al. v. City of Philadelphia, et al., in the United States District Court for the Eastern District of Pennsylvania, C.A. No.10-5952*. Retrieved from
- Simon, D., & Burns, E. (1997). *The Corner: A Year in the life of an inner-city neighborhood*. New York: Broadway Books.
- Sunshine, J., & Tyler, T. (2003). The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing. *Law & Society Review*, 37(3), 513-548.
- Taylor, R. B. (1994). *Research Methods in Criminal Justice*. New York: McGraw Hill.
- Taylor, R. B. (2015). *Community Criminology: Fundamentals of Spatial and Temporal Scaling, Ecological Indicators, and Selectivity Bias*. New York: New York University Press.
- Taylor, R. B., & Johnson, L. (2017). *RE: ACLU Matter vs. - REF# 1340012232: Analysis of Chicago Police Department Post-stop Outcomes during Investigatory Stops January through June 2016: Input to Hon. Arlander Keys' (Ret.) First Year Report. Unpublished Report. Available online at: <https://www.aclu-il.org/en/publications/march-2017-stop-frisk-report> ; listed as Appendix B: Analysis of CPD Post-stop Outcomes during Investigatory Stops. Also available online at: <http://www.rbtaylor.net/pubs.htm>*
- Tchakoute Tchuigoua, H. (2014). Institutional framework and capital structure of microfinance institutions. *Journal of Business Research*, 67(10), 2185-2197. doi:<https://doi.org/10.1016/j.jbusres.2014.01.008>
- Tyler, T. (1988). What is Procedural Justice? Criteria Used By Citizens to Assess the Fairness of Legal Procedures. *Law and Society Review*, 22(1), 103-135.
- Tyler, T. (1997). Citizen discontent with legal procedures. *American Journal of Comparative Law*, 45, 869-902.

- Tyler, T. (2001). Public Trust and Confidence in Legal Authorities: What Do Majority and Minority Group Members Want from the Law and Legal Institutions? *Behavioral Science and the Law*, 19, 215-235.
- Tyler, T. (2003). Procedural justice, legitimacy, and the effective rule of law *Crime and Justice: A Review of Research, Vol 30* (Vol. 30, pp. 283-357).
- Tyler, T., Fagan, J., & Geller, A. (2014). Street stops and police legitimacy: Teachable moments in young urban men's legal socialization. *Journal of Empirical Legal Studies*, 11(4), 751-785. doi:10.1111/jels.12055
- Tyler, T., & Huo, Y. J. (2002). *Trust in the Law: Encouraging Public Cooperation With the Police and Courts*: Russell Sage Foundation.
- Tyler, T., & Lind, E. A. (2001). Procedural Justice. *Handbook of Justice Research in Law*, 65-92.
- Weisburd, D., Groff, E. R., & Yang, S.-M. (2012). *The Criminology of Place: Street Segments and Our Understanding of the Crime Problem*. Oxford, UK: Oxford University Press.
- Wilson, W. J. (1996). *When Work Disappears: The World of the New Urban Poor*. New York: Knopf.
- Wolins, L. (1967). The Use of multiple regression procedures when the predictor variables are psychological tests. *Educational and Psychological Measurement*, 27, 821-827.