

Claremont Colleges

Scholarship @ Claremont

CGU Theses & Dissertations

CGU Student Scholarship

Spring 2021

Essays on Race in Policing

Hunter Robert Johnson

Claremont Graduate University

Follow this and additional works at: https://scholarship.claremont.edu/cgu_etd

Recommended Citation

Johnson, Hunter Robert. (2021). *Essays on Race in Policing*. CGU Theses & Dissertations, 214.
https://scholarship.claremont.edu/cgu_etd/214. doi: 10.5642/cguetd/214

This Open Access Dissertation is brought to you for free and open access by the CGU Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in CGU Theses & Dissertations by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.

Essays on Race in Policing

Hunter Johnson

Claremont Graduate University
2021

© Copyright Hunter Johnson, 2021.
All rights reserved

Approval of the Dissertation Committee

This dissertation has been duly read, reviewed, and critiqued by the Committee listed below, which hereby approves the manuscript of Hunter Johnson as fulfilling the scope and quality requirements for meriting the degree of Doctor of Philosophy in Economics.

Gregory DeAngelo, Chair
Claremont Graduate University
Associate Professor of Economic Sciences

Matthew B. Ross
Claremont Graduate University
Assistant Professor of Economic Sciences

CarlyWill Sloan
Claremont Graduate University
Assistant Professor of Economic Sciences

Abstract

Essays on Race in Policing

Hunter Johnson

Claremont Graduate University: 2021

In recent years, a number of high-profile policing controversies have led to global indignation over racial disparities in policing and perceived police brutality. This paper explores three different dimensions of race in policing. The first chapter of this dissertation examines whether the presence of female and minority police officers affects the likelihood of police use of force and whether officers are more or less likely to use force against civilians of a different race. Focusing on a subset of 911 calls resulting in arrest, I use an instrumental variables estimation method with dispatch availability by officer race/gender as an instrument for the presence of different officer types. I find that the presence of a female officer significantly reduces the likelihood that force is used. Calls involving white officers and black civilians are significantly more likely to result in use of force.

The second chapter uses data on 7.5 million police-civilian interactions made by 1,663 Texas Highway Patrol officers to estimate the impact of five mandatory police trainings on the racial composition of traffic stops and racial disparities in related outcomes. The five trainings considered are (1) Cultural Diversity; (2) Arrest, Search & Seizure; (3) Racial Profiling; (4) Traffic; and (5) Deescalation. We exploit quasi-random variation in the timing of when individual officers receive training and estimate a series of event study models. We find that training has little to no effect on policing behavior in terms of either racial composition or stops or related outcomes. In general, our findings cast serious doubt on the ability of policymakers to use training as an effective intervention for combatting longstanding disparities in law enforcement.

The third chapter examines whether externally-imposed affirmative action plans designed to increase the shares of nonwhite and female police officers have impacted the rates of reported offenses and/or offenses cleared by arrest. Using a series of modern econometric strategies including difference-in-differences decomposition and generalized synthetic controls, we do not find a significant effect of court-imposed affirmative action plans on the rates of reported offenses or reported offenses cleared by arrest. We also consider whether unlitigated agencies change their practices due to the threat of litigation but are unable to identify causal evidence of such threat effects.

Acknowledgments

I have received a great deal of support in writing this dissertation. My deepest gratitude goes to my advisor Greg DeAngelo without whom I would never have gone down this path of researching crime and policing. Before Greg arrived at CGU at the start of my third year I was pursuing international money and finance and had never considered applied microeconomics. He helped me pursue a new path and showed me that I was capable of producing publishable research. Greg's guidance has been invaluable in helping me grow as a researcher.

I am deeply indebted to my other committee members, Matt Ross and CarlyWill Sloan, for their helpful comments on my research and the advice they have given me since they arrived at CGU. I would like to acknowledge Matt in particular for being a great coauthor on one of the chapters of this dissertation.

I am grateful to my other coauthors Steve Mello, Stephen L. Ross, and Maryah Garner, whose contributions to two chapters of this dissertation have been invaluable. In addition to being a coauthor, Maryah has been a great friend since we both started the Ph.D. program at CGU. I would like to thank Mark Hoekstra for his valuable comments on my job market paper. His understanding of police use of force and good judgment have been a great help in carrying out this research. I would also like to thank Hisam Sabouni for helping me get started with programming and mentoring me throughout the program.

I would like to extend my gratitude to the Institute for Humane Studies and the Horowitz Foundation for their generous financial support. The Computational Justice Lab at CGU has also been a tremendous help not only financially but also as a place to discuss important new research in the world of criminal justice. In particular, I would like to acknowledge Yunie Le, Rebecca Bommarito, Josie Xiao, Morgan Stockham, Minjae Yun, Dana Avgil, Rainita Narender, and Anuar Assamidanov.

Finally, I want to thank my parents and close friends for all the support they have given me over the years. Without their help, I never could have made it this far.

Contents

1	Chapter 1: Does the Presence of Female and Minority Police Reduce the Use of Force?	1
1.1	Introduction	1
1.2	Data Sources	5
1.3	Identification Strategy & Results	6
1.3.1	Dispatch and Arrest Protocol	6
1.3.2	Models	9
1.3.3	Main Results	12
1.4	Robustness	13
1.4.1	Replication of Hoekstra & Sloan (2020)	13
1.4.2	Officer Timing	16
1.5	Heterogeneous Effects	18
1.6	Conclusion	20
2	Chapter 2: Is Police Training an Effective Intervention for Addressing Disparities?	22
2.1	Introduction	22
2.2	Data & Background	23
2.2.1	Texas Highway Patrol Data	23
2.2.2	Officer Training	24
2.3	Methods & Results	27
2.3.1	Does Cultural Diversity Training Affect Racial Disparities?	27
2.3.2	Do Other Trainings Affect Racial Disparities?	30
2.3.3	Does Training Affect Outcomes for White Motorists?	32
2.4	Parametric Models	33
2.5	Discussion and Conclusion	35
3	Chapter 3: Estimating Effects of Affirmative Action in Policing	37
3.1	Introduction	37
3.2	Background	38
3.3	Data	42
3.3.1	Affirmative Action Litigation Data	42
3.3.2	UCR and CPS Data	43

3.4	Models and Results	45
3.4.1	Two-way Fixed Effect Difference-in-Differences Models	45
3.4.2	Difference-in-Differences Decomposition	47
3.4.3	Results for DD Models	47
3.4.4	Duration Models	48
3.4.5	Results for Duration Models	49
3.4.6	Generalized Synthetic Control	50
3.4.7	Results for Generalized Synthetic Control Models	51
3.5	SUTVA Violation Test	51
3.6	Discussion	55
3.7	Conclusion	56

1 Chapter 1: Does the Presence of Female and Minority Police Reduce the Use of Force?

1.1 Introduction

On May 25, 2020, George Floyd died while in the custody of four officers of the Minneapolis Police Department. Floyd, an unarmed black male, died while a white police officer kneeled on his neck for eight minutes and 46 seconds. The autopsy performed by the Hennepin County Medical Examiner’s Office concluded that the cause of death was “cardiopulmonary arrest complicating law enforcement subdual, restraint, and neck compression.” Video of Floyd’s death recorded by bystanders led to widespread condemnation of the officers involved and worldwide protests in support of the Black Lives Matter movement. These protests have led to calls for wide-ranging police reform and heightened scrutiny on police use-of-force practices. Floyd’s death as well as the deaths of other unarmed black men such as Michael Brown have also renewed attention on racial prejudice in law enforcement.

There is a widespread sentiment among the American public that law enforcement is characterized by racial prejudice, with blacks reporting much lower confidence in police than whites (Pew Research Center, 2016; Gallup, 2016). Around 67 percent of blacks report feeling that blacks are treated less fairly than whites by police, while only 40 percent of whites feel that blacks are treated less fairly (Gallup, 2016).

Despite widespread media coverage and public interest, however, there is a dearth of empirical literature on police use of force and potential racial disparities. A paper by Fryer (2018) uses data from four different samples across the United States to study racial bias in use of force. Fryer considers the following samples: New York City’s stop-question-and-frisk (SQF) data containing detailed use of force descriptions; Police-Public Contact Survey (PPCS) data containing civilian-reported interactions with police including use of force; officer-involved shooting (OIS) data from three Texas cities, six Florida counties, and Los Angeles County; and Houston arrest data where lethal force is deemed “more likely to be justified,” coupled with Houston OIS data. Analyzing each sample separately, Fryer finds that blacks and Hispanics are more likely to experience non-lethal uses of force, but he finds no evidence of racial bias in the use of lethal force when contextual factors are accounted for. Fryer argues that his findings are consistent with a model in which some police officers have preferences for discrimination. However, Fryer’s work is complicated by a common difficulty in studying racial bias in policing, namely, the fact that officer decisions to initiate

contact are endogenous. Researchers do not observe cases in which officers choose not to make contact, which may be influenced by racial prejudice or other confounding factors (Dharmapala and Ross, 2004; Anwar and Fang, 2006). In addition, where Fryer seeks to overcome this difficulty by using a subset of Houston arrests as a benchmark case for where force might occur, he is only able to consider cases in which an officer discharges his firearm or his taser.

Recently, two papers have used samples of 911 calls to overcome the endogeneity of officer-civilian interactions. The use of this sample avoids selection effects associated with officer-initiated contact because 911 calls are made by civilians and officers are assigned to calls by dispatchers primarily based on location and availability. Weisburst (2019a) uses this type of sample to extend the work of Fryer by leveraging officer race in use-of-force incidents and by examining the relationship between arrests and use of force. Weisburst finds that conditional on arrest, use-of-force rates do not differ systematically by race.

Hoekstra and Sloan (2020) analyze over 2 million 911 calls in two different undisclosed cities to examine racial disparities in use of force. The authors use quasi-random variation in officer race to examine whether white officers are more likely to use force than Hispanic or black officers, finding that white officers are considerably more likely to use force overall. Additionally, Hoekstra and Sloan compare how white officers increase use of force when dispatched to neighborhoods with greater “minority” presence compared to how minority officers increase use of force. They find that white officers “increase use of any force much more than minority officers when dispatched to more minority neighborhoods.” In their paper, Hoekstra and Sloan rely on the assumption that 911 calls involve a civilian of the predominant race in the Census Block Group where the call was made. This allows the authors to analyze full samples of all 911 calls during their sample period for each city they consider.

In this paper, I contribute to the burgeoning use-of-force literature by examining racial disparities in use of force in the case study of Dallas, Texas. Following Weisburst (2019a) and Hoekstra and Sloan (2020), I utilize a sample of 1.4 million 911 calls made between 2014 and 2016 to avoid issues of officer-initiated contact. I begin by examining how officers of different races vary in their propensity to use force irrespective of civilian demographics. I leverage officer race to determine whether officers are more likely to use force against civilians of a different race when controlling for officer and call characteristics. Rather than making any assumptions about civilian race, I use a subsample of 911 calls resulting in arrest to benchmark cases where force is likely to occur. The arrest data identify civilian race as

well as other relevant characteristics used in my analysis.¹ In addition to looking at racial disparities in use of force, I also examine whether male and female officers are equally likely to use force.

This work builds on the larger literature of racial bias in policing which has mainly focused on vehicle drug searches (Anwar and Fang, 2006; Antonovics and Knight, 2009; West, 2018b) and stop-and-frisk policies (Weisburd et al., 2015; Goel et al., 2016). As with Fryer’s analysis, this literature is complicated by the endogeneity of officer-civilian interactions. A recent paper by West (2018b) circumvents this problem by studying the racial bias of state troopers who are randomly dispatched to highway accidents. West finds evidence of racial bias; state troopers are more likely to issue citations to opposite race drivers. Others use the “veil of darkness” provided by dusk in which officers can not see the race of the driver whom they pull over (Grogger and Ridgeway, 2006; Horrace and Rohlin, 2016). My work contributes to this literature by examining the arguably more contentious issue of use of force.

My work also builds on the relatively recent literature on the impact of women in policing and resulting improvements to policing quality. Rabe-Hemp (2008) examines under what conditions female officers behave differently from their male colleagues. Namely, her research finds that female officers are less likely to use “controlling behaviors” such as threats, physical restraint, and arrest than their male counterparts.² Miller and Segal (2018) study the integration of women into policing and its effects on violent crime reporting and domestic violence, finding that increased female representation in law enforcement leads to increases in reporting rates of violent crimes against women, particularly domestic violence. Additionally, they find significant declines in rates of intimate partner homicide and non-fatal domestic abuse. My work sheds light on whether there are benefits to more gender diversity in policing resulting from reduced likelihood of use of force in civilian encounters with female officers.

For this analysis I use an instrumental variables approach with officer availability as an instrument for the race/gender of the officer who responds to a service call. From the dispatch data, I am able to determine which officers are available to respond to individual service calls. Using this information, I identify officers as “available” if they are working a

¹Civilian race is not known for all 911 calls.

²Rabe-Hemp’s identification strategy, selection on observables, may not adequately address the endogeneity of officer-assignment. In fact, male and female officers in Rabe-Hemp’s sample differ in many observables. For example, female officers are more likely to respond to calls with female civilians and less likely to respond to calls where civilians are impaired (Rabe-Hemp, 2008).

particular shift in a certain location and are not preoccupied with another incident. For each call, I know the percentage of available officers by race and gender. The assumption is that officer availability by race/gender is exogenous to the race/gender of civilians conditional on observable call characteristics. Throughout the paper, I pay close attention to how officers are sorted to different calls for service. I conduct balance tests to demonstrate that my results are not the result of selection effects.

I find no significant difference in the likelihood of use of force across officer race overall. Calls involving female officers are 3.9 percentage points less likely to end in use of force. My results indicate that the presence of a white male officer in a 911 call involving a black civilian substantially increases the likelihood that force will be used. Specifically, such calls ending in arrest involving white officers and black civilians are 3.3 percentage points more likely to result in use of force. These results are statistically significant at the 1 percent level. The results of my replication analysis are broadly consistent with the results of Hoekstra and Sloan (2020) in interactions with white officers and black civilians. In these cases, I find that opposite-race officers are between 0.34 and 0.40 percentage points more likely to use force in general, and between 3.2 and 3.5 percentage points more likely to use force in the arrest sample.

Because the DPD data contain arrest information, I am able to evaluate the reliability of conditioning on arrest by researchers. For example, Fryer (2018) uses arrest data as a benchmark for cases where lethal force is considered more likely to be justified. In my replication analysis of Hoekstra and Sloan (2020), conditioning on arrest yields similar results for interactions between white officers and black civilians, but substantially different results for interactions between white officers and Hispanic civilians. This suggests that arrest data may not be a reliable benchmark in all cases.

In Section 2 of the paper, I discuss data sources and present summary statistics. Section 3 presents the identification strategy and main results. The section begins with a description of DPD dispatch and arrest protocol and how it informs my identification strategy. I describe the IV models and present the main results. Section 4 consists of replication of Hoekstra and Sloan (2020) and an analysis of the effect of officer timing for robustness. Section 5 contains heterogeneous effects of race and gender interactions for officers and civilians. Section 6 concludes.

1.2 Data Sources

The Dallas Police Department (DPD) serves over 1.3 million residents of the city of Dallas, 42 percent of whom are Hispanic, 24 percent black, and 29 percent white.³ In recent years, DPD has released several datasets in an effort to increase transparency. This project uses the publicly available “response to resistance” data which covers 2014 through 2016, which I refer to as the use-of-force data. The use-of-force data contain incidents in which officers report using physical force against civilians who are resisting officer authority. Officers who use force are identified by badge number. The data include descriptions of the type of force used, such as pushing, joint locks, armbars, taser use, and firearm use. The data also include descriptions of civilian resistance during the incident, such as active aggression, defensive resistance, and weapon display. Further, the data contain information on whether the officer or civilian was injured during the incident and if so, the severity and type of injuries.

I link the use-of-force data with dispatch data, arrest records, and officer demographic data obtained from records requests. Dispatch data contain information on each officer who responds to an incident including the time of dispatch, time en route, and time of arrival for each officer. The dispatch data also contain important call characteristics such as the reason for the call, call priority, date, and location. Dispatch data do not provide information regarding civilian characteristics. Arrest data contain information on each civilian arrested by DPD including race, gender, age, and charges. Arresting officers are identified by badge number. Use-of-force, dispatch, and arrest data are linked using a common identifier which I refer to as the incident number. Officer demographics including race, gender, age, and years of service are contained in separate data sets which I link to the others by badge number.

I consider only 911 calls that do not involve any officer-initiated contact.⁴ My full sample period runs from the latter half of 2014 through 2016.⁵ My data consist of 1,425,151 unique service calls. Of these calls, 37,682 involve the arrest of one or more individuals (i.e. 2.64 percent of all service calls result in arrest). There are 1,348 calls in which officers use force, or around 0.09 percent of all calls. Some incidents involve more than one civilian, and some civilians are involved in multiple separate incidents at different times.

³Throughout our analysis, black, Hispanic, and white are treated as mutually exclusive categories. Unlike many other data sources such as the U.S. Census, DPD does not treat Hispanic as an ethnicity. Thus, civilians are never classified as “white Hispanic” or likewise for other races.

⁴Most officers who respond to 911 calls are assigned by dispatchers. However, there are cases where these officers are assisted by others who are nearby and choose to respond of their own volition. I exclude these calls from my analysis.

⁵Incident numbers for use-of-force incidents before July 2014 are coded differently and I am unable to link these incidents to dispatch and arrest data.

From the complete dispatch data I identify which officers are available to respond to calls at the time they are made. I identify officers as able to respond to a call if they are on shift when the call comes in and are not preoccupied with other incidents. This allows me to calculate officer availability rates for each call by officer race and gender. Availability rates are used as a measure of the likelihood that an officer of a given race or gender responds to a call. These rates are the basis of my IV approach and are described more fully in the following section.

Table 1 presents summary statistics for call-level data divided into three samples. The first sample contains all 911 service calls in our data. The second contains only calls that result in arrest. The third contains only calls that result in force. Civilian characteristics are not available for all calls; this information is present only in arrest and use-of-force data. Significantly, more officers are dispatched to calls ending in arrest or use of force.

1.3 Identification Strategy & Results

1.3.1 Dispatch and Arrest Protocol

My research design and identification strategy are based primarily on the dispatch protocol used by DPD in the handling of 911 calls. Information on the handling of calls for service comes primarily from contacts within DPD who are familiar with the dispatch process. This information is supplemented by publicly available information from the Dallas Public Safety Committee (2016) as well as the descriptions of DPD dispatch and officer assignment by Weisburst (2019b).

Citizens in Dallas who call 911 for police assistance are connected to a 911 operator. The 911 operator creates an active call report which summarizes the basic information of the call, including location, description of events, and time. Each call report includes a categorization of the incident type that is given by a dispatch code. These call categories are simple, literal descriptions of the event, such as “active shooter,” or “domestic disturbance.”⁶ Some calls are sub-categorized further when this is relevant for dispatch purposes. For instance, a burglary may be coded as in progress, recent, or at an unknown time. The operator is also responsible for identifying the priority of each service call. Priority is measured discretely from one to four, with one being the highest priority and four being the lowest.⁷ Level-one

⁶There are 23 different dispatch codes in the data, although there are more available to call-takers when writing active call reports. I exclude certain categories such as police response to traffic accidents.

⁷In fact, there are also level-five priority calls which are described by DPD as “telephone service calls” such as reports of lost property or non-critical missing persons. There are no DPD guidelines for officer

calls are emergencies such as shootings and kidnappings in progress for which the goal is officer response in eight minutes or less. Level-two calls are considered “prompt” calls such as robberies and criminal assaults. Here, the goal is officer response in 12 minutes or less. Level-three calls are general service calls such as intoxicated person, drug house, or recent burglary. The goal for level-three calls is officer response in 30 minutes or less. Finally, level-four calls are considered non-critical incidents such as loud music, theft, criminal mischief, and animal complaints. The goal for level-four calls is officer response in 60 minutes or less. The 911 operator inputs all call information into the Computer Aided Dispatch System where it is electronically routed to the dispatch queue for use by dispatchers.

Dispatchers look at the queue of active calls and assign available officers to active incidents. Dispatchers are able to locate officers on a map and identify which officers are available based on whether the officers are involved with any other incidents at the time of the call. Officers are assigned primarily based on location and availability. If there are many more active calls than available officers, lower-priority calls are postponed until higher-priority calls are resolved. Dispatchers also decide the number of officers to assign to a call based on the type of call. More serious incidents (such as shootings and mental health calls) may involve the dispatch of multiple officers.

The primary responders to calls for service are patrol officers, who work 8-hour shifts in one of seven divisions in the city.⁸ Divisions are further divided into 234 different beats. Patrol shift schedules are set once a year based on officer seniority. Patrols are typically conducted in police cars, with officers either alone or in pairs. Officers who are geographically close to an incident are more likely to be dispatched to the incident. Proximity is especially important for urgent calls.

Overwhelmingly, calls are assigned to patrol officers who work within the division where the call occurred. Figure 1 shows the proportion of calls the typical officer is assigned to in his n th most common division. This figure shows that around 95 percent of the time, the typical officer responds to 911 calls in his most common division. Officers are rarely pulled from their primary division to respond to 911 calls in other divisions. Similarly, Figure 1 shows the proportion of calls the typical officer is assigned to in his n th most common beat. Officers are dispersed much more broadly across the beats within a division. I take this as evidence that officers are assigned to calls as-good-as-randomly within a division.

response to these calls. They are not reported in the data.

⁸These shifts are approximately 12AM to 8AM, 8AM to 4PM, and 4PM to 12AM. However, some officers work 10-hour shifts, and shifts are slightly staggered so that officers do not all change shifts at the same time.

Table 1: Call-Level Summary Statistics

	(1)		(2)		(3)	
	All Calls		Arrest Sample		Force Sample	
Number of Calls	1,425,151		37,682		1,348	
Number of Arrests	37,682		37,682		1,232	
Proportion Resulting in Arrest	2.64%		100%		91.39%	
Number of Force Incidents	1,348		1,232		1,348	
Proportion Resulting in Force	0.09%		3.27%		100%	
<i>Call-Level Variables</i>						
Priority	2.64	(0.81)	2.19	(0.65)	1.96	(0.63)
Minutes Between Call and Dispatch	38.43	(77.45)	16.22	(33.21)	12.14	(33.09)
Day of Week						
Sunday	0.15	(0.36)	0.16	(0.36)	0.18	(0.39)
Monday	0.14	(0.35)	0.14	(0.35)	0.13	(0.34)
Tuesday	0.13	(0.34)	0.13	(0.34)	0.13	(0.34)
Wednesday	0.13	(0.34)	0.13	(0.33)	0.12	(0.33)
Thursday	0.14	(0.34)	0.14	(0.34)	0.11	(0.32)
Friday	0.15	(0.36)	0.15	(0.36)	0.14	(0.35)
Saturday	0.16	(0.36)	0.16	(0.37)	0.18	(0.38)
Watch						
8AM-4PM	0.33	(0.47)	0.28	(0.45)	0.2	(0.4)
4PM-12AM	0.42	(0.49)	0.42	(0.49)	0.42	(0.49)
12AM-8AM	0.25	(0.43)	0.3	(0.46)	0.38	(0.49)
Black Officer Present	0.37	(0.48)	0.47	(0.5)	0.56	(0.5)
Hispanic Officer Present	0.32	(0.47)	0.47	(0.5)	0.59	(0.49)
White Officer Present	0.63	(0.48)	0.78	(0.42)	0.9	(0.31)
Male Officer Present	0.91	(0.28)	0.96	(0.19)	0.98	(0.13)
Female Officer Present	0.25	(0.43)	0.42	(0.49)	0.48	(0.5)
Officer Age	38.8	(8.3)	37.4	(6.6)	36.6	(5.3)
Officer Years of Service	11.5	(7.7)	10.1	(6)	9.6	(4.7)
Number of Officers Dispatched	2.21	(1.42)	4.1	(3.4)	7.04	(5.5)
Number of Officers Arrived	1.89	(1.16)	3.21	(2.49)	5.07	(4.02)
<i>Civilian Characteristics</i>						
Proportion Black Civilian	-	-	0.5	(0.5)	0.53	(0.5)
Proportion Hispanic Civilian	-	-	0.27	(0.44)	0.25	(0.43)
Proportion White Civilian	-	-	0.22	(0.41)	0.2	(0.4)
Proportion Male Civilian	-	-	0.8	(0.39)	0.82	(0.37)
Proportion Female Civilian	-	-	0.2	(0.39)	0.18	(0.37)
Civilian Age	-	-	35.1	(12.1)	31.9	(10.8)

Standard deviations in parentheses. Officers of a race other than black, Hispanic, or white are present in approximately one percent of calls. Civilian characteristics including race, gender, and age are available only in the arrest and use-of-force data. Summary statistics can not be provided for these variables for the set of all calls.

When a dispatched officer arrives at the location of the incident, he gathers information and determines whether an offense occurred. The officer assists the complainant or victim.

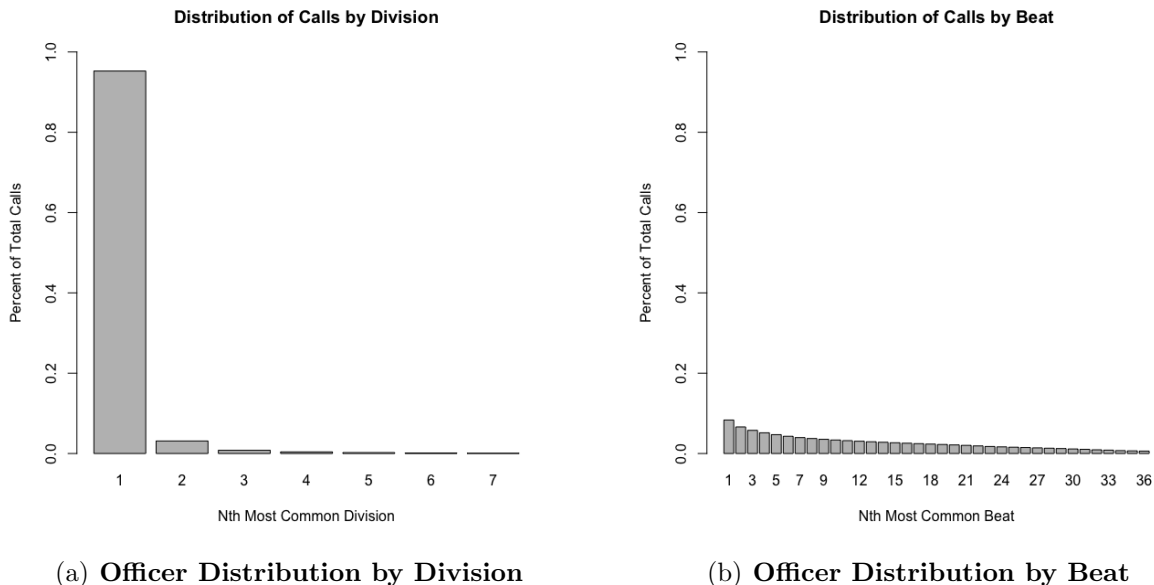


Figure 1: Percent of Calls in Average Officer's Most Common Division/Beat

In the event that an offense occurred, the officer writes an offense report and submits it for review with the DPD. Additional investigation may occur later if the offense is severe enough to warrant it. During the course of the initial officer response, the officer may identify a suspect and decide whether to make an arrest or not. Around 85 percent of all use-of-force incidents in the data occur during the course of an arrest.

1.3.2 Models

In the main analysis, I consider the effect of officer race and gender on the likelihood of use of force occurring during a 911 call. I use an instrumental variables approach to examine this outcome in two different ways. First, I consider differences in the propensity to use force across officer race and gender, irrespective of civilian demographics. Second, I consider how the likelihood of use of force varies when officer race differs from civilian race.

I use the dispatch availability of officers by race and gender as an instrument for whether an officer of that race or gender is present at an incident. More precisely, using the example of black officer availability, my instrument is constructed as the number of black officers available to respond to a service call divided by the total number of officers on shift as in Equation (1), where c indexes the service call. Officers are considered available if they are working during the shift that the call comes in and are not preoccupied with another call. The instrument is constructed similarly for Hispanic, white, male, and female officers. I use

this instrument to deal with potential selection effects from officer sorting that can not be accounted for by control variables and fixed effects.

$$BlackAvailRate_c = \frac{Number\ of\ Available\ Black\ Officers_c}{Number\ of\ Officers\ on\ Shift_c} \quad (1)$$

In support of the IV approach, I conduct balance tests to determine whether my instruments are correlated with other observable call characteristics. I regress each instrument on a set of nine call characteristics including the following: proportions of civilians by race/gender; civilian age; minutes between call and dispatch; call priority; median household income; and the proportion of residents with less than a high school diploma in the Census block group where the call occurred. Balance test results are presented in Table 2. Three of the 36 coefficients are statistically significant, a number consistent with random chance.

Table 2: **Correlation Between Instruments and Covariates**

	(1)	(2)	(3)	(4)
	Black Avail. Rate	Hisp. Avail. Rate	White Avail. Rate	Female Avail. Rate
Prop. Black Civ.	0.028 (0.048)	0.005 (0.055)	-0.056 (0.03)	-0.051 (0.065)
Prop. Hisp. Civ.	-0.01 (0.072)	-0.051 (0.048)	0.092** (0.033)	0.023 (0.072)
Prop. White Civ.	-0.033 (0.034)	0.054 (0.032)	-0.015 (0.026)	0.021 (0.042)
Prop. Female Civ.	-0.104** (0.041)	-0.014 (0.033)	0.053 (0.032)	0.071 (0.047)
Civilian Age	-0.872 (1.65)	1.677 (2.41)	-1.179 (1.614)	4.017** (1.382)
Minute Diff.	-3.738 (2.646)	5.288 (5.344)	-3.476 (3.315)	-2.13 (5.237)
Priority	-0.04 (0.051)	-0.257 (0.164)	-0.071 (0.08)	0.252 (0.152)
Median HH Income	-909.37 (2,232.28)	-1,052.02 (7,636.7)	-2,447.76 (6,318.44)	-5,020.38 (7,594.89)
Prop. Less than HS Degree	-0.015 (0.022)	0.032 (0.032)	-0.002 (0.016)	0.013 (0.029)
Number of Observations	37,359	37,359	37,359	37,359

This table reports coefficients from regressions of call characteristics on the instruments. Each call characteristic is regressed separately on each instrument. Regressions include day of week, hour of day, and division-year-week-shift fixed effects. Standard deviations in parentheses.

To see how the model addresses endogeneity, first consider the following OLS regression which estimates the difference in overall use of force between calls in which a black officer is

present and calls with no black officer:

$$\begin{aligned}
ForceUsed_c = & \beta_0 + \beta_1(BlackOfficer)_c + \delta X_c + \lambda_1(DayOfWeek)_c + \\
& \lambda_2(HourOfDay)_c + \lambda_3(DivisionYearWeekShift)_c + \epsilon_c,
\end{aligned}
\tag{2}$$

where c indexes the service call.⁹ The outcome variable, $ForceUsed_c$, is a binary variable equal to one when the call ends in use of force and zero otherwise. $BlackOfficer_c$ is a binary variable which indicates whether a black officer is present at the call. X_c is a set of call characteristics which includes the reason for the call, priority, and time between call and dispatch. The model also includes fixed effects for day of week and hour of day, as well as $Division * Year * Week * Shift$ fixed effects. The choice of these fixed effects is informed by the dispatch protocol described above.

The OLS specification will not address any remaining endogeneity that comes from officer sorting that is not already accounted for by our controls. One possibility is that there is an unwritten DPD policy affecting how officers are assigned to calls, such as a proclivity to send female officers to domestic violence calls. There may be other unobservables that influence whether black officers are likely to be dispatched to a call. These unobservables may also influence the likelihood that force is used irrespective of officer characteristics. For example, it may be the case that black officers are more likely to sort into less confrontational calls, which in turn will have a lower likelihood that force is used.

The IV approach avoids potential selection effects by considering how black officers are more likely to be assigned to calls based on availability. I run the following first stage:

$$\begin{aligned}
BlackOfficer_c = & \alpha_0 + \alpha_1(BlackAvailRate)_c + \theta X_c + \lambda_1(DayOfWeek)_c + \\
& \lambda_2(HourOfDay)_c + \lambda_3(DivisionYearWeekShift)_c + \epsilon_c,
\end{aligned}
\tag{3}$$

where c again indexes the service call. The outcome variable $BlackOfficer$ is a binary variable equal to one if a black officer shows up to a service call and zero otherwise. $BlackAvailRate$ is the instrument which is described above. Controls and fixed effects are the same as in Equation (2).

⁹Note that the model is similar for white, Hispanic, male, and Female officers, as are the rest of the equations in this chapter.

I run the following IV model:

$$ForceUsed_c = \beta_0 + \beta_1(\widehat{BlackOfficer})_c + \delta X_c + \lambda_1(DayOfWeek)_c + \lambda_2(HourOfDay)_c + \lambda_3(DivisionYearWeekShift)_c + \epsilon_c, \quad (4)$$

where $\widehat{BlackOfficer}_c$ represents the fitted values from the first stage.

1.3.3 Main Results

This section begins with results for the average effect of officer race and gender on use of force. These results are presented in Table 3. Panels A through D display results for the effect of each officer type on use of force. Each column displays the results of a different model. Column 1 presents results from the OLS estimations described in Equation (2) for each officer race and gender. Results for the first stage presented in Column 2 indicate that the instruments are strong. Reduced form and IV results are presented in Columns 3 and 4, respectively. The OLS results differ substantially from the IV results in sign, magnitude, and significance. Based on the main specification in Column 4, I do not find significant evidence that arrests involving the presence of black, Hispanic, or white officers vary in use of force. However, arrests with a female officer present are less likely to result in force.

Next, I consider the effect of officer race on use of force by civilian race. These models estimate the effect of officer race on whether force is used against a black, Hispanic, or white civilian. Each panel represents results for the presence of at least one officer of the given race (e.g. arrests in which at least one black officer is present). Each column represents a different dependent variable. For example, in Column 1 the dependent variable is a binary variable indicating whether force is used against a black civilian or not. Columns 2 and 3 present similar results for force used against Hispanic and white civilians, respectively.

These results suggest that there is a strongly significant increase in use of force against black civilians when white officers are present at arrest. There is also suggestive evidence that the presence of a Hispanic officer at an arrest reduces the likelihood that force will be used against a Hispanic civilian. These results are consistent with those of Hoekstra and Sloan (2020) and Fryer (2018).

Table 3: **Effect of Officer Race & Gender on Use of Force**

	(1)	(2)	(3)	(4)
Dependent Variable:	OLS	First Stage	Reduced Form	IV
	Force Used	B/H/W/F Officer Present	Force Used	Force Used
<i>Panel A:</i>				
Black Officer Present	0.009*** (0.002)	- -	- -	-0.018 (0.017)
Black Availability Rate	- -	0.894*** (0.13)	-0.016 (0.017)	- -
Mean of Dependent Variable:	0.033	0.474	0.033	0.033
First Stage F Statistic	-	47.01	-	-
Percent of All Arrests	47%	47%	47%	47%
<i>Panel B:</i>				
Hispanic Officer Present	0.015*** (0.001)	- -	- -	-0.000 (0.020)
Hispanic Availability Rate	- -	1.366*** (0.117)	0.000 (0.03)	- -
Mean of Dependent Variable:	0.033	0.47	0.033	0.033
First Stage F Statistic	-	135.40	-	-
Percent of All Arrests	47%	47%	47%	47%
<i>Panel C:</i>				
White Officer Present	0.021*** (0.003)	- -	- -	0.034 (0.022)
White Availability Rate	- -	0.647*** (0.071)	0.022 (0.016)	- -
Mean of Dependent Variable:	0.033	0.777	0.033	0.033
First Stage F Statistic	-	82.06	-	-
Percent of All Arrests	78%	78%	78%	78%
<i>Panel D:</i>				
Female Officer Present	0.008** (0.002)	- -	- -	-0.039*** (0.014)
Female Availability Rate	- -	1.477*** (0.094)	-0.058** (0.023)	- -
Mean of Dependent Variable:	0.033	0.421	0.033	0.033
First Stage F Statistic	-	249.44	-	-
Percent of All Arrests	42%	42%	42%	42%
Call Controls	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes
Division-Year-Week-Shift FE	Yes	Yes	Yes	Yes
Observations	37,682	37,359	37,359	37,359

This table reports coefficients from 16 different regressions. Each column represents a different model and each panel presents results for a different officer type (i.e. specific race or gender). Robust standard errors are clustered at the division level. Kleibergen-Paap first-stage F statistics are presented here. Percent of all arrests indicates the percent of arrests involving at least one officer of the given type (e.g. at least one black officer present).

1.4 Robustness

1.4.1 Replication of Hoekstra & Sloan (2020)

In this section, I evaluate the robustness of the main results. I begin by replicating the analysis done by Hoekstra and Sloan (2020). Hoekstra and Sloan estimate the effect of

Table 4: **Effect of Officer Race on Use of Force by Civilian Race**

Force Used Against:	(1) Black Civilian	(2) Hispanic Civilian	(3) White Civilian
<i>Panel A:</i>			
Black Officer Present	-0.014 (0.011)	-0.002 (0.006)	-0.000 (0.014)
Mean of Dependent Variable:	0.018	0.008	0.007
First Stage F Statistic	47.01	135.40	82.06
Percent of All Arrests	27%	12%	9%
<i>Panel B:</i>			
Hispanic Officer Present	0.001 (0.006)	-0.011* (0.007)	0.011 (0.013)
Mean of Dependent Variable:	0.018	0.008	0.007
First Stage F Statistic	47.01	135.40	82.06
Percent of All Arrests	23%	14%	10%
<i>Panel C:</i>			
White Officer Present	0.033*** (0.010)	0.005 (0.010)	-0.003 (0.016)
Mean of Dependent Variable:	0.018	0.008	0.007
First Stage F Statistic	47.01	135.40	82.06
Percent of All Arrests	38%	22%	19%
Call Controls	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes
Division-Year-Week-Shift FE	Yes	Yes	Yes
Observations	37,682	37,359	37,359

This table reports coefficients from 9 different regressions. Each column represents a different dependent variable and each panel presents results for a different officer type (i.e. specific race or gender). Robust standard errors are clustered at the division level. Kleibergen-Paap first-stage F statistics are presented here. Percent of all arrests indicates the percent of arrests involving both an officer and civilian type. For example, 27 percent of arrests involve a black officer and black civilian.

officer race on use of force using data for 911 calls from two undisclosed cities. The first city is predominantly black and white, while the second city is predominantly Hispanic and white. The race of the civilian involved in each call is not known in either city. Hoekstra and Sloan address this problem by using the race of the Census Block Group from the call location. The authors “compare how the probability of using force changes for white officers dispatched to white and black neighborhoods, compared to that of black officers.” They find in the first city that white officers use force 60 percent more often than black officers on average and that dispatching opposite-race officers increases use of force by 30 to 60 percent. In the second city, they find that white and Hispanic officers use force at similar overall rates, but white officers increase use of force more than Hispanic officers when dispatched

to Hispanic neighborhoods. They estimate that dispatching an opposite-race officer in this second city roughly doubles the likelihood of use of force.

Hoekstra and Sloan use a difference-in-differences model as their main specification. I modify their specification to account for the idiosyncrasies of DPD dispatch protocol as follows:

$$\begin{aligned}
 ForceUsed_{ic} = & \beta_0 + \beta_1(ProportionBlackCivilians)_c + Officer_i + \\
 & \beta_2(WhiteOfficer * ProportionBlackCivilians)_{ic} + \quad (5) \\
 & Division * Year * Week * Shift_c + DayOfWeek_c + HourOfDay_c + X_c + \epsilon_{ic},
 \end{aligned}$$

where $ProportionBlackCivilians_c$ is the proportion of black civilians in the Census Block Group of the call location, $Officer_i$ is an individual officer fixed effect, and X_c is a set of call and officer controls.¹⁰

These results are presented in Table 5. Because Dallas is relatively more diverse than the cities used by Hoekstra and Sloan, I stratify the results by neighborhood type. In Panel A, I use a sample of neighborhoods in Dallas that are at least 80 percent black or white. In Panel B, I use a sample of neighborhoods that are at least 80 percent Hispanic or white. Because the DPD data contain arrest records, I am able to evaluate the reliability of conditioning on arrest when civilian race data is unavailable for all officer-civilian interactions. This sheds light on whether the endogeneity of conditioning on arrest results in significantly different findings. I present the results for both samples. Following Hoekstra and Sloan, I present the results using different levels of controls. Columns 1 and 4 include $Division * Year * Week * Shift$, $DayOfWeek$, and $HourOfDay$ fixed effects. Columns 2 and 5 add officer fixed effects. Columns 3 and 6 include the full set of fixed effects and controls, adding call characteristics including call priority, latitude and longitude, time between call and dispatch, median household income, proportion with less than a high school diploma, and call type.

The results in Panel A for black and white neighborhoods are similar across specification. Estimates for the arrest sample are roughly an order of magnitude larger. These results are positive and significant at the one percent level. Consistent with the main results as well as the results of Hoekstra and Sloan, opposite-race officers are significantly more likely to use force. Specifically, opposite-race officers in these neighborhoods are between 0.34 and 0.40 percentage points more likely to use force in the full sample. Compared to the mean use of

¹⁰Specifically, I use different fixed effects. Hoekstra and Sloan use $Beat * Year * Week * Shift$, while I use $Division * Year * Week * Shift$ fixed effects as well as $DayOfWeek$ and $HourOfDay$ fixed effects.

Table 5: Replication of Hoekstra & Sloan Results

	All Calls			Arrests		
	(1)	(2)	(3)	(4)	(5)	(6)
	Use of Force	Use of Force	Use of Force	Use of Force	Use of Force	Use of Force
<i>Panel A: Black and White Neighborhoods</i>						
White Officer	0.0036398*** (0.0004256)	0.0040296*** (0.000485)	0.0033725*** (0.0090941)	0.0348389*** (0.0094677)	0.032267*** (0.0094677)	0.034825*** (0.0095866)
Mean of Dependent Variable:	0.0032672	0.0032672	0.0032672	0.0526283	0.0526283	0.0526283
Observations	830,069	830,069	809,058	48,244	48,244	46,364
<i>Panel B: Hispanic and White Neighborhoods</i>						
White Officer	0.0035101*** (0.0006915)	0.0038477*** (0.0007959)	0.0020501** (0.0007957)	-0.0333014** (0.0160475)	-0.039769** (0.0166799)	-0.0463022*** (0.0170296)
Mean of Dependent Variable:	0.0033171	0.0033171	0.0033171	0.0542537	0.0542537	0.0542537
Observations	1,250,779	1,250,779	1,205,032	72,032	72,032	69,044
Division-Year-Week-Shift FE	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Officer FE	-	Yes	Yes	-	Yes	Yes
Officer Controls	-	-	Yes	-	-	Yes
Call Controls	-	-	Yes	-	-	Yes

This table presents coefficients from regressions designed to replicate those in Hoekstra and Sloan (2020). The reported coefficients are from the regression of *Force Used* on the interaction of *White Officer* and *Block Group Proportion Black*. Panel A considers only neighborhoods which are greater than 80 percent black and white. Panel B considers only neighborhoods which are greater than 80 percent Hispanic and white. Results are presented for both the sample of all calls and the sample of arrests only. Robust standard errors are clustered at the officer level.

force, these results imply increases of 103 to 123 percent. For the arrest sample, opposite race officers are between 3.2 and 3.5 percentage points more likely to use force, an increase of 61 to 66 percent.

The results in Panel B for Hispanic and white neighborhoods are similar across specification but not across sample. Columns 1 through 3 indicate that opposite-race officers are significantly more likely to use force by between 0.21 and 0.38 percentage points, an increase of 62 to 116 percent. However, in columns 4 through 6 the effects are significant but negative, suggesting that opposite-race officers reduce the use of force by between 3.3 and 4.6 percentage points, implying a decrease of 61 to 85 percent. These results suggest that conditioning on arrest can sometimes lead to different results.

1.4.2 Officer Timing

In this section, I consider whether the main results are driven by officers who arrive first or those who are dispatched as backup. “First arrivers” are identified by the minute they were dispatched to a 911 call. This allows more detailed consideration of the conditions surrounding scenarios that result in statistically higher use of force. Recall from the main results that calls with a female officer present are significantly less likely to result in use of force. This finding could be spurious if, for example, female officers are assigned as backup to relatively benign calls.

Table 6: **Effect of Officer Race on Use of Force by Civilian Race**

Force Used Against:	(1)	(2)	(3)	(4)
	Overall	Black Civilian	Hispanic Civilian	White Civilian
<i>Panel A:</i>				
Black Officer Present	-0.016 (0.018)	-0.016 (0.010)	-0.002 (0.006)	-0.000 (0.014)
Mean of Dependent Variable:	0.031	0.018	0.008	0.007
First Stage F Statistic	47.013	47.013	47.013	47.013
Percent of All Arrests	47%	27%	12%	9%
<i>Panel B:</i>				
Hispanic Officer Present	-0.001 (0.022)	0.001 (0.006)	-0.011* (0.007)	0.011 (0.013)
Mean of Dependent Variable:	0.031	0.018	0.008	0.007
First Stage F Statistic	135.402	135.402	135.402	135.402
Percent of All Arrests	47%	23%	14%	10%
<i>Panel C:</i>				
White Officer Present	0.029 (0.022)	0.034*** (0.011)	0.005 (0.010)	-0.003 (0.016)
Mean of Dependent Variable:	0.031	0.018	0.008	0.007
First Stage F Statistic	82.056	82.056	82.056	82.056
Percent of All Arrests	78%	38%	22%	19%
<i>Panel D:</i>				
Female Officer Present	-0.034*** (0.011)			
Mean of Dependent Variable:	0.031			
First Stage F Statistic	249.442			
Percent of All Arrests	42%			
Call Controls	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes
Division-Year-Week-Shift FE	Yes	Yes	Yes	Yes
Observations	37,359	37,359	37,359	37,359

This table reports coefficients from regressions that account for officer timing. Each column represents a different dependent variable and each panel presents results for a different officer type (i.e. specific race or gender). Robust standard errors are clustered at the division level. Kleibergen-Paap first-stage F statistics are presented here. Percent of all arrests indicates the percent of arrests involving both an officer and civilian type. For example, 27 percent of arrests involved a black officer and black civilian.

The models used in this section are nearly identical to those used for the main results. However, the dependent variable here is equal to one only if force was used by an officer in the first group dispatched. The variable of interest here is whether a black/Hispanic/white/female officer is present in the first dispatched group of officers. The results of these models are presented in Table 6 and are nearly identical to those in Table 4. I find no significant difference in the likelihood of force across officer race overall. Calls involving female officers are less likely to result in use of force. Calls involving white officers significantly increase the

likelihood of force against black civilians.

1.5 Heterogeneous Effects

In this section, I explore the heterogeneous effects of race and gender for both officers and civilians. Model specification is the same as in Equations (3) and (4) except the models now also account for officer and civilian gender. Results for the 36 unique types of interactions (e.g. black male officer and black male civilian, black male officer and black female civilian,...) are presented in Tables 7 through 9. Each table presents results for a particular officer race further divided into two panels by officer gender. Certain officer-civilian interactions occur infrequently. For example, only 1 percent of all arrests involve a black female officer and Hispanic female civilian. One should be cautious in drawing inferences from these relatively rare cases.

Table 7 presents suggestive evidence that arrests involving black male officers and Hispanic male civilians are less likely to result in force, while those involving black male officers and white male civilians are more likely to result in force. These results are significant at the 10 percent level. There is no evidence that arrests involving black female officers vary in use of force across civilian race/gender. Results in Table 8 suggest that arrests with Hispanic male officers and black male civilians are more likely to result in force, the only statistically significant finding for Hispanic officers.

Table 7: Effect of Black Officers on Use of Force by Civilian Race/Gender

Use of Force Against:	(1) Black Male	(2) Black Female	(3) Hisp. Male	(4) Hisp. Female	(5) White Male	(6) White Female
<i>Panel A: IV</i>						
Black Male Present	0.006 (0.017)	0.003 (0.01)	-0.025* (0.014)	-0.011 (0.008)	0.02* (0.01)	0.007 (0.006)
Mean of Dependent Variable	0.014	0.004	0.008	0.001	0.005	0.002
Percent of All Arrests	18%	5%	9%	2%	6%	2%
<i>Panel A: First Stage</i>						
Black Male Availability	0.451*** (0.103)	0.451*** (0.103)	0.451*** (0.103)	0.451*** (0.103)	0.451*** (0.103)	0.451*** (0.103)
First Stage F Statistic	19.18	19.18	19.18	19.18	19.18	19.18
Observations	37,329	37,329	37,329	37,329	37,329	37,329
<i>Panel B: IV</i>						
Black Female Present	0.03 (0.043)	0.001 (0.028)	0.026 (0.039)	0.003 (0.009)	0.03 (0.024)	0.014 (0.015)
Mean of Dependent Variable	0.014	0.004	0.008	0.001	0.005	0.002
Percent of All Arrests	6%	3%	2%	1%	1%	1%
<i>Panel B: First Stage</i>						
Black Female Availability	0.108*** (0.014)	0.108*** (0.014)	0.108*** (0.014)	0.108*** (0.014)	0.108*** (0.014)	0.108*** (0.014)
First Stage F Statistic	61.96	61.96	61.96	61.96	61.96	61.96
Observations	32,265	32,265	32,265	32,265	32,265	32,265
Call Controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Division-Year-Week-Shift FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 8: **Effect of Hispanic Officers on Use of Force by Civilian Race/Gender**

	(1)	(2)	(3)	(4)	(5)	(6)
Use of Force Against:	Black Male	Black Female	Hisp. Male	Hisp. Female	White Male	White Female
<i>Panel A: IV</i>						
Hisp. Male Present	0.026*	0.004	-0.014	-0.008	0.00	0.007
	(0.014)	(0.004)	(0.015)	(0.005)	(0.013)	(0.014)
Mean of Dependent Variable	0.014	0.004	0.008	0.001	0.005	0.002
Percent of All Arrests	16%	4%	11%	2%	7%	3%
<i>Panel A: First Stage</i>						
Hisp. Male Availability	0.537***	0.537***	0.537***	0.537***	0.537***	0.537***
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)
First Stage F Statistic	148.15	148.15	148.15	148.15	148.15	148.15
Observations	37,343	37,343	37,343	37,343	37,343	37,343
<i>Panel B: IV</i>						
Hisp. Female Present	-0.04	-0.017	-0.029	-0.004	0.03	0.027
	(0.052)	(0.016)	(0.019)	(0.009)	(0.015)	(0.019)
Mean of Dependent Variable	0.014	0.004	0.008	0.001	0.005	0.002
Percent of All Arrests	4%	2%	3%	1%	1%	1%
<i>Panel B: First Stage</i>						
Hisp. Female Availability	0.119***	0.119***	0.119***	0.119***	0.119***	0.119***
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
First Stage F Statistic	27.12	27.12	27.12	27.12	27.12	27.12
Observations	30,118	30,118	30,118	30,118	30,118	30,118
Call Controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Division-Year-Week-Shift FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 9: **Effect of White Officers on Use of Force by Civilian Race/Gender**

	(1)	(2)	(3)	(4)	(5)	(6)
Use of Force Against:	Black Male	Black Female	Hisp. Male	Hisp. Female	White Male	White Female
<i>Panel A: IV</i>						
White Male Present	0.06*	0.011**	-0.002	-0.003	0.021	0.02*
	(0.031)	(0.004)	(0.015)	(0.007)	(0.018)	(0.011)
Mean of Dependent Variable	0.014	0.004	0.008	0.001	0.005	0.002
Percent of All Arrests	28%	8%	18%	3%	13%	5%
<i>Panel A: First Stage</i>						
White Male Availability	0.633***	0.633***	0.633***	0.633***	0.633***	0.633***
	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)	(0.075)
First Stage F Statistic	72.05	72.05	72.05	72.05	72.05	72.05
Observations	37,359	37,359	37,359	37,359	37,359	37,359
<i>Panel B: IV</i>						
White Female Present	0.044***	-0.028*	0.006	-0.007	0.02**	-0.003
	(0.01)	(0.015)	(0.013)	(0.006)	(0.01)	(0.013)
Mean of Dependent Variable	0.014	0.004	0.008	0.001	0.005	0.002
Percent of All Arrests	6%	3%	4%	1%	3%	2%
<i>Panel B: First Stage</i>						
White Female Availability	0.216	0.216	0.216	0.216	0.216	0.216
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
First Stage F Statistic	92.51	92.51	92.51	92.51	92.51	92.51
Observations	35,933	35,933	35,933	35,933	35,933	35,933
Call Controls	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Division-Year-Week-Shift FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 9 presents results for white officers. These results suggest that arrests with a white male officer present are more likely to result in use of force against black males and black females. For arrests involving white female officers, force is more likely to be used against black males but less likely to be used against black females. Additionally, arrests with a white female officer present result in an increased likelihood of use of force against white males.

1.6 Conclusion

Recent high-profile policing controversies have heightened attention on police use of force and racial discrimination by law enforcement. Despite the widespread attention these issues receive from media and the public, little empirical research has been done on this topic. Two recent articles stand out. First, Fryer (2018) uses data from four different samples across the United States to study racial bias in police use of force. Fryer finds that blacks and Hispanics are more likely to experience non-lethal uses of force. However, he finds no evidence of racial bias in use of lethal force when contextual factors are taken into account. Fryer argues that his results are consistent with racial discrimination by some police officers. However, Fryer's paper is complicated by the fact that his samples contain incidents initiated by officers. The decision by officers to initiate contact with civilians may be motivated by racial bias, thereby complicating the analysis. Fryer seeks to address this problem by using a subset of Houston arrests as a benchmark case for situations where force might be expected. However, Fryer is only able to consider cases in which an officer discharges his firearm or his taser.

A recent paper by Hoekstra and Sloan (2020) examines officer use of force in two undisclosed cities, one which is mainly black and white and the other which is mainly Hispanic and white. Using a sample of 911 calls, Hoekstra and Sloan avoid the problems associated with officer-initiated interactions. They find that white officers are considerably more likely to use force overall. Additionally, they find that white officers increase use of force much more than minority officers when dispatched to minority neighborhoods. Specifically, they find that dispatching opposite-race officers to more minority neighborhoods increases use of force by 30 to 60 percent in the first city. In the second city, they find that dispatching an opposite-race officer roughly doubles the likelihood of use of force.

This paper builds on the nascent use of force literature using a case study of 911 calls in Dallas, Texas from 2014 through 2016. Using an instrumental variables method with officer availability by race/gender as the instrument, I consider two main research questions. First, I examine whether officers vary in their overall propensity to use force by race and gender. Second, I examine the effect of the interaction of officer and civilian race on the likelihood of use of force. For the former, I find no significant differences in the propensity to use force by officer race. However, I find that calls involving female officers are less likely to result in use of force by roughly 3.3 percentage points. Results regarding the interaction of officer and civilian race are broadly consistent with those of Hoekstra and Sloan (2020). Namely, I find that calls involving white officers are significantly more likely to result in use of force against black civilians. Conditional on arrest, calls involving white officers are between 3.2

and 3.5 percentage points more likely to result in use of force (an increase between 61 and 67 percent over the mean). Unconditionally, replicating the results of Hoekstra and Sloan (2020), I find that calls involving white officers increase the likelihood of force by 0.34 and 0.4 percentage points (an increase of roughly 103 to 123 percent over the mean).

A unique contribution of this paper is the inclusion of arrest data. This arrest data allows me to shed light on the reliability of using arrests as a benchmark when civilian race is not otherwise known. In replicating the results of Hoekstra and Sloan (2020), I find that my results are generally consistent across sample for neighborhoods that are predominantly black and white.¹¹ These results are also consistent with the main results from my IV method. However, the results of the analysis change significantly when considering neighborhoods that are predominantly Hispanic and white. This suggests that researchers must be careful in using arrests as a benchmark case.

The main findings of this paper are consistent with the popular perception that police officers are racially biased in their use of force. However, this is not the only possible interpretation. In this paper, I identify the effect of the interaction of officer and civilian race on the likelihood of use of force. Without a better understanding of civilian behavior in these encounters, there is not sufficient evidence to support the claim that officers are racially biased. For instance, civilians may be more likely to resist arrest or exhibit confrontational behavior towards officers of a different race. To address this complication, more granular data including civilian resistance are necessary.

¹¹While the results for the arrest sample are roughly an order of magnitude larger, this is to be expected given that the baseline use of force conditional on arrest is also an order of magnitude higher.

2 Chapter 2: Is Police Training an Effective Intervention for Addressing Disparities?

Coauthored with Steve Mello, Matthew B. Ross, and Stephen L. Ross

2.1 Introduction

The possibility that police treat minorities differently than their non-Hispanic white peers has recently become a source of political protest and social unrest in the United States. The sentiment that law enforcement in America is characterized by racial prejudice is widely held, with blacks reporting much lower confidence in police than whites (Pew Research Center, 2020). Police training is often cited by policymakers and the policing community as a key approach to reducing police violence and disparate treatment. The need for more and better police training is one of the few areas where advocates and the policing community can agree on a potential solution. However, there is surprisingly little empirical evidence on the impact of formalized training on police behavior. With an estimated average of \$2.2 million spent by individual municipal police departments on training annually (\$3.6 million by state police departments), the dearth of empirical research on what works is alarming (Bureau of Justice Statistics, 2006). Not surprisingly, there is even less evidence that training interventions could potentially help to mitigate disparate treatment on the part of police towards minorities.

A number of criminal justice scholars have examined specific one-off police training interventions focused on narrow subjects like procedural justice, social interaction and communication, deescalation, mental health, and others.¹ Although many of those studies have found training to be effective at changing an officer's self-reported attitude, few have examined actual enforcement outcomes or explored their usefulness at mitigating disparities. Recent studies of force training have examined enforcement outcomes and found that these trainings had a large and persistent impact on police behavior (Lim and Lee (2015); Ba and Grogger (2018)). Basic training interventions in the developing world have also been shown to reduce

¹This literature has generally found that enrollment in procedural justice training has positive impacts on enforcement activity and outcomes (see Wheller et al. (2013); Graziano et al. (2014); Skogan et al. (2014); Thompson (2016); Nagin and Telep (2017); Nix et al. (2017); Mazerolle and Terrill (2018); Murphy and Mazerolle (2018); McLean et al. (2020a); McLean et al. (2020b); Rosenbaum and Lawrence (2017); Owens et al. (2017)). For example, find evidence of improved attitudes among treatment group officers toward procedurally fair communication in hypothetical encounters, though they find no evidence of behavioral changes in use of force. examine the effects of a one-day racial bias training on police stops and searches.

crime and improve overall police effectiveness (Banerjee et al. (2012); Garcia et al. (2013)). However, several recent RCTs conducted in the UK by criminal justice scholars have found that social interaction and racially biased search training did not affect the racial composition of traffic stops and searches (See McLean et al. (2020a) and Miller et al. (2020)). In our study, we focus on the Texas Highway Patrol (a very large U.S. policing agency) and examine a broad swath of legislatively-mandated trainings on cultural diversity, arrest protocol, search & seizure, racial profiling, traffic enforcement, and deescalation training.

Using data on 7.5 million police-citizen interactions made by 1,663 newly hired police officers from the 2010 to 2018 cohorts, we exploit quasi-random variation in the timing of when individual officers receive training and estimate a series of event study models. We use this framework to explore whether any of these trainings impacted related enforcement outcomes for a subsample of white non-Hispanic motorists as well as whether they affected the overall racial composition of traffic stops, searches, arrests, or citations. Across the board, we find that training has little to no effect on policing behavior both in terms of related enforcement outcomes and in terms of the racial composition of these outcomes. Given the substantial annual investment in training by the Texas Highway Patrol as well as other U.S. policing agencies, these findings are particularly important from a policy perspective. Further, there is no evidence supporting policy proposals aimed at combatting longstanding policing disparities through more police training. Given recent studies by West (2018a) and Horrace and Rohlins (2018) which find that search disparities decline with experience, our findings suggest that alternative interventions like apprenticeship models might be more effective relative to formalized training.

2.2 Data & Background

2.2.1 Texas Highway Patrol Data

In our study, we construct a unique linked dataset that allows us to track the enforcement impact of five legislatively mandated police trainings including (1) Cultural Diversity; (2) Arrest, Search, and Seizure; (3) Racial Profiling; (4) Traffic; and (5) Deescalation training. Through a series of public information requests, we have obtained data on over 25 million traffic stops, searches, and arrests made by approximately 6,500 Texas Department of Public Safety (DPS) peace officers from 2006 to 2019. This data represents the universe of traffic stops made by Texas Highway Patrol during this period. We have also obtained course-level training records from the Texas Commission on Law Enforcement (TCOLE) for individual

police officers as well as their administrative data on rank, educational attainment, and demographics. Any peace officer is required to take these particular trainings at least once during their time in “basic proficiency” which ranges from a period of 2 to 6 years of service depending on when they graduate from the academy relative to the ongoing training cycle. The courses range in length and are typically taken in person at a training facility managed by the Texas Commission on Law Enforcement or the Texas Law Enforcement Education Administration. We examine how each of these five trainings impacts the overall racial composition of traffic stops as well as racial differences in related enforcement outcomes (searches, arrests, and numbers of citations).

Key to our analysis, we link the traffic stop data to 800,000 individual course enrollments for the entire career of all officers in our data. We also link the individual officers in our traffic stop data to human resource records that describe their demographics as well as their job title and pay over time. These data allow us to assess whether training is an effective intervention for combating disparities in the application of stops, searches, arrests, and citations by police on minority citizens. We focus on a subset of 1,663 officers from the 2010 to 2018 academy cohorts who were responsible for approximately 7.5 million traffic stops since being hired.² We limit our focus to officers observed making traffic stops for at least one year, and who made at least one stop per day on average. From the traffic stop sample, we omit stops of commercial vehicles and stops which resulted in an arrest being the only violation since these were potentially motivated by a call for service or a warrant. Our sample for our main analysis consists of over 4.1 million traffic stops made by officers who received their first Cultural Diversity training within the sample period.³ The empirical strategy for our analysis relies on an event study design that examines enforcement outcomes immediately before and after a particular officer receives training relative to a control.

2.2.2 Officer Training

Our primary interest is the effect of mandatory Cultural Diversity training on the racial composition of traffic stops made by Texas Highway Patrol officers. Officers are required to take a minimum of eight credit hours of Cultural Diversity training during their “basic

²We focus our analysis on officers from the 2010 to 2018 cohorts due to changes in training requirements after 2009. This change can be observed in Table 1. Notice that in pre-2010 cohorts it was far more common for officers to have their first Cultural Diversity training outside the sample window and no Cultural Diversity training within the sample window.

³Note that the sample for the main analysis involving Cultural Diversity training is not the same as any sample used for analysis of the other trainings. This is because officers can receive different trainings at different times.

Table 1: Officer Cohort Descriptive Statistics

	Overall	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Number of Officers	2,274	312	166	133	136	76	128	181	196	211	393	273	69
<i>1st CD Outside Sample; >1st CD w/in</i>													
Total	107	26	14	12	17	2	6	8	5	14	3	3	0
First CD at Academy	0	0	0	0	0	0	0	0	0	0	0	0	0
First CD w/in 0-2 Years	19	3	2	0	1	0	0	0	0	10	3	3	0
First CD w/in 2-4 Years	30	0	0	1	13	0	1	6	5	4	0	0	0
First CD w/in 4-6 Years	32	14	9	0	0	2	5	2	0	0	0	0	0
<i>1st CD Outside Sample; No CD w/in</i>													
Total	804	276	148	119	30	16	23	21	20	45	38	57	11
First CD at Academy	355	173	108	54	0	0	1	2	0	7	4	5	1
First CD w/in 0-2 Years	385	98	40	65	14	3	8	10	13	38	34	52	10
First CD w/in 2-4 Years	12	1	0	0	1	2	2	0	6	0	0	0	0
First CD w/in 4-6 Years	33	2	0	0	1	8	12	9	1	0	0	0	0
<i>1st CD w/in Sample</i>													
Total	696	3	4	2	77	53	91	139	123	93	92	19	0
First CD at Academy	0	0	0	0	0	0	0	0	0	0	0	0	0
First CD w/in 0-2 Years	132	0	0	0	1	1	4	0	6	29	72	19	0
First CD w/in 2-4 Years	294	0	0	0	6	5	8	77	114	64	20	0	0
First CD w/in 4-6 Years	192	0	0	0	11	37	79	62	3	0	0	0	0
<i>No CD Yet</i>													
Total	667	7	7	7	12	5	8	13	48	59	260	197	58
First CD at Academy	0	0	0	0	0	0	0	0	0	0	0	0	0
First CD w/in 0-2 Years	0	0	0	0	0	0	0	0	0	0	0	0	0
First CD w/in 2-4 Years	0	0	0	0	0	0	0	0	0	0	0	0	0
First CD w/in 4-6 Years	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 1 presents descriptive statistics for officer cohorts. Officers are separated by when they took their first Cultural Diversity (CD) training. For officers who have had at least one Cultural Diversity training, results are further disaggregated by whether they received this training at the academy or years after their Basic Peace Officer training.

proficiency” period. The training is divided into two four-hour blocks. The first block consists of two modules that all officers are required to take: “Introduction to Diversity” and “Cultural Diversity”. For the second block, officers must choose another four-hour course from the remaining set of topic modules in the Cultural Diversity curriculum which include: Generational Diversity; Workplace Diversity; Gender Diversity; and Law Enforcement as a Diverse Culture.⁴ These trainings are designed to be “hands on, interactive, and scenario based,” with scenarios oriented to day-to-day experiences both on and off the job (Texas Commission on Law Enforcement, 2008). In addition to the standard Cultural Diversity training, we also include closely related trainings such as “Cultural Awareness.”

We also examine the effects of four other mandatory trainings on related outcomes and the racial composition of stops. First, we consider the effect of Arrest, Search, and Seizure training for which officers are required to have a minimum of 15 credit hours. This course is designed to teach officers the grounds for reasonable suspicion, probable cause, detention, search, seizure, and arrest. Second, officers receive Racial Profiling training in several different courses, only some of which are mandatory. For instance, TCOLE course number 3256,

⁴Of these secondary modules, “Workplace Diversity” and “Law Enforcement as a Diverse Culture” are most closely related to our outcome of interest. For example, “Law Enforcement as a Diverse Culture” discusses police-civilian interactions including racial issues.

Table 2: Stop-Level Descriptive Statistics

	Overall	Black Motorists	Hispanic Motorists	White Motorists
Observations	4,079,131	432,260	1,312,186	2,115,871
<i>Motorist Characteristics</i>				
White Motorist	0.519 (0.50)	0.000 (0.00)	0.000 (0.00)	1.000 (0.00)
Black Motorist	0.106 (0.31)	1.000 (0.00)	0.000 (0.00)	0.000 (0.00)
Hispanic Motorist	0.322 (0.47)	0.000 (0.00)	1.000 (0.00)	0.000 (0.00)
Male Motorist	0.638 (0.48)	0.638 (0.48)	0.698 (0.46)	0.645 (0.48)
In-State Vehicle	0.878 (0.33)	0.857 (0.35)	0.904 (0.29)	0.887 (0.32)
Car	0.445 (0.50)	0.669 (0.47)	0.441 (0.50)	0.393 (0.49)
Red/Yellow/Orange Vehicle	0.090 (0.29)	0.078 (0.27)	0.104 (0.31)	0.085 (0.28)
<i>Officer Characteristics</i>				
Black Officer	0.066 (0.25)	0.130 (0.34)	0.045 (0.21)	0.064 (0.25)
Hispanic Officer	0.313 (0.46)	0.167 (0.37)	0.506 (0.50)	0.227 (0.42)
White Officer	0.598 (0.49)	0.666 (0.47)	0.431 (0.50)	0.687 (0.46)
Other Race Officer	0.022 (0.15)	0.038 (0.19)	0.019 (0.14)	0.021 (0.14)
Male Officer	0.950 (0.22)	0.965 (0.18)	0.928 (0.26)	0.958 (0.20)
Female Officer	0.050 (0.22)	0.035 (0.18)	0.072 (0.26)	0.042 (0.20)
CD Credits	1.010 (0.16)	1.000 (0.13)	1.009 (0.12)	1.013 (0.18)
Arrest Credits	1.194 (0.17)	1.170 (0.11)	1.239 (0.26)	1.178 (0.12)
Use of Force Credits	1.044 (0.15)	1.063 (0.19)	1.043 (0.12)	1.041 (0.15)
Traffic Credits	1.051 (1.48)	1.044 (1.46)	0.875 (1.35)	1.130 (1.53)
Deescalation Credits	1.122 (0.03)	1.120 (0.03)	1.124 (0.04)	1.121 (0.02)
<i>Traffic Stop Outcomes</i>				
Number of Violations	1.450 (0.77)	1.584 (0.90)	1.557 (0.83)	1.365 (0.69)
Number of Citations	0.424 (0.65)	0.516 (0.75)	0.523 (0.72)	0.345 (0.58)
Number of Warnings	1.026 (0.68)	1.068 (0.72)	1.034 (0.70)	1.020 (0.66)
Arrest	0.018 (0.13)	0.040 (0.20)	0.019 (0.13)	0.015 (0.12)
Search	0.025 (0.16)	0.051 (0.22)	0.030 (0.17)	0.017 (0.13)

Table 2 presents descriptive statistics for the set of traffic stops used in our analysis broken down by motorist race. In the officer characteristics panel, we present the mean number of credit hours obtained in each training scaled by the required minimum to advance. Standard deviations in parentheses.

“Racial Profiling,” is optional, but officers are required to have a minimum of two credit hours in TCOLE course number 3257, “Combined Asset Forfeiture and Racial Profiling.” Third, officers are required to take a substantial amount of Traffic training. TCOLE trainings are unclear on the minimum number of credit hours officers must spend in the various Traffic trainings, but we observe that the average officer in our sample has taken 24 hours. Finally, officers are required to have a minimum of eight credit hours of Deescalation training that is designed to limit the use of force in public interaction, particularly in cases where

individuals are in crisis or behaving erratically (Texas Commission on Law Enforcement, 2021).

2.3 Methods & Results

We divide our models and results into three different sections. In the first section, we examine the effect of the main training of interest, Cultural Diversity, on racial disparities. We begin by looking at the effect of Cultural Diversity training on the racial composition of traffic stops, specifically the share of blacks and Hispanics stopped. We then consider whether Cultural Diversity training affects racial differences in two post-stop outcomes: the probability of an arrest or search and the number of citations issued. In the second section, we examine whether any of the other trainings affect either the racial composition of stops or racial differences in post-stop outcomes. Finally, we examine whether training has any effect on post-stop outcomes using a sample of white motorists only.

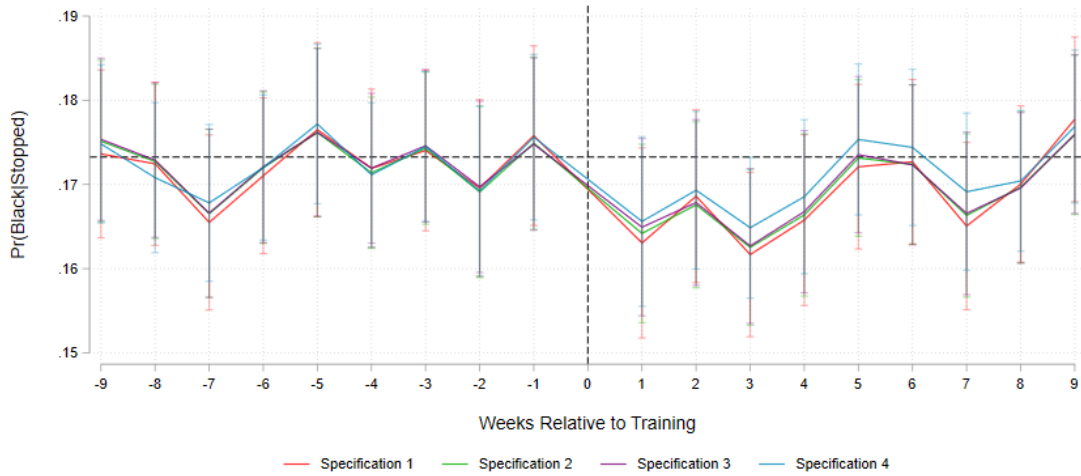
2.3.1 Does Cultural Diversity Training Affect Racial Disparities?

For the estimates that examine the racial composition of enforcement outcomes for black motorists, we limit our sample to both black and white non-Hispanic motorists and estimate a linear probability model of the form

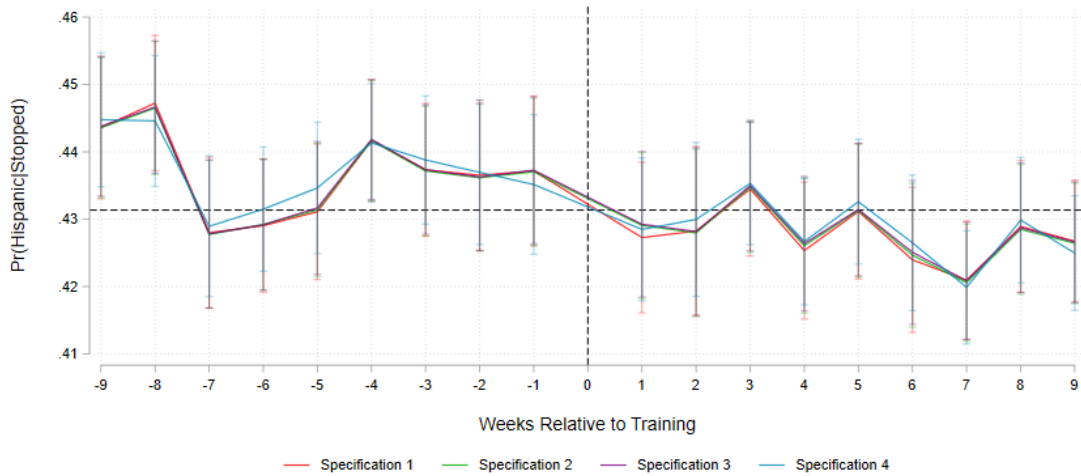
$$1[\text{black}_{i,o}] = \beta_0 + \sum_{t=1}^T \beta_t 1[\text{run}_{i,o,t}] * \text{credits}_{t,o} + \text{credits}_{t,o} + \gamma_t + h_i + l_i + \lambda_{t,o} + \eta_{i,t}, \quad (1)$$

where the dependent variable $1[\text{black}_{i,o}]$ indicates that the motorist stopped by officer o in incident i was black. The primary explanatory variables $\sum_{t=1}^T 1[\text{run}_{i,o,t}]$ are a series of event time dummies indicating that the stop was made nine weeks before the first time an officer receives training in a given area. The event time dummies are interacted with the variable $\text{credits}_{t,o}$, which indicates the credit hours an officer has received in Cultural Diversity training at time t .⁵ We include all stops made by the officer earlier than nine weeks before their first training as the omitted category but drop those later than nine weeks after the training. All specifications also include controls for date γ_t , hour of day h_i , county by route l_i and the number of credits in other non-focal training courses $\lambda_{t,o}$.

⁵In the uncommon event that an officer receives two trainings during the event study window, we take a weighted average of training credits. Only two officers in our data have two Cultural Diversity trainings in the event study window.



(a) **Black**

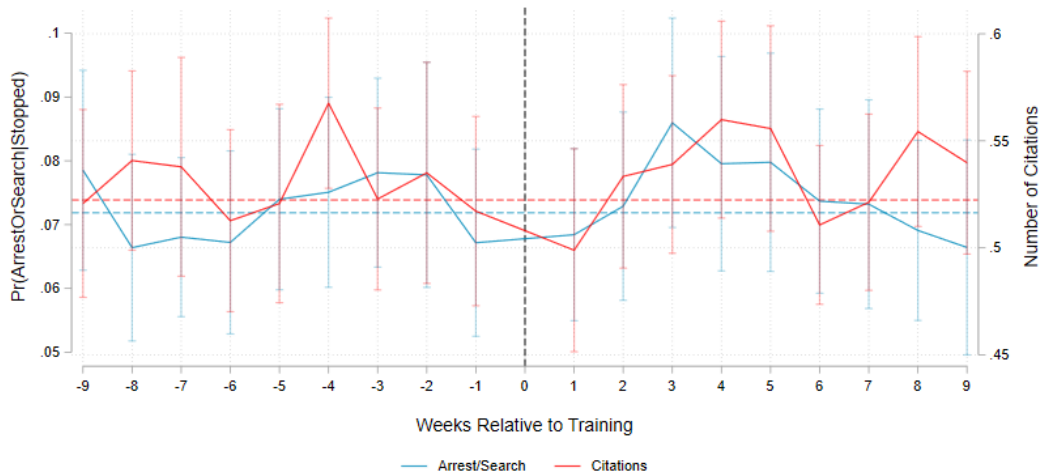


(b) **Hispanic**

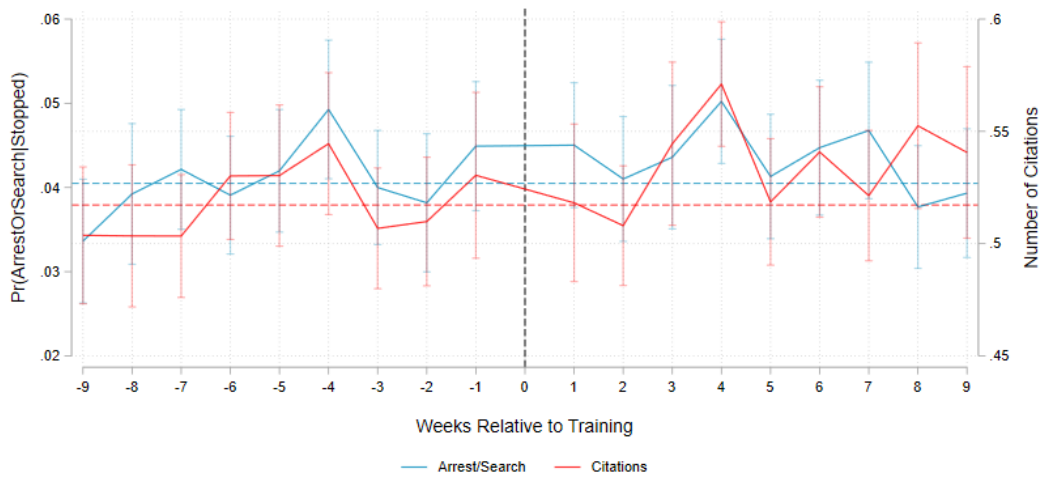
Figure 1: **Effect of CD Training on Racial Composition of Stops**

Additional more rigorous specifications also include controls for motorist characteristics, officer characteristics, and officer fixed effects. Estimation for Hispanic motorists is analogous to the model described by Equation 6.

Figure 1 presents point estimates and 95 percent confidence intervals from the application of Equation 6 to the legislatively mandated Cultural Diversity training. As discussed above, the baseline model includes controls for date, hour of day, county by route, and number of other training courses. The baseline model is represented in Figure 1 as Specification 1 (red), which estimates the effect of the training on the black (Hispanic) share of stopped motorists



(a) **Black**



(b) **Hispanic**

Figure 2: **Effect of CD Training on Racial Differences in Post-Stop Outcomes**

before and after the training. Specification 2 (green) introduces controls for motorist characteristics (gender and vehicle characteristics) and Specification 3 (purple) also includes officer characteristics (race and gender). Finally, Specification 4 (blue) replaces the officer characteristics with officer fixed effects. Although the likelihood that a stopped motorist is black appears to drop by a small amount after the training, the effect is statistically insignificant and the standard errors are consistently around about half a percentage point. Similarly, we find no evidence of an effect of Cultural Diversity training on the likelihood that a stopped motorist is Hispanic. We conclude that the Cultural Diversity training appears to have little

if any impact on the overall racial composition of traffic stops.⁶ Because results are similar across specifications we choose to use Specification 4 as our main model for subsequent results.

Next, we examine the effect of Cultural Diversity training on racial differences in post-stop outcomes. Again, we limit the sample to both black (Hispanic) and white non-Hispanic motorists and estimate the following linear probability model:

$$\begin{aligned}
 1[\textit{outcome}_{i,o}] = & \beta_0 + \sum_{t=1}^T \alpha_t 1[\textit{run}_{i,o,t}] * \textit{credits}_{t,o} + 1[\textit{black}_{i,o}] * \textit{credits}_{t,o} \\
 & + 1[\textit{black}_{i,o}] + \sum_{t=1}^T \beta_t 1[\textit{run}_{i,o,t}] * 1[\textit{black}_{i,o}] * \textit{credits}_{t,o} + \gamma_t + h_i + l_i + \lambda_{t,o} + \eta_{i,t},
 \end{aligned} \tag{2}$$

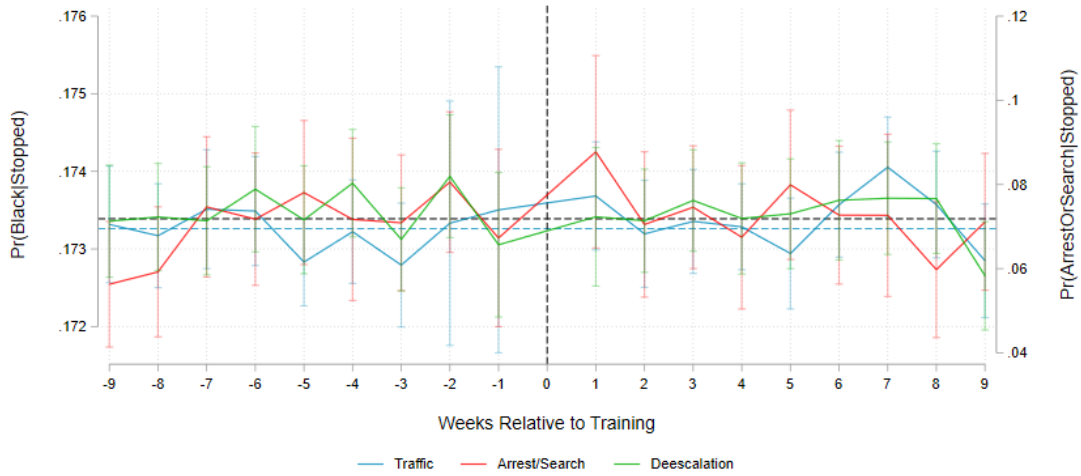
where the dependent variable $1[\textit{outcome}_{i,o}]$ indicates that the motorist stopped by officer o in incident i experienced one of the two post-stop outcomes we consider. Namely, we explore whether training has any effect on the probability of arrest/search or number of citations. Figure 2 presents results for the effect of Cultural Diversity training on these other outcomes. We find no compelling evidence that Cultural Diversity training has any impact on racial differences in post-stop outcomes for either black or Hispanic motorists. Taken together with our previous results for the racial composition of stops, these results suggest that the effectiveness of Cultural Diversity training in mitigating racial disparities is limited. Again, estimation for Hispanic motorists is analogous to the model described by Equation (2).

2.3.2 Do Other Trainings Affect Racial Disparities?

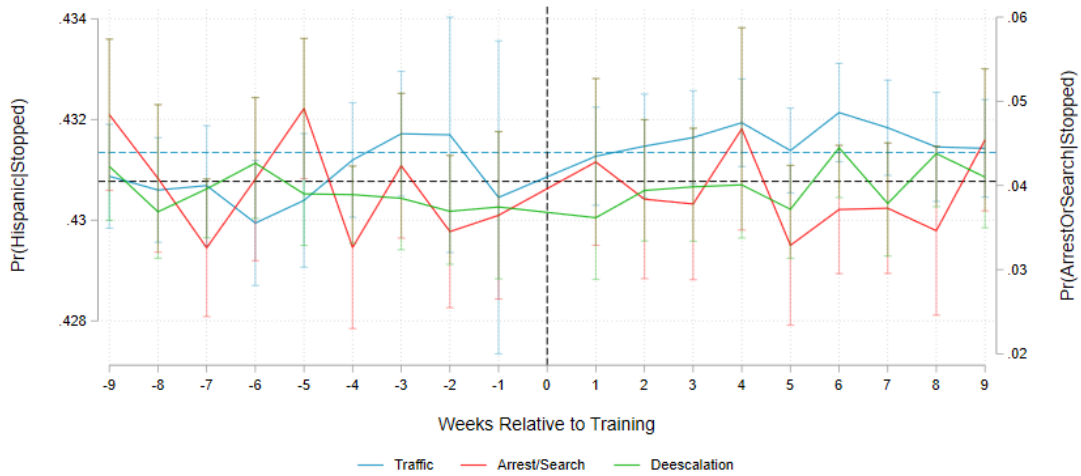
Finding no effect for Cultural Diversity training, we consider whether other trainings affect relevant racial outcomes. In particular, we begin by estimating the effect of Traffic training on the racial composition of stops using Equation (1). These results are presented in Figure 3 where the effect of the Traffic training is represented in blue and the outcome is measured by the left axis. As with Cultural Diversity, there is no indication of any effect of Traffic training on the racial composition of stops for black or Hispanic motorists.

Although neither Cultural Diversity nor Traffic training appear to effect the racial composition of stops, one might reasonably expect that trainings do affect related post-stop outcomes. For instance, it is plausible that Arrest/Search training could affect the share of

⁶For robustness we run these same event studies using white officers only and find that the results are similar to those in Figure 1.



(a) **Black**

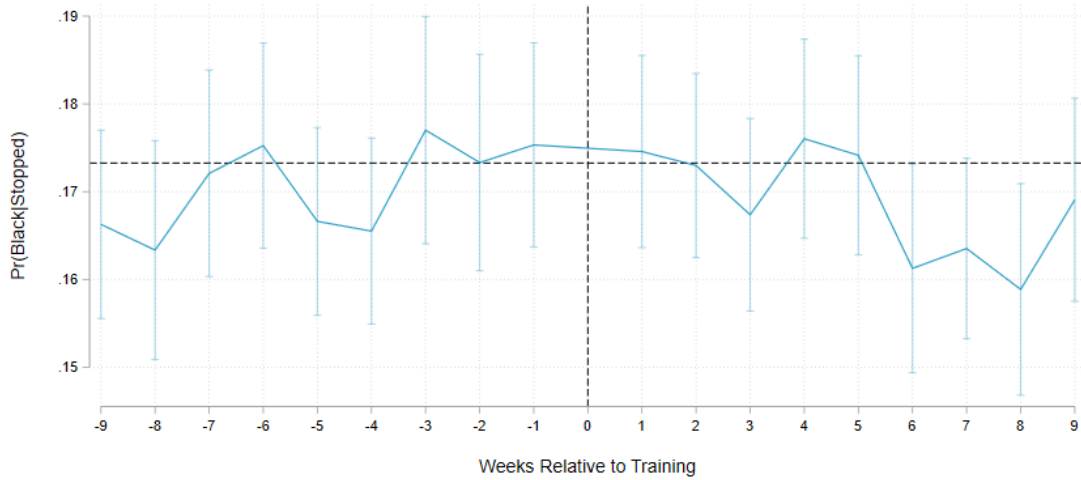


(b) **Hispanic**

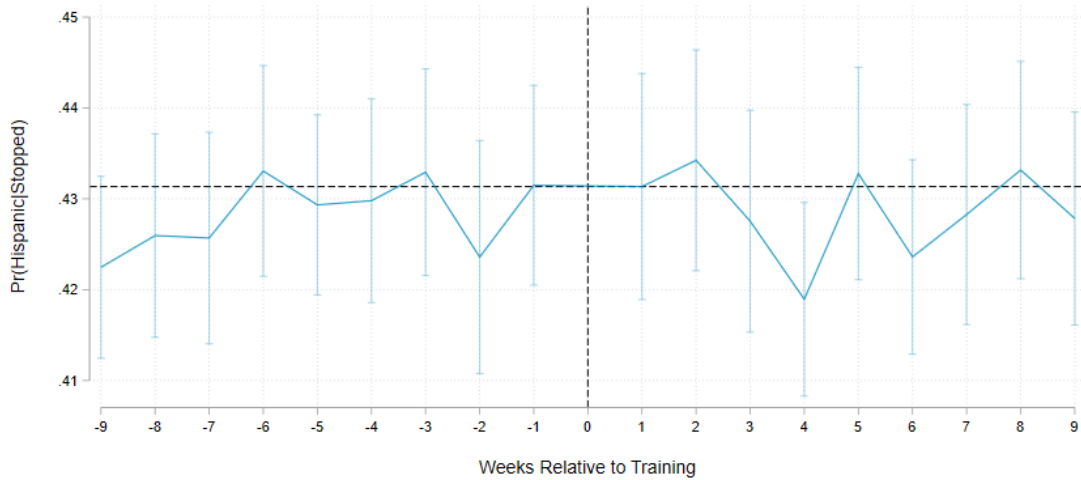
Figure 3: **Effect of Other Trainings on Racial Composition of Stops**

black (Hispanic) motorists arrested or searched if not the overall black (Hispanic) share of stopped motorists. We examine this issue by applying Equation (2) to the Arrest/Search and Deescalation trainings. These results are also presented in Figure 3, where Arrest/Search training effects are represented with red and Deescalation training effects are represented in green. The arrest/search outcome is measured by the right axis. We find no discernible impact of either training on racial disparities in the arrest/search outcome.

Finally, we estimate the effect of Racial Profiling training on the racial composition of traffic stops using Equation (1). These results are presented in Figure 4. As with Cultural



(a) **Black**



(b) **Hispanic**

Figure 4: **Effect of RP Training on Racial Composition of Stops**

Diversity and Traffic, we find no effect on the racial composition of stops.⁷

2.3.3 Does Training Affect Outcomes for White Motorists?

Having found no evidence of training effects on racial outcomes, we consider whether these trainings have any effect at all on relevant outcomes. We do so by restricting our sample to

⁷In addition to the trainings described above, we also examine the effect of Use of Force training on these outcomes. Officers are required to take a minimum of 13 credit hours of Use of Force training. We find no effect on racial outcomes.



Figure 5: **Effect of Other Trainings for White Motorists**

white motorists only and estimating a linear probability model of the form

$$1[Arrest/Search]_{i,o} = \beta_0 + \sum_{t=1}^T \beta_t 1[run]_{i,o,t} * credits_{t,o} + credits_{t,o} + \gamma_t + h_i + l_i + \lambda_{t,o} + \eta_{i,t}, \quad (3)$$

where $1[Arrest/Search]_{i,o}$ is an indicator for whether an arrest or search occurred during the stop. Results for these models are presented in Figure 5 for Arrest/Search and Traffic trainings. The trainings do not appear to have any effect on the arrest/search outcome for white motorists.

2.4 Parametric Models

Finally, we estimate a series of parametric models to obtain more precise estimates of the null effects from the preceding models. We estimate three categories of models. The first is the Event Study model given by Equation (4):

$$1[black]_{i,o} = \beta_0 + \beta_1 1[Treatment]_{i,o,t} * credits_{t,o} + credits_{t,o} + \gamma_t + h_i + l_i + \lambda_{t,o} + \eta_{i,t}, \quad (4)$$

which eliminates the $1[run]_{i,o,t}$ term from Equation (1) and ignores the possibility of a pre-trend or trend. This model introduces the term $1[Treatment]_{i,o,t}$ which is equal to zero

Table 3: Results for Parametric Models

Training:	Cultural Diversity			Traffic	Arrest/Search	Deescalation
Outcome:	Stops	Arrest/Search	Citations	Stops	Arrest/Search	Arrest/Search
Black Motorists						
<i>Model:</i>						
Event Study	-0.00272 (0.00202)	0.00213 (0.00319)	0.0110 (0.00829)	0.000717 (0.000959)	0.00118 (0.00361)	0.000317 (0.00238)
Pre-Trend	-0.00221 (0.00305)	0.000757 (0.00518)	0.000179 (0.0149)	0.00189 (0.00155)	0.000302 (0.00534)	-0.00281 (0.00479)
Trend	-0.00492 (0.00391)	0.00272 (0.00671)	-0.00874 (0.0200)	0.00234 (0.00194)	0.00421 (0.00724)	-0.00191 (0.00614)
Observations	2,548,021	2,547,815	2,547,815	2,134,119	1,571,523	2,185,490
Hispanic Motorists						
<i>Model:</i>						
Event Study	-0.00392 (0.00264)	0.00252 (0.00161)	0.0177* (0.00800)	0.00183 (0.00185)	-0.00144 (0.00197)	-0.000157 (0.00127)
Pre-Trend	-0.0104** (0.00327)	0.000371 (0.00246)	0.0105 (0.0104)	0.00655* (0.00315)	0.00132 (0.00304)	0.00226 (0.00226)
Trend	-0.0101* (0.00424)	0.00124 (0.00314)	0.00175 (0.0136)	0.00621 (0.00332)	0.00249 (0.00408)	0.00153 (0.00291)
Observations	3,427,956	3,427,771	3,427,771	2,702,872	1,962,276	3,269,010

prior to the training week and 1 after an officer receives training. Second, we estimate the following Pre-Trend model:

$$\begin{aligned}
 1[black_{i,o}] = & \beta_0 + \beta_1 1[Treatment_{i,o,t}] * credits_{t,o} + credits_{t,o} + \\
 & + \sum_{t=1}^T \beta_2 1[Pretrend_{i,o,t}] * credits_{t,o} + \gamma_t + h_i + l_i + \lambda_{t,o} + \eta_{i,t},
 \end{aligned} \tag{5}$$

where the term $1[Pretrend_{i,o,t}]$ is equal to zero prior to the training week window, increments by one for each of the nine weeks prior to the training week, and remains at nine in the following weeks. Third, we estimate the following Trend model:

$$\begin{aligned}
 1[black_{i,o}] = & \beta_0 + \beta_1 1[Treatment_{i,o,t}] * credits_{t,o} + credits_{t,o} + \\
 & + \sum_{t=1}^T \beta_2 1[Trend_{i,o,t}] * credits_{t,o} + \gamma_t + h_i + l_i + \lambda_{t,o} + \eta_{i,t},
 \end{aligned} \tag{6}$$

where $1[Trend_{i,o,t}]$ is equal to zero prior to the training week window, increments by one for each of the 18 weeks of the training window (nine weeks before training and nine weeks

after), and remains at 18 for the following weeks.⁸

Results for these models are presented in Table 3. As expected, the treatment effects are more precisely estimated and still nearly all null. However, there is suggestive evidence of a pre-trend in the model estimating the effect of Cultural Diversity training on stops of Hispanic motorists.

2.5 Discussion and Conclusion

In this paper, we use a sample of 7.5 million police-citizen interactions made by 1,663 Texas Highway Patrol officers to explore the effects of five mandatory police trainings on a variety of enforcement outcomes. We exploit quasi-random variation in the timing of when individual officers receive training and estimate a number of event study models. We begin by focusing our attention primarily on the eight credit hour Cultural Diversity training and its effect on the racial composition of traffic stops and other racial outcomes involving arrest and search and the number of citations issued. Finding no evidence that Cultural Diversity training affects racial outcomes, we ask whether other trainings (namely, Arrest/Search, Racial Profiling, Traffic, and Deescalation) affect racial outcomes. Again, we find no evidence that training affects either the black (Hispanic) share of stopped motorists or racial differences in the probability of arrest/search or the number of citations. Further, we find no evidence of an effect of training on these training-related outcomes for a subsample of white motorists.

Several possible explanations could explain our null results. Each possibility has different implications for policy. First, it may be the case that officers are simply not receiving enough training to impact their behavior. For each of the trainings considered here (except traffic), officers are required to take between 8 and 13 credit hours. In practice, this may be between one and two days worth of training. If an officer is racially biased, the length and intensity of training may be insufficient to overcome biases developed during the course of his lifetime. It is possible that more training is required to have a meaningful impact on officer behavior.

Importantly, we cannot determine whether officers are racially biased here. Observed racial disparities in policing outcomes do not necessarily imply racial bias if motorists in our sample vary by race in their propensity to violate traffic or criminal law. Though we cannot test this hypothesis directly, it has important policy implications. If officers are already enforcing the law impartially and continue to do so after receiving training, we should not

⁸The three equations described above are for models of the racial composition of traffic stops, analogous to Equation (1). We similarly extend Equation (2) for the parametric models of other outcomes with race on the right-hand side, but do not present the equations here in the interest of space.

expect to observe any impact on racial outcomes.

A third possibility, consistent with the previous two, is that officers do not learn primarily by training and instead learn by experience or mentoring. West (2018a) and Horrace and Rohlins (2018) observe that search disparities decline with experience. In light of our null results for police training, policymakers seeking to mitigate racial disparities might consider alternatives such as apprenticeship models.

A final possibility that must be addressed is that the specific training models used by the Texas Commission on Law Enforcement are not effective. In this case, more effective trainings could possibly be devised. Given the substantial amount of investment and heightened public interest in police training, our findings suggest that policymakers should exercise caution in expanding training as a means of mitigating racially disparate treatment.

3 Chapter 3: Estimating Effects of Affirmative Action in Policing

Coauthored with Maryah Garner and Anna Harvey

3.1 Introduction

In the aftermath of Michael Brown's death in Ferguson, Missouri, renewed attention was directed to the demographic composition of law enforcement agencies. In the summer of 2015, 2 out of 3 residents of Ferguson, but only three out of its 53 police officers, were black (U.S. Department of Justice, 2016). This disparity gave rise to questions about whether increasing the proportion of black officers on the Ferguson police force, or in law enforcement agencies more generally, would lead to different public safety outcomes. On the one hand, increasing the proportion of nonwhite police officers may lead to an increased acceptance of policing, fewer instances of the use of force, and fewer citizen complaints in nonwhite communities (Tyler, 2005), leading in turn to increases in the reporting of crime by nonwhite victims, and in the willingness of civilians to cooperate with police investigations into these crimes (Miller and Segal, 2018). Nonwhite officers may also exert greater effort, or be better equipped, to clear crimes in largely nonwhite neighborhoods. These mechanisms might lead to decreases in nonwhite crime victimization. On the other hand, increasing the proportion of nonwhite police officers may lead to decreases in policing quality if nonwhite applicants to police agencies are less qualified than white applicants, and/or if the effort of white officers is reduced due to morale effects (Lott, 2000). These mechanisms might lead to increases in nonwhite and/or white crime victimization. Detecting the presence of these potentially offsetting and racially heterogeneous effects is a challenging exercise.

Existing work on this question has generally not attempted to identify the specific causal pathways through which agency racial composition may affect policing outcomes, seeking instead to estimate average effects of agency racial composition on outcomes such as reported offense and arrest rates [Lott, 2000, McCrary, 2007]. Yet analyzing the effect of the racial composition of law enforcement agencies on these outcomes is complicated by the endogeneity of hiring and retention practices to other agency-specific factors that may also affect outcomes. To achieve identification, researchers have sought to leverage the incidence and timing of affirmative action litigation [Lott, 2000, McCrary, 2007]. This work has yielded inconsistent results regarding the effects of racial diversity on policing outcomes. Existing work has not, however, implemented recent econometric advances in difference-in-differences

estimation.

In this paper, we replicate and extend the work of Lott (2000) and McCrary (2007) using modern difference-in-differences methods, including the difference-in-differences decomposition method developed by Goodman-Bacon [2019] as well as the generalized synthetic controls method developed by Xu [2017]. Ultimately, like McCrary (2007), we do not find a significant average effect of court-imposed affirmative action plans on the rates of reported offenses or reported offenses cleared by arrest. We also extend the effort of McCrary (2007) to analyze whether unlitigated agencies change their practices due to the threat of litigation, raising concerns of stable unit treatment value assumption (SUTVA) violations. We estimate whether the total number of agencies receiving court-imposed affirmative action plans at time t has a differential impact on the rates of reported offenses or offenses cleared by arrest for agencies that will never be litigated, relative to agencies that will eventually be litigated. Like McCrary (2007), we find little evidence of spillover effects. Finally, we return to the question of identifying the specific and possibly racially heterogeneous causal mechanisms linking affirmative action litigation and public safety outcomes, suggesting that this would be a productive avenue for future research.

3.2 Background

Police departments have experienced some of the most aggressive affirmative action programs ever implemented in the United States [McCrary, 2007, Miller and Segal, 2012]. Beginning in the late 1960s with a number of employment discrimination lawsuits, federal courts began mandating affirmative action plans with the intended effect of increasing the shares of non-white and female police officers. Court-imposed affirmative action plans often take the form of hiring quotas, but also may affect standards for promotion. Some police departments are still under affirmative action plans today, often from court-imposed plans going back to the 1970s. The justification for such affirmative action plans may be to rectify past discrimination, and/or to promote the compelling government interest in increasing the effectiveness of police departments at detecting and interdicting crime. There are several reasons one might expect more racially diverse police departments to be more effective at policing. It has long been recognized that minority groups tend to be suspicious of police and the criminal justice system more generally (U.S. Kerner Commission, 1968).¹ Further,

¹As early as 1968, the U.S. Kerner Commission’s Report of the National Advisory Commission on Civil Disorders expressed this sentiment as follows: “To some Negroes police have come to symbolize white power, white racism and white repression. And the fact is that many police do reflect and express these white

as Donohue and Levitt (2001) point out, “conflicts between police and citizens have been the flashpoint for virtually every recent urban riot.” This statement remains true in more recent years (for instance, the 2014-2015 Ferguson riots and the 2015 Baltimore riots). Lack of trust may lead nonwhite civilians to be less likely to report crimes to police, to cooperate with investigations, and to take instructions from law enforcement, in the presence of a largely white police force. Increasing the proportion of nonwhite officers in racially diverse communities may lead to nonwhite victims becoming more likely to report crimes, and to members of these communities becoming more likely to cooperate with the police to help solve crimes reported by nonwhite victims.

Nonwhite police officers may also exert greater effort to detect and clear crimes occurring in nonwhite neighborhoods. Nonwhite officers may also better understand the cultural norms of predominantly nonwhite communities. One might also expect police officers to be less likely to racially discriminate against members of their own race. Nonwhite police officers may also perceive largely nonwhite neighborhoods to be less hostile, relative to white police officers.² These mechanisms may contribute to increased reporting by nonwhite crime victims, and to decreased incidence of nonwhite crime victimization.

On the other hand, there are potential adverse effects of more racially diverse police departments. In order to implement court-imposed affirmative action plans, some police departments have had to change entrance standards. Police force entrance requirements vary by location, but generally contain several basic components, including criminal background checks and physical examinations. Prior to the wave of employment discrimination lawsuits in the 1970s, it was standard to use entrance examinations that tested the cognitive abilities of prospective officers. These exams tested aptitudes pertaining to reading comprehension, verbal reasoning, analogies, and, in some cases, I.Q. Black applicants historically performed worse than white applicants on these entrance exams (McCrary, 2007). Federal courts have intervened in the examination process since the 1970s, often mandating that entrance standards be changed if they have a disparate impact on a particular group and the standards are not shown to relate to job performance. To deal with pressures from the federal judiciary, some police departments have removed cognitive tests altogether in order to increase nonwhite recruitment. Others have simply reduced standards (Lott, 2000). The

attitudes. The atmosphere of hostility and cynicism is reinforced by a widespread belief among Negroes in the existence of police brutality and in a 'double standard' of justice and protection—one for Negroes and one for whites.”

²Groves and Rossi (1970) suggest that the perceptions of hostility toward police are projections of the fears and prejudices of white police officers themselves.

relevant question to ask is whether these aptitude tests reliably screen applicants for qualities important to police work. If affirmative action litigation leads to the lowering of entrance standards, and those standards are reliable indicators of future job performance, we would expect affirmative action to lead to worse policing outcomes.

Many employment discrimination lawsuits were brought by private litigants beginning in the late 1960s. The U.S. Department of Justice (DOJ) became involved in these discrimination lawsuits after 1972. If a court found that a police department entrance exam disproportionately affected black applicants and was not reliably related to job performance, the court would order that the department devise a new test that either did not disparately affect black applicants or was job-related. There would typically be a one- to three-year lag before a hiring quota was imposed (McCrary, 2007). In some cases, hiring quotas would end once a goal had been reached (for instance, a police department might be required to hire a certain percentage of black officers each year until the agency reached a target proportion of black officers in its force). In other cases, hiring quotas lasted until terminated by the judiciary.

Several papers have considered the employment effects of affirmative action litigation. Lott (2000) and McCrary (2007) both find significant increases in black police employment following affirmative action litigation, with McCrary finding a post-litigation 14 percentage point gain in the fraction of black officers among newly hired officers. Miller and Segal (2012) also find persistent and significant employment effects for black police officers as a consequence of litigation. Their finding holds even for departments for which affirmative action is eventually terminated. Miller and Segal (2012) also find a significant divergence in black employment between agencies that continued affirmative action and those that ended it, with larger persistent gains in black employment in the former agencies. Additionally, Miller and Segal (2012) find that there is an important distinction between simply experiencing litigation, and actually having a court-imposed affirmative action plan. Departments that are litigated but not required to implement affirmative action see increases in black employment, but at a lower rate than those with court-imposed affirmative action plans.

In addition to looking at the impacts of affirmative action on employment, several papers discuss how increased diversity affects policing outcomes. Donohue and Levitt (2001) examine how the racial composition of police departments affects racial patterns of arrest, using the racial composition of fire departments as an instrument for the racial composition of police departments. They find that increases in the percentage of black police officers lead to higher arrest rates for whites but not for nonwhites. Similarly, increases in the percentage

of white police officers lead to higher arrest rates for nonwhites, but not for whites. These patterns particularly hold for minor offenses.

Lott (2000) and McCrary (2007) consider how affirmative action litigation affects reported crime and arrest rates. Using logged annual rates of per capita reported violent and property crime between 1987 and 1994 for a sample of 495 cities, Lott (2000) estimates the impacts of 19 consent decrees signed by the Department of Justice and a city’s policing agency and still in force by 1987. In both reduced form and instrumental variable models, Lott finds that consent decrees lead to substantial increases in reported crime rates, and weakly lead to decreases in arrest rates. He interprets these effects as being the result of lower hiring standards.

However, McCrary (2007) suggests that this mechanism is implausible for several reasons. First, consent decrees typically led to hiring practices in which applicants were evaluated relative to other applicants of the same race. Under these hiring practices, no reduction of standards was required for nonblack applicants. Second, while hiring standards may have been reduced in some cases to eliminate disparate impact, in many cases entrance exams were modified to be more closely related to job performance. Third, looking at test score distributions for the New York City Police Department, McCrary (2007) finds that hiring quotas impacted test scores of new hires “only minimally.” Further, the consent decrees used in Lott’s sample include neither affirmative action plans resulting from private litigation nor externally-imposed affirmative action plans that were terminated prior to 1987.

McCrary (2007) estimates event study models of the effects of affirmative action litigation on reported offense and arrest rates. Using a sample of 314 large municipal police departments, he finds little evidence that affirmative action litigation impacted reported city-level crime and arrest rates. He suggests that there may be “a complex series of effects that offset one another.” However, McCrary (2007) does not distinguish between two types of litigated cities: those that implemented court-imposed affirmative action plans, and those that did not. In addition, difference-in-differences methods developed after the publication of McCrary (2007), including the difference-in-differences decomposition method developed by Goodman-Bacon (2019), as well as generalized synthetic controls developed by Xu (2017), may lead to different insights.

Miller and Segal (2018) use affirmative action litigation, data from the National Crime Victimization Survey, and FBI data on intimate partner homicides to examine how increasing female representation among police officers impacts the incidence of domestic violence and intimate partner homicide, and the reporting of domestic violence. They find that violent

crimes against women are reported at higher rates when female representation increases in law enforcement agencies. Further, greater female representation in policing agencies leads to significant declines in the rates of intimate partner homicide and non-fatal domestic abuse. In both instrumental variables and reduced form models, they find similar effects from affirmative action litigation. They find no effects of affirmative action litigation or increased female officer shares on the reporting or incidence of crimes committed against male victims.

Our paper primarily focuses on replicating and extending the estimates of the impacts of affirmative action litigation on crime reported by Lott (2000) and McCrary (2007). However, in the discussion section we return to the question of identifying the causal pathways by which race-based affirmative action may affect policing outcomes, including possibly racially heterogeneous treatment effects.

3.3 Data

3.3.1 Affirmative Action Litigation Data

Our data on affirmative action litigation are sourced from Miller and Segal (2012), who constructed the most complete currently available legal database of affirmative action litigation involving police departments. Previous data sets, such as the one used by McCrary (2007), did not distinguish between unsuccessful litigation and successful litigation leading to court-imposed affirmative action plans. Additionally, the data set constructed by Miller and Segal (2012) looks at affirmative action litigation specifically addressing the employment of police officers, rather than the employment of all police department employees (which includes clerical and janitorial positions).

Miller and Segal constructed their litigation data by first looking at employment data from confidential EEO-4 reports filed with the Equal Employment Opportunity Commission (EEOC) between 1973 and 2005. They examined reports from 479 of the largest state and local law enforcement agencies in the United States. Miller and Segal then searched for legal records pertaining to discrimination in employment for each agency using the LexisNexis and Westlaw federal case databases. They gathered information on the actual litigation, including whether affirmative action was implemented and when it ended (if applicable), and the protected group. This information was cross-referenced with data from the U.S. Department of Justice (DOJ) and the databases used in McCrary (2007) and Lott (2000).

Miller and Segal (2012) find that of the 479 agencies examined, 140 affirmative action

cases were brought either by private plaintiffs or by the DOJ between 1969 and 2000. Of the 140 agencies which experienced litigation, 117 saw the implementation of affirmative action plans, and 23 saw litigation that did not result in court-imposed affirmative action. Among the 117 agencies that experienced court-imposed affirmative action, 67 of these agencies saw the eventual termination of the program. The mean duration of the 67 plans that terminated was 16 years. Miller and Segal (2012) also report that 96% of the affirmative action plans for which the protected group can be determined involve the employment of black officers.

3.3.2 UCR and CPS Data

We match the agencies in Miller and Segal’s affirmative action database to the agencies reporting annual crime and arrest data in the FBI’s Uniform Crime Reporting (UCR) program.³ Agencies that report zero crimes or arrests in any year are treated as missing data in that year; all agencies missing annual data are dropped from the sample. We identify the county within which each agency is located, and match counties to annual metropolitan statistical area (MSA) population and demographic data sourced from the U.S. Census Bureau’s Current Population Survey (CPS).⁴

We are left with a sample of 99 agencies in the Miller and Segal affirmative action database that are located within an MSA and that consistently report annual crime and arrest data in the UCR between 1964 and 2011. Of these 99 agencies, 27 agencies implement court-imposed affirmative action programs at some point. Among these 27 agencies, 12 agencies have programs that terminate during our sample period, and 15 have programs that do not. The mean time treated is approximately 25.7 years (20.4 years for agencies whose affirmative plans end; 30.8 years for agencies whose affirmative action plans do not end). Figure 1 reports a histogram of the years in which affirmative action plans were imposed. Figure 2 shows where treated and untreated agencies are located.

Following Lott (2000) and McCrary (2007), we focus on two sets of outcome variables: the natural logarithms of rates of reported violent and property crime offenses per 100,000 in population, and arrest rates, defined as the numbers of violent and property crime offenses cleared by arrest, divided by the numbers of violent and property crime index offenses, between 1964 and 2011. Violent crime is the aggregation of four index crimes in the UCR

³UCR data, as cleaned and compiled by Kaplan (2019), were downloaded from ICPSR. The UCR program is voluntary; not all agencies participate, or participate consistently, in the program.

⁴This data was compiled by Flood et al. (2018). The CPS data are not publicly available on an MSA level prior to 1977. We assume that the demographic data for 1964-1976 resembles the 1977 CPS data. CPS data are also not available for the MSAs in our data from 1990 to 1999, so we use the closest previous year of data for the missing year observations.



Figure 1: **Timing of Court-Imposed AA Plans**

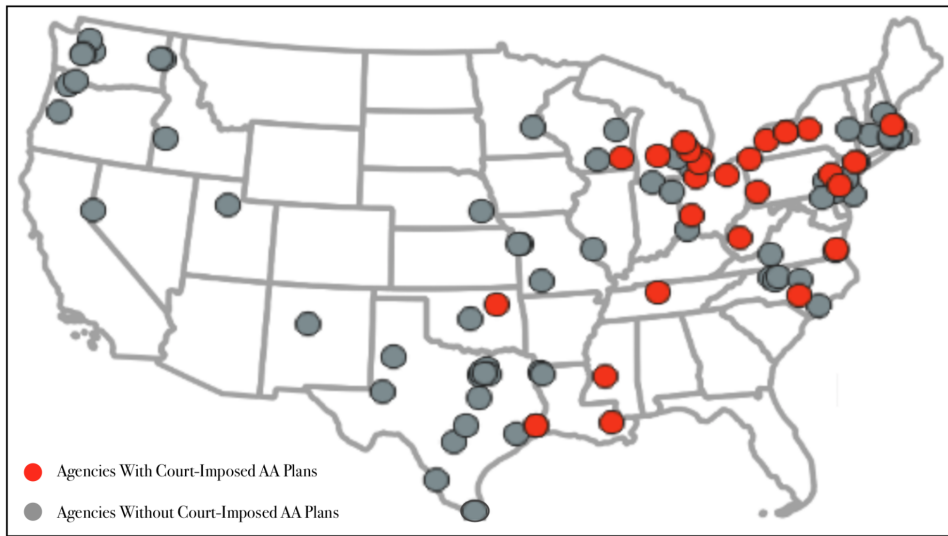


Figure 2: **Location of Treated and Untreated Agencies**

data: murder, rape, robbery, and aggravated assault. Property crime is the aggregation of burglary, theft, and motor vehicle theft. Summary statistics for the departments in our sample are presented in Table 1.

We use the MSA-level demographic data from the CPS only to match departments in the

Table 1: **Summary Statistics, Offense, and Arrest Rates**

	All Agencies	Treated Years	Untreated Years
ln(Violent Offenses Per 100,000)	6.13 (1.00)	6.70 (0.72)	6.03 (1.00)
Violent Crime Arrest Rate	0.51 (0.34)	0.44 (0.15)	0.52 (0.36)
ln(Property Offenses Per 100,000)	8.48 (0.57)	8.66 (0.48)	8.45 (0.58)
Property Crime Arrest Rate	0.18 (0.08)	0.18 (0.10)	0.18 (0.08)
Number of Observations	4,752	722	4,030

Table 1 reports the means for each annual per capita offense/arrest rate variable from 1964 to 2011. *Treated Years* reports means for agencies that had court-imposed AA plans during that year. *Untreated Years* reports means for agencies that did not have court-imposed AA plans during that year. Standard deviations in parentheses.

generalized synthetic control analyses. Summary statistics for the demographic variables for all, treated, and untreated agencies are presented in Table 2.

3.4 Models and Results

3.4.1 Two-way Fixed Effect Difference-in-Differences Models

The canonical difference-in-differences model compares pre-post changes in outcomes in treated units to pre-post changes in outcomes in untreated units for a single treatment [Goodman-Bacon, 2019]. In our data, 72 agencies are untreated units not subject to externally imposed affirmative action and 27 agencies are treated. However, treatment timing varies across treated units.

The accepted empirical strategy in this context is the two-way fixed effect difference-in-differences model (2WFE DD), as in Equation (1):

$$OffenseRates/ArrestRates_{it} = \beta_0 + \beta_1 Treat_{it} + \theta_t + \alpha_i + \epsilon_{it}, \quad (1)$$

where $Treat$ is a binary variable equal to one when a unit is subject to a court-imposed affirmative action plan and equal to zero otherwise; θ_t is a time vector containing indicators for the 48 years from 1964 to 2011; and α_i is a unit vector containing indicators for the 99 agencies. Standard errors are clustered by agency. The average treatment effect on the treated (ATT) is given by β_1 . We estimate the ATT of externally-imposed affirmative action on logged violent and property crime rates per capita and on violent and property crime

Table 2: **Summary Statistics of Demographic Variables**

	All Agencies	Treated	Untreated
Proportion White	0.84 (0.09)	0.82 (0.11)	0.85 (0.08)
Proportion Black	0.12 (0.08)	0.15 (0.10)	0.12 (0.08)
Proportion Other	0.03 (0.03)	0.03 (0.03)	0.03 (0.03)
Proportion Female	0.52 (0.02)	0.53 (0.02)	0.52 (0.02)
Proportion Not in Labor Force	0.27 (0.04)	0.28 (0.04)	0.27 (0.04)
Share of Population Age 18-21	0.05 (0.01)	0.05 (0.01)	0.05 (0.01)
Share of Population Age 21-24	0.06 (0.02)	0.06 (0.02)	0.06 (0.02)
Percent Below Poverty Line	0.11 (0.04)	0.11 (0.04)	0.11 (0.04)
City Population	210,723 (236,833)	324,626 (335,717)	190,317 (207,920)
Number of Observations	4,752	722	4,030
Number of Agencies	99	27	72

Table 2 reports the means for each demographic variable between 1964 and 2011. The *Treated* group consists of agencies which had court-imposed AA plans at least for some part of our sample period. The *Untreated* group consists of agencies that did not have court-imposed AA plans during the sample period. The *Untreated* group includes agencies that faced unsuccessful litigation as well as agencies that never faced litigation.

arrest rates between 1964 and 2011.

The 2WFE DD model specified above captures average treatment effects on the treated, but does not allow us to consider time-varying treatment effects. There are several reasons to expect the effects of affirmative action plans to vary over time. First, court-imposed affirmative action plans often take time to implement. Further, once implementation begins it takes time for the racial composition of police departments to change significantly, due to the nature of hiring quotas. To account for potentially time-varying treatment effects, we implement difference-in-differences decomposition [Goodman-Bacon, 2019].

3.4.2 Difference-in-Differences Decomposition

The 2WFE DD estimate is composed of a weighted average of treatment effects estimated from a series of 2x2 treatment/control groups, some of which compare agencies treated at the same time to untreated agencies, and some of which compare agencies treated at the same time to agencies treated at another time (earlier or later).⁵ As reported in Figure 1, there are 12 timing groups in our data, or groups of agencies which experience the imposition of externally-imposed affirmative action in the same year. There are thus 144 distinct 2x2 treatment/control comparison groups from which the 2WFE DD estimate is constructed: 132 groups in which earlier-treated agencies are compared to later-treated agencies, or vice versa, and 12 groups in which treated agencies are compared to untreated agencies. In the presence of time-varying treatment effects, comparisons between earlier and later treated units may introduce bias into the 2WFE DD estimate. The extent of the bias depends on the share of the 2WFE DD estimate that is derived from these earlier-later comparisons, which in turn depends on group size and the variance of the treatment [Goodman-Bacon, 2019].

Goodman-Bacon [2019] has developed a method to decompose the 2WFE DD estimate into the 2x2 weighted estimates from which it is derived. Using this difference-in-differences decomposition model, we can uncover the extent to which the 2WFE DD estimate depends on 2x2 DD estimates which compare earlier to later treated agencies. The Goodman-Bacon decomposition model is currently only available for strongly balanced panels in which treatment only changes from 0 to 1 over time. To estimate the decomposition model, we define treatment as a binary variable that is equal to one in all years after an affirmative action plan is imposed on an agency, and is equal to zero otherwise.⁶ We report both the 2WFE DD estimate for this model, as well as the DD estimates and weights for the three categories of treatment/control comparison groups from which the 2WFE DD estimate is derived.

3.4.3 Results for DD Models

Results for our difference-in-differences models for offense and arrest rates are presented in Table 3. The DD Model reports the 2WFE DD estimate where treatment is defined as years in which an agency is subject to an externally-imposed affirmative action plan. The GB

⁵Comparisons may also be made to always-treated units; however, we have no always-treated agencies in our data.

⁶We also estimated a version of the decomposition model that includes an indicator for years after the termination of an affirmative action plan; results were qualitatively similar to those reported here.

Table 3: **Treatment Effects for DD Models: Offense and Arrest Rates**

	Ln(Violent Crime PC)	Ln(Property Crime PC)	Violent Crime Arrest Rate	Property Crime Arrest Rate
DD Model	0.088 (0.067)	0.035 (0.046)	0.004 (0.026)	0.013 (0.015)
GB Model	0.02 (0.09)	0.02 (0.05)	-0.001 (0.027)	0.011 (0.016)
<i>Avg. DD Decomp. Estimates</i>				
Earlier T vs. Later C (Weight = 0.03)	0.01	-0.05	0.01	-0.01
Later T vs. Earlier C (Weight = 0.08)	0.16	0.01	0.05	0.05
Treated vs. Never Treated (Weight = 0.89)	0.01	0.03	-0.01	0.01
Number of Observations	4,752	4,752	4,752	4,752
Number of Agencies	99	99	99	99
Year Fixed Effects	Yes	Yes	Yes	Yes
Agency Fixed Effects	Yes	Yes	Yes	Yes

Table 3 summarizes average treatment effects for two different DD models of four different dependent variables. Each cell in the table represents the treatment effect for one model. Standard errors are clustered on agency.

Model reports the 2WFE DD estimate where treatment is defined as all years subsequent to the imposition of an affirmative action plan. For the GB Model, we also report average DD decomposition estimates and weights for 2x2 treatment/control groups.⁷ Each model is estimated separately using four different dependent variables: the natural log of per capita violent crime offenses, the natural log of per capita property crime offenses, violent crime arrest rates, and property crime arrest rates. Neither model yields any statistically significant results, for any outcome variable. These results support the findings of McCrary (2007) and contrast with the findings of Lott (2000).

The Goodman-Bacon decomposition estimates also allow us to see that there is little evidence that bias introduced by time-varying treatment effects is driving the null 2WFE DD estimates. The latter are largely driven (total weight = 89%) by comparisons between treated and untreated agencies, as in the canonical 2x2 DD model. Although there is some evidence of time-varying treatment effects, with 2x2 DD estimates signed in the opposite direction for some timing groups, relative to other treatment/control groups, little weight is placed on the timing group 2x2 DD estimates in the construction of the 2WFE DD estimate.

3.4.4 Duration Models

We also estimate models of duration or dosage effects of externally imposed affirmative action plans. For these models we replace the variable *Treat* with the variable *Years of Treatment*

⁷For the 2WFE models we use *xtreg* in Stata 16; for the decomposition model we use the *bacondecomp* Stata 16 command developed by Goodman-Bacon et al. (2019).

Table 4: **Duration Models: Offense Rates**

	Duration Model 1		Duration Model 2	
	Ln(Violent Crime PC)	Ln(Property Crime PC)	Ln(Violent Crime PC)	Ln(Property Crime PC)
Years of Treatment	0.005 (0.003)	0.003 (0.003)	0.006 (0.008)	0.005 (0.007)
Number of Observations	4,752	4,752	4,752	4,752
Number of Agencies	99	99	99	99
Year Fixed Effects	Yes	Yes	Yes	Yes
Agency Fixed Effects	Yes	Yes	Yes	Yes
Agency Time Trends	Yes	Yes	Yes	Yes

Table 4 summarizes results for the duration models described by Equations (2) and (3). Standard errors are clustered on agency.

(YOT_{it}) in our 2WFE DD model. In our first duration model we control for agency and year fixed effects:

$$OffenseRates/ArrestRates_{it} = \beta_0 + \beta_1 YOT_{it} + \theta_t + \alpha_i + \epsilon_{it}, \quad (2)$$

where β_1 is the effect of an agency being subjected to one more year of affirmative action. In our second duration model, we add the vector λ_{it} , which captures agency-specific linear time trends in offense and arrest rates. This model takes the form:

$$OffenseRates/ArrestRates_{it} = \beta_0 + \beta_1 YOT_{it} + \theta_t + \alpha_i + \lambda_{it}\epsilon_{it} \quad (3)$$

3.4.5 Results for Duration Models

Table 4 reports estimates from Equations (2) and (3) for logged offense rates. We find no effects of years of affirmative action exposure on logged offense rates, in any model. Table 5 reports estimates for Equations (2) and (3) for violent and property crime arrest rates. We again find no statistically significant effects of externally imposed affirmative action on either violent crime or property crime arrest rates. These estimates are generally consistent with our other DD estimates, as well as with the results reported by McCrary (2007).

Although we are able to control for an extensive amount of exogenous variation using agency and year fixed effects, there may still be selection effects confounding our estimates. The parallel trends assumption, on which DD models depend, states that offense and arrest

Table 5: **Duration Models: Arrest Rates**

	Duration Model 1		Duration Model 2	
	Violent Crime Arrest Rate	Property Crime Arrest Rate	Violent Crime Arrest Rate	Property Crime Arrest Rate
Years of Treatment	0.0003 (0.0015)	0.0002 (0.0007)	0.0002 (0.0019)	-0.0002 (0.0007)
Number of Observations	4,752	4,752	4,752	4,752
Number of Agencies	99	99	99	99
Year Fixed Effects	Yes	Yes	Yes	Yes
Agency Fixed Effects	Yes	Yes	Yes	Yes
Agency Time Trends	Yes	Yes	Yes	Yes

Table 5 summarizes results for the duration models described by Equations (2) and (3). Standard errors are clustered on agency.

rates within treated agencies should have changed at the same rate as offense and arrest rates within untreated agencies, had treatment not occurred. Yet if there was nonrandom selection into treatment, and treated agencies differ from untreated agencies, the parallel trends assumption may be violated. Since the 1990s, synthetic control methods have been used to construct more appropriate comparison units for treated units that differ from untreated units (Abadie et al., 2010; Card, 1990). We likewise use the generalized synthetic control method [Xu, 2017] to construct more appropriate control units for treated agencies.

3.4.6 Generalized Synthetic Control

The generalized synthetic control (GSC) method introduced by Xu [2017] addresses the case when treatment is imposed at different times for different units. This approach allows for multiple treated units and variable treatment periods.⁸ The GSC method allows us not only to match units on pretreatment observables but also to model unobserved time-varying heterogeneities using interactive fixed effects.

GSC first estimates an interactive fixed effects (IFE) model using only the police departments that were never treated and obtains a fixed number of time-varying coefficients (latent factors). It then estimates department-specific intercepts (factor loadings) for each treated police department by linearly projecting pretreatment outcomes for treated units onto the space spanned by the factors. Finally, it generates synthetic control units based on the estimated factors and factor loadings. The method is described as a “bias correction

⁸This method also has several other advantages. It includes a built-in cross-validation procedure and is easier to implement than other synthetic control methods.

procedure for IFE models when treatment data is heterogeneous across units.”⁹

The *gsynth* package requires at least seven years of pretreatment data for each treated agency, and two of our treated agencies did not meet this requirement. We are left with 25 treated agencies for the GSC models. The pretreatment covariates used in constructing the synthetic controls are the following: proportion black, share of population age 18-21, share of population age 21-24, proportion in the labor force, and percent below the poverty line.

3.4.7 Results for Generalized Synthetic Control Models

GSC models allow us to estimate treatment effects as they evolve over time, without imposing linearity. We display our results as graphs of the average treatment on the treated (ATT) over time. Each ATT is calculated by taking the difference between the treated unit and the synthetic control for a given unit, and then averaging the difference across all treated units. We do this each year over a 27 year period. Results are displayed in Figure 3. Each figure shows seven years of pretreatment effects and 20 years of posttreatment effects. The 90% confidence interval is in gray. Confidence intervals are constructed using parametric bootstrapping.¹⁰

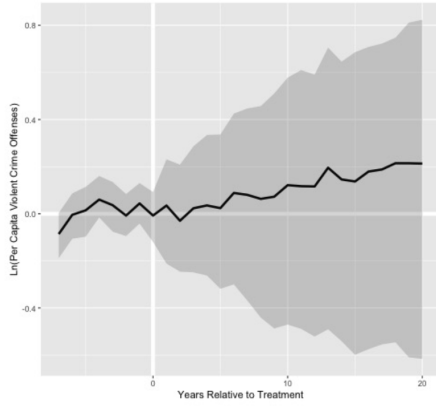
Figure 3 reports the estimates for the natural logs of violent and property crime per 100,000 in population and the estimates for violent and property crime arrest rates. From these estimates, we cannot infer that court-imposed affirmative action plans led to changes in either offense or arrest rates. Overall, these results corroborate the findings from our DD models, which are generally in agreement with McCrary (2007).

3.5 SUTVA Violation Test

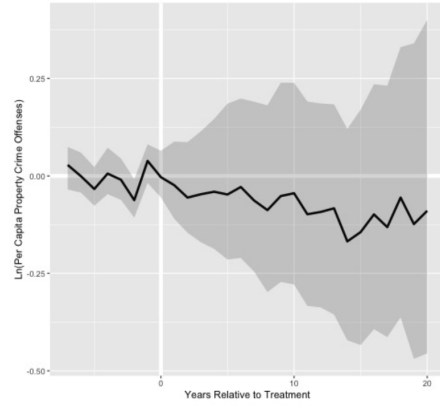
A final concern is that the law enforcement reaction to the threat of potential affirmative action litigation may have violated the stable unit treatment value assumption (SUTVA). As police departments across the country began experiencing litigation for discriminatory hiring practices, other departments may have made preemptive changes to avoid litigation. Miller and Segal (2012) find that police departments that were unsuccessfully litigated for affirmative action experienced an increase in black employment prior to litigation. This suggests that departments whose leaders believed they were likely to experience litigation

⁹To estimate this model we use the *gsynth* package developed by Xu and Liu (2018).

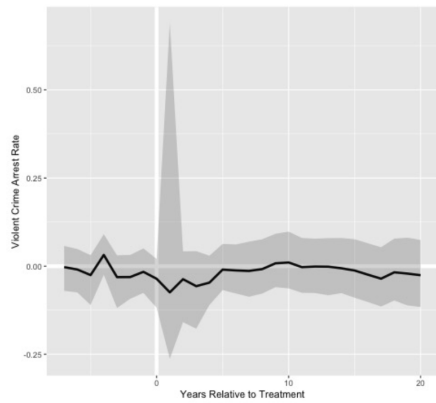
¹⁰We do not include tables showing the relative weights assigned to each agency included in the construction of each synthetic control, because these tables would be unwieldy. In most cases, the synthetic control group for a treated agency is comprised of a large number of agencies, each with a small assigned weight.



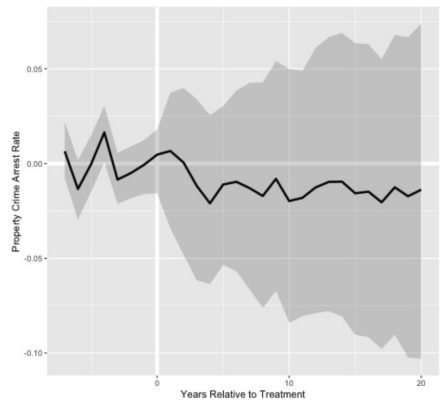
(a) ATT for ln(PC Violent Crime Offenses)



(b) ATT for ln(PC Property Crime Offenses)



(a) ATT for Violent Crime Arrest Rates

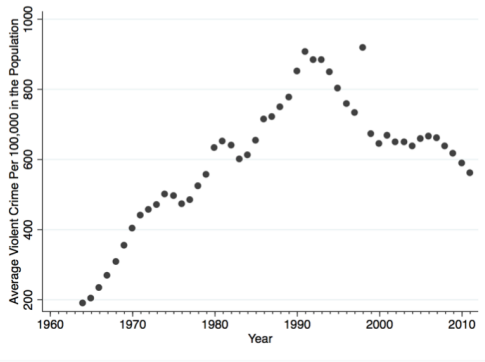


(b) ATT for Property Crime Arrest Rates

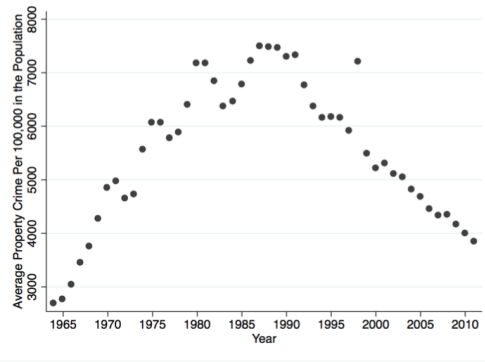
Figure 3: **Treatment Effects for GSC Models**

may have changed their behavior, implying that the treatment of affirmative action may have affected the untreated units as well as the treated. This would be a direct violation of SUTVA.

McCrary (2007) estimated whether agencies were adjusting their behavior due to the threat of litigation by matching unlitigated agencies with litigated agencies within the same federal district. He assumed the matched unlitigated agencies to be threatened agencies. He then assigned to the threatened agencies neighbor-litigation dates which were equivalent to the litigation dates of their counterparts. He identified unthreatened agencies as those agencies that had no litigated agencies within their federal district. Using an event study, he estimated the effect of neighbor-litigation dates on the minority employment gap for both threatened and unthreatened agencies. He found that there was not a significant



(a) Average Violent Offenses PC



(b) Average Property Offenses PC

Figure 4: **Offenses per 100,000 in Population**

difference in the hiring behavior of threatened and unthreatened agencies as a consequence of neighbor-litigation dates. However, McCrary (2007) looked only at litigation, not at successful litigation resulting in court-imposed affirmative action. Including unsuccessful litigation in these analyses would likely attenuate the results. Second, geographical proximity might not be a good matching mechanism.

To further explore the possibility of a SUTVA violation, we examine the effect of the total number of agencies receiving court-imposed affirmative action plans on offense and arrest rates both for agencies that will never be litigated and for agencies that will eventually be litigated.¹¹ In order to isolate the pretreatment effect, we drop litigated agencies from our sample once they have been litigated. Since the total number of agencies receiving court-imposed affirmative action plans at any point in time does not vary across agencies, we cannot use time fixed effects to control for the trends in offense and arrest rates over time. Looking at Figure 4, we can see that there are obvious non-linear time trends in both violent and property crime rates. To control for these time trends, we implement polynomial transformations of our year variable, with the optimal number of degrees determined by minimizing the Bayesian Information Criterion (BIC) for each model. We also control for agency-specific linear time trends in crime and arrest rates.

Our model takes the following form:

¹¹We assume that agencies that will one day face litigation are aware that they have a higher probability of litigation, and thus are more likely to adjust their behavior.

Table 6: **Treatment Effects on the Untreated**

	Ln(Violent Crime PC)	Ln(Property Crime PC)	Violent Crime Arrest Rate	Property Crime Arrest Rate
AA Total	-0.0004 (0.0029)	-0.0011 (0.0015)	-0.0005 (0.0006)	0.0009 (0.0005)
AA Total * Ever Litigated	-0.0120 (0.0067)	0.0000 (0.0025)	-0.0002 (0.0021)	-0.0009 (0.0011)
$\beta_1 + \beta_2$	-0.0124 (0.0059)	-0.0020 (0.0021)	-0.0006 (0.0020)	0.0000 (0.0009)
Number of Agencies	99	99	99	99
Agency Fixed Effects	Yes	Yes	Yes	Yes
Agency Time Trends	Yes	Yes	Yes	Yes
Time Polynomial Degree	3	3	1	3

Table 6 summarizes results for our models of potential SUTVA violations described by Equation (4). Standard errors are clustered on agency.

$$\begin{aligned}
 \text{OffenseRates}/\text{ArrestRates}_{it} = & \beta_0 + \beta_1 AATotal_t + \beta_2 (AATotal * EverLitigated)_{it} + \\
 & + \sum_{p=1}^P \xi_p t^p + \alpha_i + \lambda_i t + \epsilon_{it}, \quad (4)
 \end{aligned}$$

where $AAtotal_t$ is the total number of police departments with court-imposed affirmative action plans at time t , calculated from the data reported by Miller and Segal (2012); $EverLitigated$ is a binary variable that is equal to one (in all time periods) for agencies that will eventually experience affirmative action litigation; and P is the optimal number of degrees of the polynomial transformation of the year variable t , as reported in Table 6. β_1 is the effect of increasing the total number of court-imposed affirmative action plans on agencies that will never be litigated, while $\beta_1 + \beta_2$ estimates the effect for agencies that will eventually be litigated, and β_2 is the difference in the effect.

Table 6 reports these estimates. We do not find evidence of a possible SUTVA violation. The number of agencies subjected to court-imposed affirmative action plans at any point in time does not appear to be associated with significant changes in offense or arrest rates among unlitigated agencies, either in absolute terms or relative to litigated agencies.

3.6 Discussion

In this paper, we replicated and extended the work of Lott (2000) and McCrary (2007) using modern difference-in-differences methods. These two studies sought to estimate the average effects of race-based affirmative action litigation on outcomes such as reported offense and arrest rates. In our replications and extensions, which also focused on offense and arrest rates, we found results generally consistent with the null results reported by McCrary (2007). Also like McCrary (2007), we do not find evidence that agencies changed their behavior in anticipation of affirmative action litigation.

Yet it is not clear that analyzing only the average effects of court-imposed race-based affirmative action is the most productive empirical strategy. These effects may be heterogeneous by victim race. As noted previously, race-based affirmative action leading to increased proportions of nonwhite police officers may lead to increases in reporting by nonwhite crime victims, and in the willingness of civilians to cooperate with police investigations into crimes committed against nonwhite victims, while having little effect on crimes experienced by white victims. Nonwhite officers may also exert greater effort, or be better equipped socially and culturally, to clear crimes in largely nonwhite communities. Race-based affirmative action may then lead to decreases in the incidence of crimes experienced by nonwhite victims, and/or increases in the reporting of crimes experienced by nonwhite victims while having little effect on the incidence and reporting of crimes experienced by white victims.

These potentially racially heterogeneous treatment effects may also offset each other. For example, race-based affirmative action leading to increases in the share of nonwhite officers may both increase the reporting of offenses experienced by nonwhite victims, and decrease the number of offenses experienced by these victims, with a net null effect on reported crimes experienced by nonwhite victims. Likewise, race-based affirmative action leading to increases in the number of nonwhite officers could both increase the effort devoted to clearing offenses experienced by nonwhite victims, but simultaneously decrease the number of these offenses through deterrence effects, again leading to a null effect on offenses cleared by arrest for nonwhite victims.

In order to identify and disambiguate these potentially racially heterogeneous and offsetting effects, researchers will need to look for data beyond the widely used UCR data. For example, using the National Crime Victimization Survey, which allows for the identification of the race and gender of crime victims, and for the measurement of crimes both unreported and reported to law enforcement, Miller and Segal (2018) are able to identify gender-specific causal effects of affirmative action on both the actual incidence of violent crimes and the

reporting of those crimes. Likewise, the identification of the potentially racially heterogeneous and offsetting effects linking race-based affirmative action to public safety outcomes is a productive avenue for future research.

3.7 Conclusion

Affirmative action was aggressively implemented in police departments beginning in the 1970s. This implementation usually took the form of court-imposed hiring quotas. We examined how court-imposed race-based affirmative action has impacted both offense and arrest rates, seeking to replicate and extend the results of Lott (2000) and McCrary (2007). Our estimates from difference-in-differences, DD decomposition, duration, and generalized synthetic control models generally support the null results of McCrary (2007).

We then analyzed potential spillover effects, to examine whether untreated agencies were changing their behavior as they observed other agencies receiving court-imposed affirmative action plans. Like McCrary (2007), we did not find evidence supporting the claim of spillover effects of treatment on the untreated.

Finally, we concluded with a discussion of the importance of identifying the possibly racially heterogeneous and offsetting causal effects of race-based affirmative action on public safety outcomes.

References

- Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.
- Antonovics, K. and Knight, B. G. (2009). A new look at racial profiling: Evidence from the Boston Police Department. *The Review of Economics and Statistics*, 91(1):163–177.
- Anwar, S. and Fang, H. (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review*, 96(1):127–151.
- Ba, B. and Grogger, J. (2018). The introduction of tasers and police use of force: Evidence from the Chicago Police Department. *NBER Working Paper No. 24202*.
- Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D., and Singh, N. (2012). Can institutions be reformed from within? Evidence from a randomized experiment with the Rajasthan police. *NBER Working Paper No. 17912*.
- Bureau of Justice Statistics (2006). State and local law enforcement training academies, 2006. <https://www.bjs.gov/content/pub/pdf/slleta06.pdf>.
- Card, D. (1990). The impact of the Mariel boatlift on the Miami labor market. *Industrial and Labor Relations Review*, 43(2):245–257.
- Dallas Public Safety Committee (2016). Communications Operations Center (Handling Calls for Service).
- Dharmapala, D. and Ross, S. L. (2004). Racial bias in motor vehicle searches: Additional theory and evidence. *B.E. Journal of Economic Analysis & Policy*, 3(12).
- Donohue, J. J. and Levitt, S. (2001). The impact of race on policing and arrests. *The Journal of Law and Economics*, 44(2):367–394.
- Flood, S., King, M., Rodgers, R., Ruggles, S., and Warren, J. (2018). *Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset]*.
- Fryer, R. G. (2018). An empirical analysis of racial differences in police use of force. *NBER working paper No. 22399*.

- Gallup (2016). Public opinion context: Americans, race and police.
- Garcia, J. F., Mejia, D., and Ortega, D. (2013). Police reform, training and crime: Experimental evidence from Columbia’s Plan Cuadrantes. *CAF Working Paper No. 2013-01*.
- Goel, S., Rao, J. M., and Shroff, R. (2016). Precinct of prejudice? Understanding racial disparities in New York City’s stop-and-frisk policy. *The Annals of Applied Statistics*, 10(1):365–394.
- Goodman-Bacon, A. (2019). Difference-in-differences with variation in treatment timing. Working paper.
- Goodman-Bacon, A., Goldring, T., and Nichols, A. (2019). *bacondecomp*: Stata module for decomposing difference-in-differences estimation with variation in treatment timing.
- Graziano, L. M., Rosenbaum, D. P., and Schuck, A. M. (2014). Building group capacity for problem solving and police-community partnerships through survey feedback and training: A randomized control trial within Chicago’s community policing program. *Journal of Experimental Criminology*, 10(1):79–103.
- Grogger, J. and Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association*, 101(475):878–887.
- Groves, E. and Rossi, P. (1970). Police perceptions of a hostile ghetto. *American Behavioral Scientist*, 13(5):727–744.
- Hoekstra, M. and Sloan, C. (2020). Does race matter for police use of force? Evidence from 911 calls. *NBER Working Paper No. 26774*.
- Horrace, W. C. and Rohlin, S. M. (2016). How dark is dark? Bright lights, big city, racial profiling. *The Review of Economics and Statistics*, 98(2):226–232.
- Horrace, W. C. and Rohlins, S. (2018). Police officer learning, mentoring and bias in traffic stops. *Syracuse University Working Paper*.
- Kaplan, J. (2019). Uniform Crime Reporting program data: Offenses known and clearances by arrest, 1960-2017.
- Lim, H. and Lee, H. (2015). The effects of supervisor education and training on police use of force. *Criminal Justice Studies*, 28(4):444–463.

- Lott, J. (2000). Does a helping hand put others at risk?: Affirmative action, police departments, and crime. *Economic Inquiry*, 38(2):239–277.
- Mazerolle, L. and Terrill, W. (2018). Can you build a better cop? Making every police-citizen interaction count. *Criminology & Public Policy*, 17(1):89–96.
- McCrary, J. (2007). The effect of court-ordered hiring quotas on the composition and quality of police. *American Economic Review*, 97(1):318–353.
- McLean, K., Wolfe, S. E., Rojek, J., Alpert, G. P., and Smith, M. R. (2020a). Randomized controlled trial of social interaction police training. *Criminology & Public Policy*, 19(3):805–832.
- McLean, K., Wolfe, S. E., Rojek, J., Alpert, G. P., and Smith, M. R. (2020b). Social interaction training to reduce police use of force. *The ANNALS of the American Academy of Political and Social Science*, 687(1):124–145.
- Miller, A. R. and Segal, C. (2012). Does temporary affirmative action produce persistent effects? a study of black and female employment in law enforcement. *The Review of Economics and Statistics*, 94(4):1107–1125.
- Miller, A. R. and Segal, C. (2018). Do female officers improve law enforcement quality? effects on crime reporting and domestic violence. *The Review of Economic Studies*, 0:1–28.
- Miller, J., Quinton, P., Alexandrou, B., and Packham, D. (2020). Can police training reduce ethnic/racial disparities in stop and search? Evidence from a multisite UK trial. *Criminology & Public Policy*, 19(4):1259–1287.
- Murphy, K. and Mazerolle, L. (2018). Policing immigrants: Using a randomized control trial of procedural justice policing to promote trust and cooperation. *Australian & New Zealand Journal of Criminology*, 5(1):5–32.
- Nagin, D. and Telep, C. (2017). Procedural justice and legal compliance. *Annual Review of Law and Social Sciences*, 13:5–28.
- Nix, J., Pickett, J. T., Wolfe, S. E., and Campbell, B. A. (2017). Demeanor, race, and police perceptions of procedural justice: Evidence from two randomized experiments. *Justice Quarterly*, 34(7):1154–1183.

- Owens, E., Weisburd, D., Amendola, K. L., and Alpert, G. P. (2017). Can you build a better cop? Experimental evidence of supervision, training, and policing in the community. *Journal of Experimental Criminology*, 13:293–319.
- Pew Research Center (2016). The racial confidence gap in police performance.
- Pew Research Center (2020). A month before George Floyd’s death, black and white Americans differed sharply in confidence in the police. Online; accessed 2 February 2020.
- Rabe-Hemp, C. E. (2008). Female officers and the ethic of care: Does officer gender impact police behaviors? *Journal of Criminal Justice*, 36(5):426–434.
- Rosenbaum, D. and Lawrence, D. (2017). Teaching procedural justice and communication skills during police–community encounters: Results of a randomized control trial with police recruits. *Journal of Experimental Criminology*, 13:293–319.
- Skogan, W. G., Van Craen, M., and Hennessy, C. (2014). Training police for procedural justice. *Journal of Experimental Criminology*, 11.
- Texas Commission on Law Enforcement (2008). Cultural diversity: Course number 3939.
- Texas Commission on Law Enforcement (2021). Specific course reporting numbers. <https://www.tcole.texas.gov/content/specific-course-reporting-numbers>.
- Thompson, I. (2016). The impact of procedural justice training on first year constables’ interactions with citizens: A randomized controlled trial. *Master’s Thesis at Fitzwilliam College*.
- Tyler, T. (2005). Policing in black and white: Ethnic group differences in trust and confidence in the police. *Police Quarterly*, 8(3):322–342.
- U.S. Department of Justice (2016). Advancing diversity in law enforcement.
- U.S. Kerner Commission (1968). Report of the national advisory commission on civil disorders.
- Weisburd, D., Wooditch, A., Weisburd, Sarit, Y., and Sue-Ming (2015). Do stop, question, and frisk practices deter crime? *Criminology & Public Policy*, 00(0):1–26.
- Weisburst, E. K. (2019a). Police use of force as an extension of arrests: Examining disparities across civilian and officer race. *AEA Papers and Proceedings*, 109:152–156.

- Weisburst, E. K. (2019b). Whose help is on the way? The importance of individual police officers in law enforcement outcomes. Working paper.
- West, J. (2018a). Learning by doing in law enforcement. *University of Santa Cruz Working Paper*.
- West, J. (2018b). Racial bias in police investigations. Working paper.
- Wheller, L., Quinton, P., Fildes, A., and Mills, A. (2013). The Greater Manchester Police procedural justice experiment: Technical report. *College of Policing London Working Paper*.
- Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.
- Xu, Y. and Liu, L. (2018). *gsynth*: Generalized synthetic control method.