Title

Whose help is on the way?: The importance of individual police officers in law enforcement outcomes.

Journal

The Journal of Human Resources, 59(4)

ISSN

0022-166X

Author

Weisburst, Emily

Publication Date

2024-07-01

DOI

10.3368/jhr.0720-11019R2

Peer reviewed

HHS Public Access

Author manuscript

J Hum Resour. Author manuscript; available in PMC 2024 August 14.

Published in final edited form as:

J Hum Resour. 2024 July ; 59(4): 1122–1149. doi:10.3368/jhr.0720-11019R2.

"Whose help is on the way?":

The importance of individual police officers in law enforcement outcomes

Emily K. Weisburst[*]

Emily Weisburst is an assistant professor of public policy at the Luskin School of Public Affairs, University of California, Los Angeles

Abstract

Police discretion has large potential consequences for public trust and safety; however, little is known about the extent of this discretion. I show that arrests critically depend on which officer responds to a 911 call; 1 standard deviation increase in officer arrest propensity raises arrest likelihood by 40%. High arrest officers are more likely to be white and have less experience. I find mixed evidence that arrest propensity is related to arrest quality. High arrest officers use force more often and make more low-level arrests, while they also have a higher share of low-level arrests that result in conviction.

Keywords

Police Discretion and Decision-Making; Economics of Crime; K40; K42; J45

1 Introduction

In 2020, the horrific death of George Floyd at the hands of a police officer in Minneapolis drew widespread attention to issues of police use of force, the effectiveness of police practices, and the stark racial disparities in policing. This death sparked a powerful public reaction in part because it was witnessed in the context of a number of high profile use of force fatalities in recent years. Following each incident, there has been a debate about individual officer actions as well as broader police practices. In many cases, pundits have made conflicting claims about the nature of police behavior and the importance of individual officer decisions, sometimes asserting that "any officer would have responded in the same way" in a given situation and at other times claiming that the events are "isolated incidents attributable to 'bad actors' that do not reflect the rest of a department."[1]

Police officer discretion has large potential consequences for civilian trust, public safety and individuals that interact directly with the police. However, on a basic level, there

[*]Contact address: Department of Public Policy, UCLA Luskin School of Public Affairs, 337 Charles E. Young Dr E., Los Angeles, CA 90095. weisburst@ucla.edu.

[1]This phrasing is not directly attributed to any single pundit or public figure. An example of the first argument can be found in opinion pieces through the organization *Blue Lives Matter*, which was established as a reaction to the *Black Lives Matter* movement (BlueLivesMatter, 2016). The second argument was recently invoked by Attorney General Sessions as a reason to cease the Department of Justice's enforcement of consent decree agreements with police departments, which were established to address civil rights concerns related to law enforcement actions (Department of Justice, 2017).

is surprisingly limited evidence about whether officer decisions actually matter to the outcomes of police interactions after considering the context of an incident. If there are differences in police officer behavior, how large are these differences? Further, how do differences in police officer behavior relate to measures of policing quality?

In this paper, I estimate the degree and importance of police discretion in arrest decisions across officers, conditional on incident context. Within particular incident types, behavioral differences across officers are likely to result from differences in officer skills, experience and preferences. These differences in behavior are most important when they relate to the quality of policing outcomes. For example, if officers who make more arrests make arrests that are relatively “worse” than those made by other officers, this suggests that reducing the discretion of these officers could improve policing quality. More generally, the ability to observe these differences could permit targeted policy interventions, such as monitoring, training, or promotion/demotion, towards officers that display particular behaviors. In this project, I relate individual officer arrest behavior to multiple indicators of arrest quality, such as use of force, measures of charge severity and conviction likelihood.

I analyze police officer arrest decisions using a sample of nearly 2 million calls for service (or 911 calls) and over 1,600 police officers from the Dallas Police Department in Texas (DPD). I estimate individual officer arrest propensity, controlling for detailed information on the characteristics of calls, including call urgency and dispatch code, peer responders and time and geographic factors.

There is sizable variation in individual officer responses to calls for service. I find that a 1 standard deviation increase in an officer’s arrest propensity corresponds to a 40% increase in the likelihood of arrest or 1 additional arrest per 100 calls. This magnitude is comparable to the difference in arrest likelihood between a high priority major disturbance call and a low priority criminal mischief call, or the difference between a call that is instantaneously dispatched and a call that is dispatched 90 minutes after a call is made. This estimate is also larger than comparable measures of judge leniency on conviction, pretrial detention and juvenile detention rates in the literature, which find that a 1 standard deviation increase in judge leniency corresponds to a 5-20% increase in these outcomes (e.g. Bhuller et al., 2020; Dobbie et al., 2018; Aizer and Doyle, 2015). Moreover, the variance in arrests explained by individual officers is larger than the variance explained by geographic police beats, which is notable given that geographic sub-regions within a city vary substantially in demographics, crime rates and police intensity.

Throughout the analysis, I pay particular attention to patterns of officer sorting to calls for service and conduct a number of robustness checks to verify that the observed dispersion in arrest behavior is not an artifact of selection. These checks include verifying that officer fixed effect estimates are similar in a model that excludes call-specific covariates and are similar in settings that are less likely to be affected by sorting, such as urgent calls and calls placed when fewer officers are available to respond.

After estimating the arrest propensity of each officer, I investigate relationships between officer arrest propensity and multiple indicators of arrest quality. I estimate officer arrest propensity in a training sample, and then relate estimates of officer arrest propensity to characteristics of each officer's arrests in a test sample, conducting this analysis using 100 iterations of random partitions of the data. While higher and lower arrest propensity officers face comparable crime offending environments, higher arrest propensity officers are significantly more likely to use physical force against a civilian during an arrest, and make arrests for less severe crimes, including drug offenses, resisting arrest, and failure to provide identification. Among these misdemeanor arrests, higher arrest type officers are more likely to have their arrests result in conviction. The results provide a mixed picture of the relationship between arrest propensity and arrest quality. Higher arrest officers have a lower severity threshold for making arrests and appear to be more physically aggressive; however, these officers may also be more effective at obtaining convictions in court.

Linking the officer arrest propensity estimates to officer demographics, I find that white officers and less experienced officers have moderately higher arrest propensities. In an illustrative mechanisms analysis, I project officer arrest propensity on officer demographics to decompose the relationship between officer arrest propensity and arrest outcome characteristics. I find that both officer experience and race are important drivers of this relationship, which suggests that policies that accelerate officer learning or increase officer diversity could reduce the incidence of use of force and increase the severity threshold for making an arrest. This finding complements a growing body of work investigating the relationship between police officer behavior and experience (Deangelo and Owens, 2017; West, 2019). In particular, consistent with the pattern of results in this study, West (2019) finds that more experienced officers are more likely to find contraband when they search vehicles during traffic stops.

This paper contributes to the literature in economics and criminology[2] by providing novel evidence of the extent and importance of officer-level police discretion. A major advantage of this study is the ability to study police discretion using 911 call for service data, a setting where each observed call response is originally initiated by a civilian and not by a police officer. Researchers in economics have typically restricted their attention to law enforcement interactions that are initiated by officers, such as searches in traffic stops, speeding tickets and pedestrian stop and frisk interviews (e.g. Gonçalves and Mello, 2021; Horrace and Rohlin, 2016; Anbarci and Lee, 2014; Antonovics and Knight, 2009; Ridgeway and MacDonald, 2009; Gelman et al., 2007; Anwar and Fang, 2006; Grogger and Ridgeway, 2006; Knowles et al., 2001). Importantly, these interactions are a choice variable of the police officers involved. A growing body of research finds that race can also be a factor in police decisions to make traffic (or pedestrian) stops and that studies that focus only on the outcomes of these interactions neglect to consider police discretion that contributes

[2]Criminologists and sociologists have documented several dimensions of police discretion, including variation in work-related decisions, interpretation and implementation of the law and the use of extra-legal factors, such as suspect race, in decision-making (e.g. Fagan et al., 2016; Nickels, 2007; Frydl and Skogan, 2004; Mastrofski, 2004; Walker, 1993; Reiss, 1971). Scholars have long noted that resource and managerial oversight constraints are particularly acute in the public services sector, including the high-stakes setting of policing (e.g. Lipsky, 1980). However, in the current police environment, it is unclear whether police discretion remains a dominant force, as recent technological advancements, such as automated data systems and body cameras, have the potential to increase police oversight and reduce discretion (Ridgeway, 2018), or could exacerbate differences in police treatment of civilians (Brayne, 2017; Joh, 2016). I contribute to this literature by investigating the importance of police discretion in the current environment using evidence from high frequency 911 call data.

to sample selection (Knox et al., 2020; Horrace and Rohlin, 2016; Gelman et al., 2007; Grogger and Ridgeway, 2006). This paper addresses these important selection issues by restricting attention to the setting of 911 calls. A parallel strategy is employed by West (2018), who studies racial bias among state troopers who are randomly dispatched to motor vehicle accidents, and in Hoekstra and Sloan (2022), who study racial bias in officer shootings using a sample of 911 calls for service.

Further, while the literature has frequently exploited aggregate officer demographic characteristics, nearly all of the work in this space does not incorporate officer identity. An exception is work by Ridgeway and MacDonald (2009) examining officer-initiated pedestrian stops in New York City, comparing the race distribution of individual officer stops to stops made by similar peer groups. This paper also complements work by Gonçalves and Mello (2021) that measures individual officer effects in a racial bias test applied to Florida Highway Patrol officer decisions to issue speeding tickets.

Other contributions of this study include the richness of the data and the variety of policing interactions that I examine.[3] I draw on multiple data sources to construct several measures of arrest characteristics and outcomes, covering officer use of force, arrest charge type and severity, and court dismissal or conviction outcomes. These measures of arrest characteristics inform our understanding of the multiple dimensions of officer arrest decisions and provide insight about the attributes of marginal arrests. Further, the analysis sample of 911 calls covers a diverse cross-section of police work, allowing examination of responses to a broad swath of incidents and offenses.

2 Institutional Background and Description of Data

2.1 Protocols for Dispatch and Call Responses

When a civilian calls DPD for police assistance, they are connected to a 911 call-taker. The call-taker creates an active call report that summarizes important facts related to the incident, including location and relevant descriptions of the events. Active call reports also include a dispatch code that categorizes the incident type and priority level for response. Given a set of open active calls, DPD dispatchers assign available officers to calls. Calls are dispatched according to their priority, or their level of severity and urgency. When there is a long call queue, responses to low priority calls are postponed until more serious calls have been resolved. The pool of available officers when a call is received depends on patrol responses to other incidents at the time. (Figure A1 depicts the response process in Dallas).

Patrol officers are the primary responders to calls for service. Officers are assigned to work in 1 of 7 police divisions in the city for 8-hour shifts, or watches, from 12am-8am, 8am-4pm and 4pm-12am.[4] Regular patrol shift schedules are set once a year, based on the seniority

[3]A recent working paper by Weisburd (2019) uses similar 911 call data from Dallas, to answer a different research question: whether police patrol presence reduces crime. Work in criminology has also used the setting of 911 call responses to evaluate body worn camera field experiments (e.g. Braga et al., 2018). Further, criminologists have qualitatively examined police responses to 911 calls for service (Reiss, 1971) and measured the aggregate arrest outcomes of 911 calls (Spelman and Brown, 1984).
[4]The three 8-hour shifts used in the analysis are approximate, in practice some officers work 10 hour shifts and other officers have start and end times that are slightly staggered across police shifts.

of officers.[5] Calls are assigned to patrol officers who work within the geographic police division where the call incident occurred.

When officers are not responding to calls, they can use their discretion to determine how to patrol their assigned police beat, within an assigned geographic police sector. This may include circling the area, engaging with community members, investigating suspicious activity, initiating civilian stops, and making arrests. Officers inform the dispatcher when they are engaged in an officer-initiated activity, and in most cases, the dispatcher will consider the officer unavailable for any call responses until the activity has concluded. The dispatcher may divert an officer from an officer-initiated activity to a call in cases where there is an urgent or high priority call and/or no other officers are available to respond. In practice, this means that officers who are highly active in initiating activity while on patrol may end up responding to fewer calls than less active officers. As discussed above, this project focuses on measuring police discretion only in civilian-initiated calls rather than other officer-initiated activities.

Officers typically conduct patrol in police cars, alone or in pairs. At the beginning of each shift, officers may choose to patrol with another officer, depending on the number of cars available for that shift. Each car is considered an “element” that can be dispatched to an incident. Paired officers respond to all calls together throughout a shift.

If more than one patrol element is available to respond to an incident at the time of dispatch, dispatchers consider a number of factors in their assignment of available officers. More serious incidents may require or benefit from a response by multiple officers or cars. Additionally, officers who are geographically close to an incident are more likely to be dispatched to the incident, especially if the call is urgent. At the same time, when a large pool of officers is available to respond to a call within a division, officers may volunteer to take particular calls as they are posted, a potential source of selection.

The estimates in the study will impacted by selection bias if officers choose to respond to calls based on incident characteristics that are unobservable. To address this concern, I conduct a series of tests to verify that selection does not affect the empirical estimates (see Section 3.3).

This project focuses only on the *first or initial group* of officers to be dispatched to a call, based on the call dispatch time stamp. Multiple officers may be dispatched simultaneously when officers are either patrolling together or multiple cars are dispatched for serious calls.

When the assigned patrol element arrives at the scene of the incident, the responding officer(s) determines if an offense occurred, gather information, investigates the scene and assists the complainant or victim. If an officer determines that an offense occurred, the officer submits an offense report to a staff reviewer at DPD who examines the report for completeness. After this initial review, the offense may be assigned to a detective in

[5]Depending on the needs of the department, officers may choose to work overtime patrol shifts outside of their regular shift schedules, though these shifts are also set in advance, typically a month or a week prior.

an investigative unit based on the offense type. The assigned detective will then pursue additional investigation of the offense, if warranted.

Over the course of a call response, officers may identify a suspect and/or make an arrest. Alternatively, an arrest may be made at a later date after a detective assumes responsibility for a follow-up investigation of the offense. Individual responding officers have the ability to influence arrests directly, by making the decision to apprehend an individual at the scene of the incident, or indirectly, by laying the groundwork for an investigation through gathering information for the initial offense report. In practice, nearly all arrests occur within a day of the initial call response and the responding officer is involved in the arrest.[6]

Officers do not face explicit incentives to make a large number of arrests and DPD does not impose a policy of arrest quotas. Like most police departments in the U.S., DPD does aim to solve as many reported crimes as possible through arresting suspects associated with reported crime incidents. This department goal largely applies to more serious or violent crimes and is less likely to be a factor for more minor offenses.

2.2 Dallas Police Department Data

The setting for this study is the Dallas Police Department (DPD) in Dallas, Texas. This project uses administrative DPD data covering dispatched calls for service (911 calls), records of arrests and the names and badge numbers of responding officers between June 2014 and January 2019. This data is supplemented with DPD and Dallas County information on arrest charges, non-shooting use of force incidents, and court records. I construct the primary arrest outcome using a liberal definition of arrests, coding a call incident as having an arrest if any of the DPD data files obtained for the study indicate that an arrest was made. Additionally, I merge the DPD data sets with demographic information on police officers obtained through an open records request to the city of Dallas. The officer data includes officer race, gender, age, experience, salary and job title. (See the Online Data Appendix A4 for more detail).

While the data is pulled from the call dispatch system, the data includes some call events that are officer-initiated rather than complainant-initiated. I clean the data to exclude calls listed as officer “Mark-Outs”, or records where an officer initiates an investigation and then convey their location to the dispatcher, as well as traffic stops, calls where officers respond to assist other officers in the field, fire department related calls, and other call types associated with officer-initiated investigations. I also restrict attention to the first group of officers dispatched to a call, excluding responders that are sent to the scene after the initial group.

Next, I trim the sample to include only calls that are likely to involve patrol officers or a sample of routine patrol calls. I do this by first excluding calls that are dispatched after unusually long delays, or 5 hours or more after the initial call. I also exclude calls that include officers who do not appear to be regular patrol officers, as they have few

[6]Of the data with information on arrest dates (87% of arrests), 97% of arrests occur within a day of the incident. When there is information on the arresting officer (85% of arrests), 82% of arrests involve an original responding officer.

call responses in the sample. Specifically, I exclude calls where any officer responding to a particular call has fewer than 1000 observations in the raw data (inclusive of all original records, including officer-initiated interactions). In practice, the characteristics of the trimmed sample and unrestricted sample are similar (see Table A2).

2.3 Summary Statistics

While this study is restricted to a single city, Dallas is representative of other policing contexts in similarly sized cities within the U.S. Dallas is a large and diverse urban center, with over 1.3 million residents and a population that is 42% Hispanic/Latinx, 24% Black and 29% white (Census, 2017). The city of Dallas experienced 775 violent crimes and 3,182 property crimes per 100,000 residents in 2017. These crime rates are slightly higher than other cities with over 1 million residents, which averaged 720 violent crimes and 2,643 property crimes per 100,000 residents in the same year (FBI, 2018).

Table 1 summarizes the analysis data at the call level. The analysis sample includes over 1.9 million calls and over 3 million call by officer records for over 1,600 officers. Table A2 shows that the analysis sample and the untrimmed sample are very similar; this consistency suggests it is suitable to generalize results from the analysis sample.

On average, it takes 34 minutes for a patrol officer to be dispatched to an incident after a call is made, with a standard deviation of 53 minutes. The variation in this dispatch time lag highlights the fact that dispatchers prioritize calls based on severity and that officers cannot immediately respond to all incidents. On average, 32% of officers on a shift are unavailable to respond, or are engaged in another police response, at the time of each dispatched call. The most common dispatch codes are for major disturbances, burglaries and criminal mischief. At the time of dispatch, only a small number of incidents are designated as violent offenses; robberies, criminal assaults, and armed encounters collectively comprise less than 3% of calls. Other call types such as major disturbances, injured person, and accidents may also ultimately be associated with a violent crime, though these outcomes are not readily apparent at the time of dispatch.

Approximately 3% of call responses result in an arrest. DPD makes approximately 64 thousand total arrests per year, of which approximately 27 thousand arrests, or 42%, result from 911 call responses. While the court data in this study covers Dallas County, an area with over twice the population of the city of Dallas, arrests resulting from DPD call responses nevertheless represent substantial portion of the charges and convictions in county court, or 13% of total charges and convictions.

White officers and Black arrestees are over-represented relative to the population of Dallas. White patrol officers respond to 48% of call incidents, while Black and Hispanic/Latinx officers respond to 26% and 21% of incidents, respectively. Relative to the sample of arrests with demographic information for arrestees, 50% of arrests have a Black arrestee, 24% have a white arrestee, and 23% have a Hispanic/Latinx arrestee.[7] 7% of call responses involve a police officer in training, less than 1% involve a police sergeant, and 14% involve a female

[7]Demographic information is not available for 13% of arrests in the sample.

officer. Averaged across call responses, DPD patrol officers earn approximately $57,000 per year. Summary statistics tabulated at the officer level are similar and are displayed in Appendix Table A1.

3 Police Discretion in Arrests

3.1 Empirical Model

I use a predictive model of arrests to estimate each officer's arrest propensity and measure the dispersion in this propensity across officers.

As a first step, I estimate the following linear probability model,

$$Arrest_{ikgt} = \theta_i + \pi X_{kt} + \delta_{dt} + \phi_{g(k)} + \psi_{g(i)} + \varepsilon_{ikgt}$$

where i indexes the responding officer within a group of I_k officers responding to a call (if other officers are present), k indexes the incident, d indexes geographic police divisions, g indexes police beat or police sector location, and t indexes time. The outcome $Arrest_{ikgt}$ is the primary focus of the analysis and denotes whether an arrest was made in association with a 911 call. X_{kt} are a set of incident specific characteristics, including 20 aggregated dispatch codes and indicators for hour within a shift. X_{kt} includes controls for the number of other officers responding to the call with the focal officer in the same car and the number of other officers responding from different car patrol units. These controls account for the fact that some calls require multiple responders and that some officers choose to patrol with a partner at the beginning of their shift. X_{kt} also includes controls for the *urgency or severity* of the call, defined as the number of minutes that pass between when a call occurs and the time of dispatch (entered as a linear and quadratic term). Further, X_{kt} includes an indicator for whether an address has received more than 1,000 dispatch responses in the sample period. Additionally, X_{kt} includes the proportion of officers available to be dispatched (relative to those working a shift) at the time of each call event. Lastly, X_{kt} includes an indicator for whether the call response is in the same geographic police sector as an officer's prior call response.

$\phi_{g(k)}$ are indicators for police beat locations of calls and $\psi_{g(i)}$ are indicators for an officer's assigned police patrol sector. These variables control for time-constant differences in arrest patterns across geography. There are 234 beats in Dallas and each is fully contained within 1 of the 7 police divisions and 1 of the 35 police sectors in the city. $\phi_{g(k)}$ capture time constant differences in the locations of calls or incidents. $\psi_{g(i)}$ captures time constant differences in the locations of officer patrol assignments, which capture aspects of officer peers and patrol location norms. Officers need not respond to calls exclusively within their assigned patrol sector location.

δ_{dt} are shift indicators that capture time-varying location-specific arrest patterns that are associated with specific shift assignments. These variables are Police Division*Day-of-the-Week*8-hour Shift*Month*Year fixed effects. To increase power, the baseline model does not include a separate indicator for each individual shift, but rather aggregates shifts into month by year groups.[8] For example, the four Tuesday evening shifts in the Central Division are grouped in January 2016. In Section 3.3, I conduct a number of robustness specifications that vary the fixed effects of the model and find similar results.[9]

θ_i is the key variable in the model and measures the time-invariant or permanent arrest propensity of officer i. Given the numerous controls in the empirical model, θ_i represents an officer specific effect that is measured within dispatch call type, shift cell and geographic location. In practice, observations with multiple responders are duplicated for each responder, and regressions are weighted by the inverse of the number of responders for each call, $1/I_k$. Given this weighting, the analysis is conducted at the call-level.

Using this model, I calculate the dispersion in officer-level permanent arrest propensity as the standard deviation of the distribution of θ_i across officers. In order to establish a conservative estimate of police officer dispersion, I adjust the estimates of θ_i terms using Empirical Bayes techniques.

I calculate the adjusted estimates of θ_i using the following steps. First, I estimate officer fixed effects $\hat{\theta}_{i,raw}$ using the baseline specification. I use the full fixed effects model to calculate these effects, allowing the officer effects to be arbitrarily correlated with the covariates in the model. Next, I compute a composite residual, $\hat{r}_{ikgt} = \hat{\theta}_{i,raw} + \hat{\varepsilon}_{ikgt}$, and estimate its variance, $\sigma_r^2 = E[r_{ikgt}^2]$ using the sample analog of the squared residual, as well as the within officer residual variance, $\sigma_\varepsilon^2 = E[\varepsilon_{ikgt}^2]$. Across officer variance is calculated as $\sigma_A^2 = \sigma_r^2 - \sigma_\varepsilon^2$. I then calculate the individual adjusted officer arrest propensities using the following transformation: $\hat{\theta}_i^{EB} = \sigma_A^2/(\sigma_A^2 + \frac{\sigma_{\varepsilon,i}^2}{N_i}) \cdot \hat{\theta}_{i,raw}$ The estimate of a standard deviation in the officer arrest propensity distribution is given by the empirical estimate of $\hat{\sigma}_A$. (See Online Appendix A3 for more detail.)

This procedure produces a “shrinkage factor,” $\sigma_A^2/(\sigma_A^2 + \frac{\sigma_{\varepsilon,i}^2}{N_i})$, which adjusts individual officer effects toward zero when the number of observations per officer, N_i, is small, or the variation in the officer effect, $\sigma_{\varepsilon,i}^2$, is large. Throughout this paper, I focus on results using the estimate of a standard deviation in the officer effect distribution, $\hat{\sigma}_A$, and the adjusted estimates of individual officer effects, $\hat{\theta}_i^{EB}$, and refer to these adjusted estimates as $\hat{\theta}_i$. The results are not an artifact of this precision adjustment and are comparable when unadjusted fixed effects from the first stage are used (see Section 3.3 for a discussion of alternate precision adjustments).

3.2 Results

Individual police officers vary substantially in their arrest behavior. Figure 1 shows the estimated distribution of officer effects, $\hat{\theta}_i$. For each officer, $\hat{\theta}_i$ represents his/her permanent or time-invariant arrest propensity, conditional on time and geography controls,

[8]A version of the model with controls for individual 8-hour shifts is estimated as a robustness check in column (4) of Table A7.
[9]There are 1,608 officers, 8,232 shift categories, 234 police beats, and 43 assigned police sectors. The number of assigned police sector categories includes groups where this information is missing; when possible it is replaced by police division.

call characteristics, and peer influence. This estimated distribution has a longer right tail, showing that a small number of officers have especially high arrest propensities.

Swapping an officer that has a low arrest propensity with one that has a high arrest propensity can critically change the outcome of a call response. A 1 standard deviation in $\hat{\theta}_i$ corresponds to 0.07 standard deviations in the total arrest outcome. In percentage terms, a 1 standard deviation increase in an officer's arrest propensity corresponds to a 41% increase in the likelihood that a given call incident results in an arrest, relative to the mean arrest to call rate of 2.6%. Similarly, a 1 standard deviation increase in the officer arrest propensity distribution corresponds to 1 additional arrest per 100 calls.

It is useful to benchmark this estimate against other factors that contribute to arrests in call responses. A 1 standard deviation increase in an officer's arrest propensity is approximately equivalent to the difference in arrest likelihood between: a high priority major disturbance call and a low priority criminal mischief call, a robbery call and a call classified as "other" that is low priority, and a call that is dispatched 90 minutes after it is placed versus a call that is instantaneously dispatched. (See Online Appendix A2 for results of the first stage).

A more formal way to consider the contribution of officer effects is to estimate a decomposition of the variance of arrests. Table 2 displays two different decompositions of the outcome variance. Panel A estimates the ensemble variance decomposition proposed in Card et al. (2018) where each variance component is the covariance of that model component with the arrest outcome. Officer fixed effects comprise 0.5% of the total outcome variance, where the total variance of arrests in the sample is 0.026. While the share of total variance is the primary metric of the contribution of officer effects, it is also instructive to compare the contribution of officers to the variance *explained* by the model. This benchmark is useful given that the bulk of the arrest outcome cannot be explained by observable characteristics, despite the rich detail in the control variables. Officer fixed effects comprise 20% of the variance explained by the model, given a total model R^2 of 0.026.

In Panel B of Table 2, I decompose the outcome variation by estimating the incremental increase in R^2 that results from adding a component to the model. These measures are estimated as $R^2_{total} - R^2_{-j}$ for each model component j. Here, I find similar results; the focal officer effects account for 0.37-0.42% of the total variation and 16-18% of the explained variation using the R^2 and adjusted R^2 metrics, respectively. In both decompositions, the variance explained by differences across individual officers is larger than the variance explained by geographic police beat fixed effects; a result that is striking given that geographic sub-regions within a city vary substantially in demographics, crime rates and police intensity.

Lastly, Table 2 shows the correlation between the estimated components of the model at the observation level in Panel C. Each of the correlations between officer effects, θ_j, and the other model components is close to zero. While the absolute magnitude of these correlations are small, the sign of correlations indicates that officers with higher arrest propensity are more likely to respond to calls in police beats where arrests are more likely. Further, higher arrest officers are more likely to respond to call types where arrests are more likely.

A natural next step is to consider how the estimated officer fixed effects, $\hat{\theta}_i$, are associated with officer demographic characteristics. Appendix Table A1 tabulates demographic characteristics of officers across officers with low, medium or high arrest propensities (terciles of $\hat{\theta}_i$). These descriptive statistics show that officers with lower levels of experience, younger officers, and trainee officers are more likely to have higher arrest propensities. Officers with lower arrest propensities are more likely to be Black and less likely to be white.

More formally, Table 3 shows the results of regressing $\hat{\theta}_i$ terms on officer race, gender, age, trainee or sergeant status and experience.[10] These regressions offer information about whether officers with specific traits systematically differ in their arrest propensities.

Overall, the results imply that white officers with less experience have higher arrest propensities, on average. All else equal, the likelihood of arrest is 21% lower when a responding officer has 10 years of experience instead of 5 years of experience. Black (Hispanic/Latinx) officers are 13% (5%) less likely to make arrests relative to white officers. Male and female officers do not have statistically different arrest propensities. Conversely, the small share of sergeants in the sample have a higher arrest propensity, with 26% higher likelihood of arrests than non-sergeants. Demographic characteristics collectively explain 22% of the variation in officer arrest propensity.

3.3 Model Validity and Robustness

3.3.1 Officer Sorting—While the baseline model includes controls for a rich array of observable call characteristics, the estimates of individual officer arrest propensity could be biased if officers systematically respond to calls based on unobservable call characteristics. The identification assumption of the model is that *conditional on observable call characteristics, the location of calls, officer sector and shift assignments, officer assignment to calls is idiosyncratic or as-if random.*

There are two potential patterns of selection. First officers with a high arrest propensity may be more *active*, both in the sense that they could respond to more calls or that they could be more likely to respond to marginal or less serious calls. This pattern could create a negative correlation between officer effects, $\hat{\theta}_i$, and the error terms, ε_{ikgt}. This negative selection bias would deflate the dispersion in officer fixed effects and lead to an underestimate of this parameter.

To address this concern, I test whether officers with higher arrest propensities respond to more calls for service. The correlation between arrest propensity and the number of calls per officer is −0.12 (Table 4). This small negative relationship suggests that high arrest officers are not more *active* in responding to calls than low arrest officers. This negative relationship may be attributable in part to the fact that high arrest officers might be more active in conducting officer-initiated investigations or arrests while on patrol, rendering these officers less available to respond to calls, on average. This dynamic means that it is unlikely that

[10] Salary is omitted from this regression because it is nearly perfectly correlated with experience, given the compensation formulas used by the department.

high arrest officers are "lying in wait" to respond to marginal calls that are less likely to result in arrest.

Alternatively, officers who have a high arrest propensity could prefer to respond to incidents with a higher unobservable likelihood of arrest, and officers who have a low arrest propensity could prefer to volunteer for incidents with a lower unobservable likelihood of arrest. In either case, the estimates of $\hat{\theta}_i$ will be positively correlated with the error terms ε_{ikgt} and this would inflate the estimate of dispersion in officer fixed effects.

First, I consider the importance of observable call characteristics that may affect officer response choices. Officers may choose to respond to calls based on call severity, dispatch code, availability of other officers working their shift and time within shift as well as the beat of the call, or factors captured in X_{kt} and $\phi_{g(k)}$. I estimate $\hat{\theta}_i^{'}$ from a model that retains only controls that are pre-determined at the time of a response, namely shift fixed effects, assigned police sector, and the number of other officers patrolling in the same vehicle during a shift. This "Without Controls" version of the model is displayed in Column (2) of Table 4 and Panel (A) of Figure 2.

Perfect correlation between these estimates would imply that officer effects are orthogonal to the set of call characteristics in the model, or are nearly randomly assigned to calls. This figure shows that the $\hat{\theta}_i^{'}$ is similarly distributed relative to the $\hat{\theta}_i$ estimates from the full model, with a standard deviation of 0.012 that is similar to the baseline standard deviation of 0.01. The individual estimates across these models have a very high correlation of 0.98. Further, I fail to reject equality of the distributions using a Kolmogorov-Smirnov test. Overall, this suggests that observable call choice characteristics do not bias the estimation of the officer effects distribution.[11]

Next, I conduct a balance test by regressing officer fixed effects, $\hat{\theta}_i$, on observable call characteristics and calculate a joint F-test of the significance of these characteristics (Table 4). I find that call characteristics are jointly significant with an F-statistic of 3.17. However, this statistic is small when benchmarked against a regression of actual arrest outcomes on call characteristics, where the corresponding F-statistic is over 10 times larger or 43.3. Moreover, as discussed above, removing these observable call characteristics from the model does not affect the distribution of officer fixed effect estimates.

I further test for the importance of officer sorting by focusing attention on two settings where officer sorting is less likely to impact the results, calls that occur when few officers are available, a "Low Availability" sample, and urgent calls, a "High Urgency" sample. In both settings, officers will be more constrained in their ability to volunteer to respond to a call. I define the "Low Availability" sub-sample by splitting the sample at the median officer availability rate. Likewise, I define "High Urgency" calls as those that have below median time between when a call is received and dispatched. Appendix Tables A3 and A4 show that

[11]Figure A3 plots the raw or unconditional officer arrest propensity (centered) against the covariate-adjusted effects. It is notable that there is a very high correlation between these estimates of 0.9, even without including any pre-determined covariates (or factors that officers cannot choose at the call level) in the comparison estimates.

these sub-samples do not meaningfully differ from the baseline sample in dimensions other than officer availability or call urgency.

Columns (3) and (4) of Table 4 and Panels (B) and (C) of Figure 2 show the results of restricting the observations to the "Low Availability" and "High Urgency" sample. If dispersion in officer behavior is increased by officer sorting, we would expect the estimates of dispersion to be larger in the baseline model relative to these robustness samples, where sorting is constrained. However, the graphs show a strikingly close match between the distributions. The estimated dispersion in the "Low Availability" and "High Urgency" samples is nearly identical to the baseline, with standard deviations of 0.01 and 0.12 respectively. While the distributions statistically differ from the baseline model distribution using a Kolmogorov-Smirnov test, the individual estimates from these alternative samples have a high correlation with the baseline sample of 0.92 - 0.95.

These sub-sample analyses focus on median splits of the data in officer availability and call urgency in order to maintain a comparable call sample in other dimensions of calls, such as call type and call severity. In the appendix, I conduct additional versions of this test using cuts of the data that likewise maintain a comparable call samples across officer availability and call urgency. I repeat this officer availability and call urgency sub-sample exercise within only high priority calls in Appendix Figure A4, within only low priority calls in Figure A5, and within calls that have a high predicted arrest likelihood (excluding officer effects) in Figure A6. Across each of these finer cuts of the data, the distributions in the sub-samples where sorting is constrained look similar. Collectively, these figures show that within call sub-types estimates look similar in sub-samples where officers are constrained in their choices to sort to calls.

As a final test of officer sorting, I examine how the impact of officer demographics changes in across the "Without Controls" specification, and the "Low Availability" and "High Urgency" sub-samples. In Appendix Table A5, I regress the officer effects estimated from each of these robustness models on officer demographic characteristics, and in Appendix Table A6, I regress the arrest outcome directly on officer demographic characteristics in each robustness model. In both tables, the coefficients on officer demographic characteristics are remarkably stable across the baseline model and the robustness specifications. Similar to the tests above, this evidence provides support that officer sorting is not driving the results.

3.3.2 Precision and Specification Tests—Separate from concerns about officer sorting, there could be concerns about the precision of the officer fixed effect estimates. Even in the absence of true officer differences, there will be some measured variation in outcomes across officers, simply due to idiosyncratic variation in the error term. To address this issue, I randomly re-assign call responses to placebo officers, in a manner that preserves the total call response distribution across officers in the data. Figure A7 displays the results of 100 replications of this placebo test. The actual model estimate is well outside the 95% confidence interval given by the estimated distribution from the placebo test, confirming that the estimated variation in officer effects is not simply due to noise in the data.[12]

In the Online Appendix Table A7, I additionally show that the results are robust to several alternate specifications of the model, which include narrower geographic, time, and call type controls (including individual 8-hour shift fixed effects, fixed effects for >1000 geographic police reporting areas which sub-divide police beats, controls for assigned beat rather than assigned police sector, the full set of disaggregated dispatch call codes, and adding fixed effects for the dispatchers and call-taker associated with each call). I also consider alternate methods of adjusting the estimates for precision, including weighting the unadjusted estimates by the number of calls per officer, and comparing the Empirical Bayes adjusted estimates to unadjusted officer fixed effects. Lastly, I estimate officer effects using a random effects model rather than a fixed effects model for comparison.

Across these alternative specifications and precision methods, the dispersion estimates are very similar to the base model, with a 1 standard deviation in officer effects corresponding to a 34-47% increase in arrest probability. With no adjustment for precision, this standard deviation estimate is not substantively larger than the base model, corresponding to a 47% increase in arrest probability. Further, I fail to reject that the distribution of officer effects is not the same as the baseline model for nearly all of the robustness specifications.

Most notably, the random effects estimates are not statistically different from the fixed effects estimates in the baseline model. Because a random effects model assumes that officer assignment to calls is independent from the observable characteristics of the model, this test provides additional evidence that officer sorting is not driving the results. Moreover, across all robustness specifications in Appendix Table A7, the correlation between adjusted and unadjusted officer effects is greater than 0.94.

4 Arrest Propensity and Arrest Outcome Characteristics

4.1 Analysis Framework

Identifying meaningful variation in arrest outcomes across individual officers provides evidence about the existence and extent of police discretion in the field. However, arrests are not a normative outcome. An arrest may have positive or negative welfare consequences depending on the incident context, culpability of the arrestee, severity of the offense, implycations for public safety, as well as the subsequent burden for the arrested individual and his/her family.

In this section, I examine relationships between officer arrest propensity and characteristics of arrest outcomes, in order to better understand the margins of arrest decisions across officers. One possibility is that officers with higher arrest propensities may simply be more *productive* than lower arrest propensity officers. Alternatively, higher arrest propensity officers could have a lower severity or evidence threshold for making an arrest and/or be relatively more *aggressive*.

[12]This test is similar to the exercise used in Guell et al. (2015), which randomly re-assigns surnames to individuals in a manner that preserves the skewed distribution of names in their earnings data. To be conservative, the officer effect estimates in the placebo distribution are not adjusted toward 0 using Empirical Bayes' shrinkage techniques.

Specifically, I identify a number of different arrest outcome characteristics that are associated with the quality of arrests to compare performance across officer types. These include court conviction, dismissal, and non-conviction outcomes for arrest charges. The measures also include adverse consequences of arrest events, such as officer use of force. No single factor is a definitive marker of a “good” or “bad” arrest, but collectively, the analysis illustrates relative differences in arrest quality across higher and lower arrest propensity officers.

I draw on multiple data sources from DPD and Dallas County on arrests, offense reports, county court outcomes, and non-shooting use of force incidents to construct measures of arrest outcome characteristics. Each of these characteristics measures the outcomes of arrests that result from the 911 call responses (see description in Table 5 and Online Data Appendix A4 for additional details on data features and cleaning).

I examine associations between officer arrest propensity and indicators of arrest quality by regressing various arrest outcomes on estimates of officer arrest fixed effects, $\widehat{\theta}_i$. I use estimates of $\widehat{\theta}_{i,train}$ from a training sample period and outcomes from a non-overlapping test sample period in order to address concerns about the joint determination of these variables. I determine test and training samples by randomly partitioning the dates in the study sample. Within this partition, I first estimate $\widehat{\theta}_{i,train}$ using the model described in Section 3.1. I then estimate the following model within the *test* sample:

$$ArrestOutcome_{ikgt} = \beta\widehat{\theta}_{i,train} + \pi X_{kt} + \delta_{dt} + \phi_{g(k)} + \psi_{g(i)} + e_{ikgt}$$

The controls in this model are the same as the full set of controls in the baseline model, though here the sample is restricted to individual arrests that result from call responses rather than all call observations.[13] I focus on a base of arrests in order to estimate outcomes as a share of arrests. In contrast to the alternative of using a base of all calls, this setup avoids measuring mechanical increases in any arrest outcome that results from the fact that higher arrest officers simply make more total arrests. For example, officer use of force requires interaction with a civilian, and therefore officers who make more arrests may be automatically be more predisposed to use force as a function of call responses; however, this fact is not informative about whether high arrest officers are more aggressive conditional on making an arrest. Focusing on a base of arrest observations addresses this issue.

In this exercise, the test and training sample designations are inherently arbitrary. To provide robust estimates that account for the fact that $\widehat{\theta}_{i,train}$ is an estimated regressor, I randomly partition the dates in the sample into 100 test and training samples and estimate 100 iterations of the regression above. This procedure results in 100 estimates β^j. I report the average and standard deviation of these estimates as the preferred measure of $\widehat{\beta}$ and $se(\beta)$, similar to a bootstrap simulation. Likewise, I use the distribution of estimated t-statistics, t^j, in each iteration to determine the p-values associated with each arrest outcome regression.[14]

[13]As in the baseline model with arrest outcomes, the model is estimated at the arrest by officer level and weighted by the inverse number of officers involved in the initial call response that led to an arrest. This weighting results in estimates that correspond to the arrest level.

I report the magnitude of a percent change in each arrest outcome given a 1 standard deviation increase in officer arrest propensity as part of this analysis. This percent change relative to the mean of each arrest outcome (as a share of arrests) is useful to compare effect magnitudes, as each outcome has a different mean and support. Additionally, given that the $\hat{\theta}_{i,train}$ and $ArrestOutcome_{ikgt}$ are calculated within different samples, calls and arrests, respectively, the magnitude of β does not have an obvious interpretation on its own.

As discussed above, $\hat{\theta}_{i,train}$ can be viewed as estimates of each officer's permanent underlying arrest propensity, as they are derived from a sample of interactions that are initiated by civilians and not officers, and adjusted for incident characteristics, geography and time factors.

4.2 Results

As a first step, I check whether the training sample $\hat{\theta}_{i,train}$ are reasonably associated with baseline characteristics of officer-level outcomes in the test sample in Panel A of Table 5. This first set of outcomes is measured at the officer-level in simple bivariate regressions of the outcome in the test-sample on $\hat{\theta}_{i,train}$. The first row shows that $\hat{\theta}_i$ measured in the training sample is strongly positively correlated (correlation of 0.79) to a similarly estimated officer effect, $\hat{\theta}_i^{test}$, in the test sample, indicating that officer arrest propensity type is highly persistent across the partitions of the sample. As noted in the officer sorting analysis in Section 3.3, officer arrest propensity is negatively related to the total number of calls an officer responds to (Row (3)).

I additionally test whether officer arrest propensity is related to crime conditions in Rows (4) - (6). I do this by calculating crime exposure for each officer as the average daily number of crimes reported by other officers working on the same shifts as the focal officer.[15] Officer arrest propensity is not significantly correlated with total crime exposure, and the relationships to property crime (decrease) and violent crime (increase) go in opposite directions. The size of the violent and property crime relationships are both small in magnitude with a 1 standard deviation in officer arrest propensity associated with a 1.5% more violent crime and 2.2% less property crime. These tests provide evidence that high and low arrest propensity officers do not materially differ in their arrest outcomes because they face different crime conditions.

Next, I examine several characteristics of arrest outcomes in Panel B of Table 5. Rows (6) and (7) show that high and low arrest officers a similar share of total felony and misdemeanor arrests. Appendix Table A8 expands on this analysis by showing the relationship between officer arrest propensity and the share of officer arrests by type of charge. In this disaggregated set of results, it is clear that high arrest officers make relatively more arrests for lower level offenses, and fewer arrests for serious felonies. The results in

[14]P-values are estimated as the share of centered t-statistics that exceed t^*, in absolute value. The centered t-statistics are defined as $t^j = (\beta^j - mean(\beta^j))/se(\beta^j)$, where $mean(\beta^j)$ is the average estimate across iterations, and $se(\beta^j)$ is calculated as the standard error from the regression for iteration j. t^* is the t-statistic from the preferred estimate or the statistic that corresponds to the average estimate β across iterations, or $t^* = mean(\beta^j)/stdev(\beta^j)$.

[15]Violent crimes include aggravated assault, other assault, robbery and murder/manslaughter. Property crimes include burglary, theft, vehicle theft.

this overall composition of arrest type suggest that higher arrest type officers likely have a lower severity threshold for making an arrest relative to a lower arrest type officer. Newer work shows that lower level arrests may have limited value in reducing crime rates Cho et al. (2021), implying that lower severity arrest charges could have lower net welfare benefits than higher severity arrest charges.

In particular, Table 5 shows high arrest officers have lower shares of arrests for felony robbery, burglary, motor vehicle thefts, as well as misdemeanor simple assault, and criminal trespassing. High arrest officers have higher shares of arrests for a number of misdemeanor offenses, including burglaries of vehicles, misdemeanor theft, weapons violations, and driving under the influence (DUI). These officers also make relatively more arrests for lower level misdemeanor "quality of life" offenses, including drug offenses, resisting or evading arrest, and failing to provide ID to an officer.

These last two charges "resisting/evading arrest" and "fail to provide ID," are arrest charges that are both discretionary and indicate that an adversarial interaction occurred. An officer may arrest a civilian for "resisting arrest" if an interaction escalates and becomes combative and/or the officer perceives a lack of civilian compliance. Interactions may escalate because of officer or civilian actions (or both). A 1 standard deviation increase in arrest propensity corresponds to a 44% increase in the share of an officer's arrests that include a "resisting arrest" charge.

Consistent with the results for "resisting/evading arrest" charges, Rows (9) and (10) of Table 5 show that higher arrest propensity officers are significantly more likely to use *physical force* during the course of an arrest. I construct two definitions of use of force; the first is a use of force incident where an officer used physical force or the civilian was injured, and the second "Strict Use of Force" is an incident with the additional restrictions that the civilian did not resist and was not armed, and the officer was not injured. A 1 standard deviation increase in officer arrest propensity is associated with an 11-17% increase in officer use of force rates.

Next, I examine how arrest propensity relates to the race of civilians that an officer arrests. Officers with higher arrest propensities arrest Black and white civilians at slightly higher rates (Rows (12)-(14)). While the increase in the Black and white share of arrests are small in magnitude and comparable across groups, these estimates correspond to larger total increases in arrests of Black civilians, given both the higher baseline share of Black arrests and the higher total arrests for high arrest officers. A 1 standard deviation increase in officer arrest propensity from the average officer would result in 5.4 additional arrests of Black civilians and 2.5 additional arrests of white civilians. It is important to note that these results do not correspond to a test of taste-based racial bias, but rather represent racial disparities which could result from multiple factors including statistical discrimination and institutional discrimination. Nevertheless, the marginal arrest made by a higher arrest officer is more likely to involve a Black civilian, and therefore officers who make more arrests likely contribute to racial disparities in policing outcomes.

In Panels D and E of Table 5, I relate officer arrest propensity to the court outcomes for arrest charges. Dismissals and negotiated conviction plea deals are nearly always determined by prosecutor discretionary decisions, so court outcome measures largely reflect how a prosecutor views the quality of an arrest charge that has been made by an officer. Panel D shows that the outcomes for felony arrest charges do not differ by officer arrest propensity. Accordingly, while higher arrest type officers have a lower share of arrests with the most serious charges (Table A8), the court outcomes for these arrests do not differ from those of lower arrest type officers.

Panel E shows the share of misdemeanor arrest charges that result in different court outcomes. Higher arrest type officers are more likely to have their misdemeanor arrests result in a court charge or conviction, and are less likely to have their misdemeanor arrests result in a dismissal. Recent legal and sociological research has documented the fact that misdemeanor charges are often processed in a hasty and haphazard way, with court hearings that can conclude in no more than a few minutes (Natapoff, 2018). As a result, court outcomes for misdemeanor charges may be an imperfect metric of arrest quality relative to court outcomes for felony charges. Still, the fact that higher arrest officers have “better” court outcomes for their misdemeanor arrests suggests that these officers may be more effective or productive in completing these types of arrests.

Collectively, this analysis illustrates a mixed picture of the relationship between officer arrest propensity and arrest quality. Higher arrest officers are certainly more aggressive, as they are more likely to arrest civilians for “resisting/evading arrest” and they are more likely to use physical force against a civilian during an arrest. Higher arrest officers also appear to have a lower severity threshold for making an arrest, and a larger share of their arrests are for lower-level misdemeanor offenses. At the same time, higher arrest officers are also more likely to convert their misdemeanor arrests into court convictions, which is a potential metric of arrest quality or productivity.

4.3 Arrest Outcomes and Officer Demographics

What underlies the differences in officer arrest behavior and the associated differences in arrest outcomes across officers? The analysis thus far has shown that higher arrest officers are more likely to make arrests that have characteristics associated with lower severity, higher levels of aggression, and higher likelihood of court conviction. At the same time, it appears that white officers and less experienced officers tend to have higher arrest propensities.

In this section, I explore the relationship between officer demographic characteristics and arrest outcomes through a simple projection exercise. First, I obtain a projection of officer arrest propensity on to the demographic characteristics of officers using the training sample(s):

$$\hat{\theta}_{i,train} = \alpha_0 + \alpha_{race}X_{race,i} + \alpha_{exp}X_{exp,i} + \alpha_{other}X_{other,i} + r_i$$

where X_{race} includes whether an officer is Black, Hispanic/Latinx or other race (non-white), X_{exp} is years of experience and experience squared, and X_{other} includes all other officer demographic variables in Table 3 including gender, age, and trainee status. This exercise produces predicted components of $\hat{\theta}_{i,train}$ given these demographic variables of $\hat{\theta}_{race,i}$, $\hat{\theta}_{exp,i}$, $\hat{\theta}_{other,i}$, and an estimated residual $\hat{\theta}_{residual,i}$. I then regress the set of arrest outcomes examined in Section 4.2 from the test sample(s) on these projection components to investigate the relative importance of these different components. As in the above analysis, I present estimates from the distribution of 100 random partitions of dates in the sample which create 100 versions of the test and training sample.

Figure 3 and Appendix Figure A8 plot the results of this analysis for the primary set of arrest outcome characteristics from Table 5 and arrest offense type outcomes from Appendix Table A8. I present the results in percentage changes to make the estimates comparable across outcomes, and I display only the estimates that are significant at least at the 10% level for clarity. The black bars in these figures show the percentage increase in an arrest outcome characteristic from a 1 standard deviation increase in officer arrest propensity, corresponding to the final columns of Table 5 and Appendix Table A8. The colored bars in the figures show the impact of a 1 standard deviation in a projected component of $\hat{\theta}_i$ on an arrest rate outcome.

The officer race and officer experience projections are both largely consistent with the direction of the total effect of $\hat{\theta}_i$ on the arrest outcome characteristics. This consistency is strongest for misdemeanor arrest types of drug offenses, resisting/evading arrest, and failure to provide ID. The consistency in race and experience projection with the total effect is also strong for use of force outcomes and misdemeanor court conviction outcomes. The results thus imply that *both* officer race and officer experience partially explain the relationship between officer arrest propensity and arrest outcomes.

5 Conclusion

Individual police officers are critical to the outcomes of police work. This paper finds substantial variation in arrest behavior across officers responding to civilian-initiated 911 calls, even after controlling for detailed characteristics of call incidents. Analyzing high frequency data on calls for service in Dallas, Texas, I find that a 1 standard deviation increase in officer arrest propensity corresponds to a 40% increase in the likelihood of an arrest.

Higher arrest propensity officers differ from lower arrest propensity officers in the types of arrests they make. While higher and lower arrest propensity officers face comparable crime offending environments, higher arrest propensity officers are more likely to arrest individuals for lower level offenses and to use physical force during an arrest. At the same time, higher arrest propensity officers are more likely to have their misdemeanor arrests result in conviction, a potential metric of arrest productivity. These results suggest that higher arrest propensity officers are likely more aggressive and have a lower severity threshold for making arrests, though these lower level charges are also more likely to result in conviction.

In turn, officers who make more arrests are more likely to be white and have lower levels of experience. The findings suggest that interventions that either increase racial diversity of officers or accelerate officer learning and expertise could potentially reduce the dispersion in officer arrest behavior while also increasing the severity threshold of officers making arrests and reducing the incidence of use of force.

This project provides new evidence about the extent of discretion in law enforcement and that this discretion is related to indicators of policing quality. Future research should extend these findings to quantify the welfare costs and benefits of different types of arrests. Moreover, investments in reducing dispersion in officer behavior could have the potential to yield benefits in the form of increased trust in law enforcement and equal access to police protection services. Future work should also assess the costs and benefits of different policy interventions that may be used to increase uniformity in officer behavior, including additional police training, monitoring procedures, mentorship programs, and targeted hiring and firing of officers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

I am grateful to professors Sandra Black, Jason Abrevaya, Leigh Linden, Becky Pettit, Rich Murphy, Steve Trejo and Brendan Kline for their helpful comments and feedback. I also thank members of the Dallas Police Department that shared information about their operations, protocols and data practices for this project. This project received funding from the University of Texas Continuing Graduate Fellowship. The views presented in this paper are my own and do not necessarily reflect the views of the Dallas Police Department.

References

Aaronson D, Barrow L, and Sander W (2007). Teachers and student achievement in the Chicago Public High Schools. Journal of Labor Economics 25(1), 95–135.

Aizer A and Doyle J (2015). Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly Assigned Judges. Quarterly Journal of Economics 130(2), 759–804.

Anbarci N and Lee J (2014). Detecting racial bias in speed discounting: Evidence from speeding tickets in Boston. International Review of Law and Economics 38, 11–24.

Antonovics K and Knight BG (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.

Bhuller M, Dahl G, Loken K, and Mogstad M (2020). Incarceration, Recidivism, and Employment. Journal of Political Economy 128 (4).

BlueLivesMatter (2016, July). Opinion: Why do cops always support other officers? https://bluelivesmatter.blue/.

Braga A, Sousa W, James Coldren J, and Rodriguez D (2018). The Effects of Body-Worn Cameras on Police Activity and Police-Citizen Encounters: A Randomized Controlled Trial. Criminology 108(3), 511–538.

Brayne S (2017). Big data surveillance: The case of policing. American Sociological Review 82(5), 977–1008. [PubMed: 38322733]

Card D, Cardoso AR, Heining J, and Kline P (2018). Firms and Labor Market Inequality: Evidence and Some Theory. Journal of Labor Economics 36(51), 13–70.

Census (2017). American community survey: Demographic and housing estimates, dallas, tx. Technical report, Census Bureau, United States.

Chetty R, Friedman JN, and Rockoff JE (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. American Economic Review 104 (9), 2593–2632.

Cho S, Gonçalves F, and Weisburst E (2021). Do Police Make Too Many Arrests? Working Paper.

Deangelo G and Owens E (2017). Learning the ropes: General experience, task-Specific experience, and the output of police officers. Journal of Economic Behavior & Organization 142, 368–377.

Department of Justice (2017, March). Supporting Federal, State, Local and Tribal Law Enforcement. Memorandum, Office of the Attorney General, Department of Justice, United States.

Dobbie W, Goldin J, and Yang C (2018). The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges. American Economic Review 108(2), 201–240.

Fagan J, Braga A, Brunson R, and Pattavina A (2016). Stops and stares: Street stops, surveillance, and race in the new policing. Fordham Urban Law Journal 43(3), 539–614.

FBI (2018). Crime in the United States: Offenses Known to Law Enforcement, 2017. Technical report, Federal Bureau of Investigation, Department of Justice, United States.

Frydl K and Skogan W (2004). Fairness and Effectiveness in Policing: The Evidence. National Academies Press.

Gelman A, Fagan J, and Kiss A (2007). An analysis of the New York City police department's "stop-and-frisk" policy in the context of claims of racial bias. Journal of the American Statistical Association 102(479), 813–823.

Gonçalves F and Mello S (2021). Racial bias in policing: Evidence from speeding tickets. American Economic Review 111 (5).

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.

Guarino C, Maxfield M, Reckase M, Thompson P, and Wooldridge J (2015). An evaluation of Empirical Bayes' estimation of value-added teacher performance measures. Journal of Educational and Behavioral Statistics 40 (2), 190–222.

Guell M, Rodriguez-Mora J, and Telmer C (2015). The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating. Review of Economic Studies 82 (693-735).

Hoekstra M and Sloan C (2022). Does Race Matter for Police Use of Force? Evidence from 911 Calls. American Economic Review (Forthcoming).

Horrace WC and Rohlin SM (2016). How dark is dark? bright lights, big city, racial profiling. The Review of Economics and Statistics 98(2), 226–232.

Joh E (2016). Surveillance discretion: Automated suspicion, big data, and policing. Harvard Law and Policy Review 10, 15–42.

Kane T and Staiger D (2008). Estimating teacher impacts on student achievement: An experimental evaluation. National Bureau of Economic Research, NBER Working Paper(14607).

Knowles J, Persico N, and Todd P (2001). Racial bias in motor vehicle searches: Theory and evidence. Journal of Political Economy 109(1), 203–229.

Knox D, Lowe W, and Mummolo J (2020). Administrative Records Mask Racially Biased Policing. American Political Science Review.

Koedel C, Mihaly K, and Rockoff JE (2015). Value-added modeling: A review. Economics of Education Review 47, 180–195.

Lipsky M (1980). Street-Level Bureaucracy. New York, NY: Russell Sage Foundation.

Mastrofski S (2004). Controlling street-level police discretion. The Annals of the American Academy of Political and Social Science 593, 100–118.

Morris C (1983). Parametric Empirical Bayes inference: Theory and applications. Journal of the American Statistical Association 78(381), 47–55.

Natapoff A (2018). Punishment without Crime: How our Massive Misdemeanor System Traps the Innocent and Makes America More Unequal (First Edition ed.). New York, NY: Basic Books.

Nickels E (2007). A note on the status of discretion in police research. Journal of Criminal Justice 35, 570–578.

Reiss A (1971). The Police and the Public. New Haven, CT: Yale University Press.

Ridgeway G (2018). Policing in the Era of Big Data. Annual Review of Criminology 1, 401–419.

Ridgeway G and MacDonald J (2009). Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops. Journal of the American Statistical Association 104 (486), 661–668.

Spelman W and Brown D (1984). Calling the police: Citizen reporting of serious crime. National Institute of Justice Report, U.S. Department of Justice.

Walker S (1993). Taming the System: The Control of Discretion in Criminal Justice, 1950-1990. Oxford University Press.

Weisburd S (2019). Police presence, rapid response rates, and crime prevention. Working Paper.

West J (2018). Racial bias in police investigations. Working Paper.

West J (2019). Learning by doing in law enforcement. Working Paper.

Author Manuscript
Author Manuscript
Author Manuscript

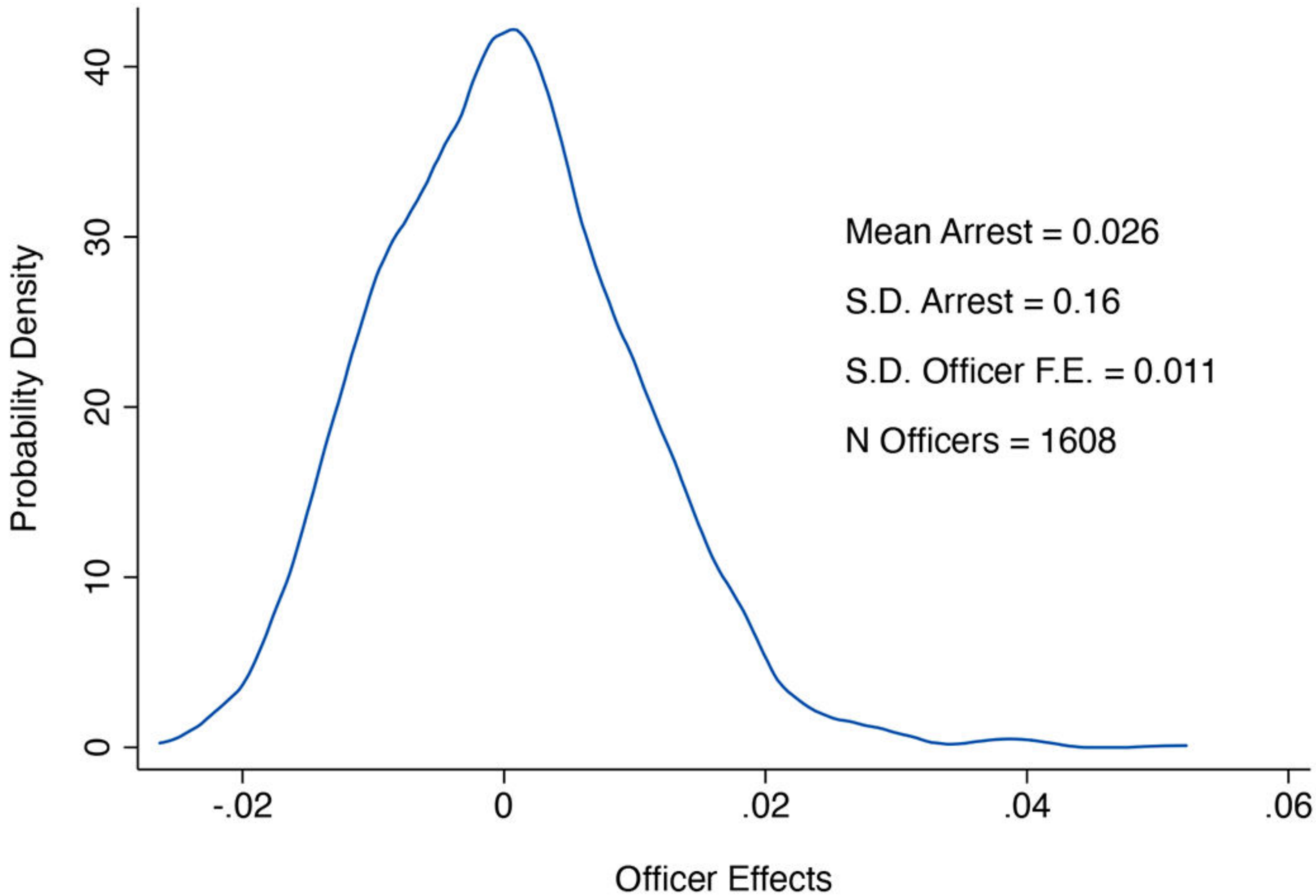

Figure 1:
Dispersion in Officer Arrest Propensity
The figure graphs the distribution of the estimated Officer Effects, $\hat{\theta}_i$, measured using the arrest outcome model on the analysis sample.

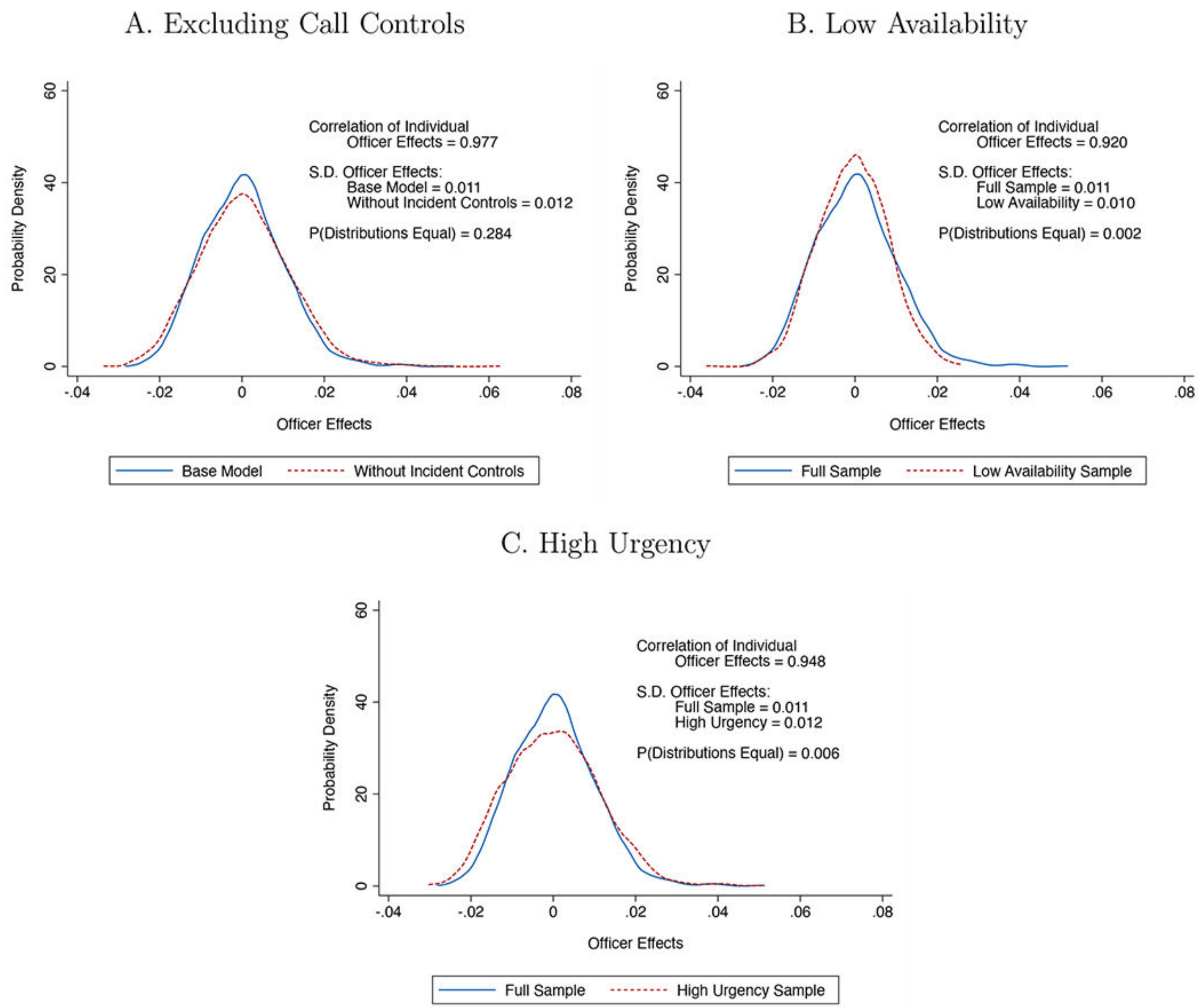

Figure 2:

Tests of the Importance of Officer Sorting to Officer Effect Distribution

These figures plot the distributions of estimates included in Table 4. The “Without Controls” model is estimating excluding any controls associated with an individual call, that are not determined at the beginning of the shift. This model includes the number of other police officers in the responding patrol car, shift effects, and assigned home sector of the officer. It excludes police beat fixed effects for the call location as well as all other call characteristics, X_{kt} and φ_g. The “Low Availability’ sub-sample is the set of observations where a greater proportion of officers are unavailable because they are responding to other call incidents at the time a call is dispatched, split at the median. The “High Urgency” sub-sample is the set of observations with a shorter time between when a call is received and when it is dispatched, split at the median. The “Correlation of Individual Officer Effects” measures the correlation between officer fixed effects in each sample to the base sample. “P(Distribution

Equal)" is the p-value from a Kolmogorov-Smirnov test of equality between officer fixed effects in each sample and the base sample.

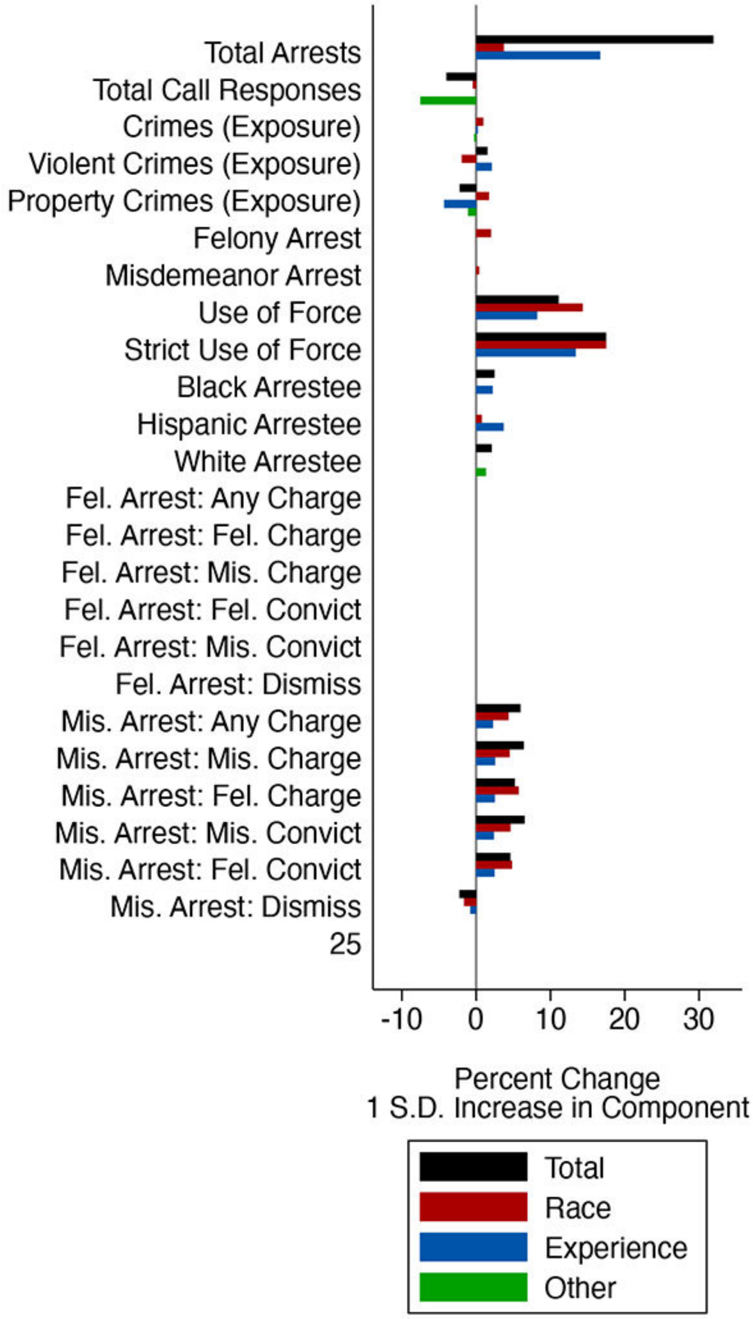

Figure 3:

Officer Effect Demographic Components and Arrest Outcome Characteristics

This figure decomposes the relationship between officer effects and arrest outcomes into the contribution of observable demographics of officers. The outcomes mirror those in Table 5. As in Table 5, estimates are averages from the distribution of 100 random partitions of dates to form a test and training sample. Within each training sample, officer effects are projected onto demographic characteristics (race, experience variables, and other demographics) and these projections are then included as the key controls for each test sample outcome.

Percent changes refer to the mean across iterations and display the impact of a one standard deviation increase in the projected component. P-values are determined from the density of t-statistics relative to the average estimate. For legibility, only effects significant at the 10% level are displayed. Black bars show the total estimated effect of a one standard deviation increase in officer effects on arrest outcomes.

Table 1:

Summary Statistics

	Mean	S.D.		Mean	S.D.
Total Officers	1608				
Total Call Responses	1901197				
Total Call-Officer Records	3289249				
Officer Characteristics			**Arrests**		
Trainee	0.070	(0.214)	Arrest	0.026	(0.160)
Sergeant	0.012	(0.094)	Black Arrestee	0.011	(0.106)
Salary (1000s)	57.48	(9.48)	Hispanic Arrestee	0.006	(0.080)
Years Experience	11.79	(7.71)	White Arrestee	0.005	(0.074)
Age	37.44	(8.66)			
Female	0.144	(0.289)			
Black	0.264	(0.393)			
Hispanic	0.212	(0.340)			
White	0.480	(0.429)			
Dispatch Code Type			**Call Characteristics**		
Assault	0.002	(0.039)	Minutes to Dispatch	34.35	(53.48)
Armed Encounter	0.007	(0.085)	Unavailable Rate	0.320	(0.162)
Robbery	0.013	(0.115)	# Officers Responding	1.730	(0.635)
Burglary - Business	0.040	(0.195)	High Priority	0.491	(0.491)
Burglary - Vehicle	0.032	(0.176)	Low Priority	0.509	(0.509)
Burglary - Residence	0.056	(0.230)	Common Location	0.053	(0.223)
Unauth. Use of Vehicle	0.019	(0.138)	Same Sector Prior Call	0.238	(0.426)
Theft	0.023	(0.150)	Assigned Sector	0.266	(0.442)
Criminal Mischief	0.108	(0.310)	Day (8am-4pm)	0.326	(0.469)
Major Disturbance	0.239	(0.426)	Evening (4pm-12am)	0.429	(0.495)
Injured Person	0.008	(0.091)	Overnight (12am-8am)	0.245	(0.430)
Accident	0.127	(0.333)			
Other	0.326	(0.469)			

This table displays summary statistics of the data used in analysis. All statistics are calculated at the call level. In cases where more than one officer is dispatched to a call, officer characteristics are averaged at the call level. The sample is restricted to exclude calls that are unlikely to be routine patrol responses; calls that are dispatched over 5 hours after they are made, as well as calls that include responding officers that are unlikely to be patrol officers, or calls with officers who have fewer than 1000 responses in the raw data. Table A2 compares this sample with the raw data.

Table 2:

Model Variance Decomposition

Full Model Variance - Arrest Outcome	0.0257			
Share Explained - R^2	2.63%			
Share Explained - Adjusted R^2	2.33%			
A. Ensemble Decomposition Covariance(Element,Y)	**% Total Variance**	**% Explained Variance**		
Officer Effects	0.54%	20.34%		
Geography Effects	0.29%	11.09%		
Shift Effects	0.56%	21.22%		
Call Characteristics	1.24%	47.22%		
B. Incremental R^2	**% Total Variance (R^2)**	**% Explained Variance (R^2)**	**% Total Variance (Adjusted R^2)**	**% Explained Variance (Adjusted R^2)**
Officer Effects	0.42%	15.84%	0.37%	17.88%
Geography Effects	0.26%	9.74%	0.25%	11.00%
Shift Effects	0.52%	19.71%	0.27%	22.26%
Call Characteristics	1.11%	42.07%	1.11%	47.51%
C. Correlation of Components	**Officer Effects**	**Geography Effects**	**Shift Effects**	**Call Characteristics**
Officer Effects	1.0000			
Geography Effects	0.0176	1.0000		
Shift Effects	−0.0063	0.0140	1.0000	
Call Characteristics	0.0486	0.0043	0.0009	1.0000

This table decomposes the variance of model components for the preferred specification. Panel A presents the ensemble variance decomposition proposed by Card et al. (2018), where each component is the covariance of a model element with the outcome. Each covariance is estimated at the call by officer observation level. Panel B shows the incremental R^2 resulting from adding the element or component to the model, defined as $R^2_{total} - R^2_{-j}$ for each component j. Panel C shows the correlation of the estimated components or coefficients of the model, at the call by officer observation level. In each section of this table, unadjusted officer fixed effects are used to calculate variance components and correlations to other components of the model at the call by officer observation level. Geography Effects includes police beat location of a call. Shift Effects include the time varying shift fixed effects as well as assigned officer police sector fixed effects. Call characteristics includes all other controls in the model.

Table 3:

Officer Effects and Officer Demographics

Outcome: Officer Effect	
Black	−0.0034*** (0.0005)
Hispanic	−0.0013** (0.0006)
Other Race	−0.0015 (0.0010)
Female	0.0002 (0.0005)
Sergeant	0.0067*** (0.0018)
Trainee	0.0008 (0.0007)
Age	−0.00001 (0.00005)
Experience	−0.0013*** (0.0001)
Experience Squared	0.00003*** (0.00000)
Observations	1599
R-Squared	0.222
Arrest Mean	0.026
Arrest S.D.	0.160

** p<0.001

** p<0.01

* p<0.05

+ p<0.1

This table shows regression results of Officer Effects, $\hat{\theta}_i$, regressed on fixed officer characteristics, at the officer level. Robust standard errors are in parentheses. White officers are the omitted race category. Officers without demographic information are excluded from the regressions.

Table 4:

Tests of the Importance of Sorting to the Officer Effect Distribution

	(1) Base	(2) Without Controls	(3) Low Availability	(4) High Urgency
Total Officers	1608	1608	1608	1608
Total Calls	1901197	1901197	928901	934340
Total Call-Officer Records	3289249	3289249	1609707	1754948
Arrest Mean	0.026	0.026	0.025	0.034
Arrest S.D.	0.160	0.160	0.157	0.181
Distribution Officer Effects				
S.D. of Officer Effects	0.011	0.012	0.010	0.012
% Change: 1 S.D. Increase in Officer Effects	41.0%	45.2%	37.8%	36.6%
% Total Variance	0.53%	0.59%	0.54%	0.51%
Comparison Officer Effects				
Correlation: Base Sample Officer Effects	-	0.977	0.920	0.948
Distribution Equality: P-Value	-	0.284	0.002	0.006
Additional Indicators				
Joint F-Test Controls: Arrest Outcome	43.29			
Joint F-Test Controls: Officer Effect Outcome	3.17			
Correlation to Number of Calls	−0.115			

This table summarizes the main analysis arrest results and tests of officer sorting. The "Without Controls" model is estimating excluding any controls associated with an individual call, that are not determined at the beginning of the shift. This model includes the number of other police officers in the responding patrol car, shift effects, and assigned home sector of the officer. It excludes police beat fixed effects for the call location as well as all other call characteristics, X_{kt} and ϕ_g. The "Low Availability' sub-sample is the set of observations where a greater proportion of officers are unavailable because they are responding to other call incidents at the time a call is dispatched, split at the median. The "High Urgency" sub-sample is the set of observations with a shorter time between when a call is received and when it is dispatched, split at the median. The "Correlation: Base Sample Officer Effects" measures the correlation between officer fixed effects in each model to Column (1). The "Distribution Equality: P-Value" is the p-value from a Kolmogorov-Smirnov test of equality between officer fixed effects in each model and those from Column (1). The "Joint F-test Controls" measures the combined significance of individual call controls that could affect an officer's decision to respond to a call, or the same set of controls omitted in Column (2). The F-tests are clustered at the level of the officer and call response. The observation count across the samples excludes singleton observations dropped from the model.

Table 5:

Officer Effects and Arrest Outcome Characteristics

	Observations	Mean	Officer Effect β	S.E.	% Change: 1 S.D. Officer Effects
A. Baseline Characteristics					
(1) Officer Effect, Test Sample	1608	−0.001	0.789***	(0.031)	-
(2) Total Arrests	1608	35.0	1100.9***	(37.9)	31.91%
(3) Total Call Responses	1608	1217.9	−4796.9***	(590.0)	−4.00%
(4) Crimes (Exposure)	1608	56.5	9.2	(19.6)	0.16%
(5) Violent Crimes (Exposure)	1608	1.1	1.7**	(0.9)	1.50%
(6) Property Crimes (Exposure)	1608	9.7	−21.4***	(3.2)	−2.23%
B. Arrest Characteristics					
(7) Felony Arrest	49262	0.112	−0.045	(0.141)	−0.40%
(8) Misdemeanor Arrest	49262	0.569	0.234	(0.254)	0.42%
(9) Use of Force	49262	0.013	0.140***	(0.048)	11.10%
(10) Strict Use of Force	49262	0.009	0.158***	(0.042)	17.48%
C. Race Share of Arrests					
(11) Black Arrestee	49262	0.439	1.065***	(0.202)	2.46%
(12) Hispanic Arrestee	49262	0.198	0.258	(0.201)	1.31%
(13) White Arrestee	49262	0.208	0.432***	(0.193)	2.10%
D. Court Outcomes: Felony Arrests					
(14) Any Charge	4563	0.535	0.412	(0.845)	0.78%
(15) Felony Charge	4563	0.461	−0.058	(0.807)	−0.13%
(16) Misdemeanor (*Down*) Charge	4563	0.173	0.969	(0.677)	5.69%
(17) Felony Convict	4563	0.439	0.080	(0.791)	0.18%
(18) Misdemeanor (*Down*) Convict	4563	0.143	0.985	(0.635)	6.96%
(19) Dismiss	4563	0.492	−0.525	(0.838)	−1.08%
E. Court Outcomes: Misdemeanor Arrests					
(20) Any Charge	39533	0.306	1.799***	(0.233)	5.96%
(21) Misdemeanor Charge	39533	0.235	1.480***	(0.218)	6.40%
(22) Felony (*Up*) Charge	39533	0.103	0.529***	(0.148)	5.21%
(23) Misdemeanor Convict	39533	0.199	1 274***	(0.211)	6.51%
(24) Felony (*Up*) Convict	39533	0.098	0.441***	(0.156)	4.59%
(25) Dismiss	39533	0.728	−1.615***	(0.230)	−2.25%

** p<0.001

** p<0.01

* p<0.05

+ p<0.1

This table displays the relationship between arrest outcomes (originating from call responses), in a test sample, and officer effects, from a training sample. 100 iterations of test and training samples are determined by randomly partitioning dates in the sample. β and S.E. are the mean and standard deviation of iteration estimates. Outcome means, observation counts, and percent changes refer to the mean across iterations. P-values are determined from the density of t-statistics relative to the average estimate. Panel (A) shows officer-level bivariate regressions of outcomes on officer effects. Crime exposure is the daily division average number of crimes on days when an officer is working, excluding crime reports resulting from call responses of the focal officer. Panels (B) - (E) measure arrest-level outcomes; and include the full set of controls from the baseline model. Officer use of force is a non-shooting incident, where "strict" refers to incidents where an unarmed civilian is injured and the officer is not injured. Panels (D) and (E) measure court charge, conviction, and dismissal outcomes among felony and misdemeanor arrests, respectively. Dismissals are either dropped/non-conviction charges, or charges that do not have a court record. An arrest may result in multiple convictions and/or dismissal.