# The Effect of Predictive Policing on Citizen Perceptions of Procedural Justice at the Example of New York City

Kerber, Vera

Thesis

# The Effect of Predictive Policing on Citizen Perceptions of Procedural Justice at the Example of New York City

Name: Vera Kerber

Course: MSc Public Administration

Track: Economics and Governance

Student Number: s2567032

Word Count: 18,620

Thesis Advisor: Dr Andrei Poama

## Abstract

When discussing predictive policing, the focus usually lies on the method's potential effect on policing outcomes. In contrast, this paper aims to add to the literature by providing an analysis of the potential effect of predictive software on the process of policing. At the centre is the question: "How does predictive policing influence citizen perceptions of procedural justice during police encounters?". For this purpose, citizen perceptions of procedural justice are operationalised as complaints and allegations against officers of the New York City Police Department (NYPD). In addition to a paired t-test, a random-effects model is used, to test for the overall effect of predictive policing and the influence of a person's race. The seven years before and after the introduction of predictive policing in 2013 thereby serve as the main independent variable. The results of the analysis do not reveal a clear negative correlation between predictive policing and citizen perceptions of procedural justice. Overall, encounters perceived as negative were less likely during the second period. However, an increase in the severity of individual complaints indicates partial support for the hypotheses derived from the literature. In addition, regarding the effect of race, the likelihood to perceive abuses of power increased disproportionately for Black and Hispanic citizens.

**Keywords:** predictive policing, procedural justice, fairness, criminal justice, New York City

**Content**

## Figures



## Tables

## Abbreviations

| | |
|---|---|
| ACLU | American Civil Liberties Union |
| ADF | Augmented Dickey-Fuller |
| ADS | Automated Decision Systems |
| BLM | Black Lives Matter |
| BPoC | Black and People of Colour |
| CCOPS | Community Control Over Police Surveillance |
| DAS | Domain Awareness System |
| FADO | Force, Abuse of authority, Discourtesy, and Offensive language |
| FBI | Federal Bureau of Investigation |
| FOIL | Freedom of Information Law |
| GLS | Generalised Least Squares |
| LAPD | Los Angeles Police Department |
| LPR | License Plate Recognition |
| NAACP | National Association for the Advancement of Coloured People |
| NYCLU | New York Civil Liberties Union |
| NYPD | New York Police Department |
| NYU | New York University |
| OLS | Ordinary Least Squares |
| SD | Standard Deviation |
| SE | Standard Error |
| US | United States |

## Introduction

After the killing of George Floyd in May 2020, protests against systemic racism and police violence swept the United States (US) and soon spread worldwide (Hill & Stein, 2020). The incident reignited the conversation about the power vested in police officers and the role of policing in society more broadly. One aspect of policing that has attracted scrutiny in recent years is the use of Automated Decision Systems (ADS) to predict spatial as well as individual patterns of criminality. It is part of a wider trend of algorithmic criminal justice, which entails the use of big data, machine learning and algorithms by public entities (Harcourt, 2007). One of the most well-known examples of ADS in law enforcement is so-called predictive policing. The use of algorithmic software to predict the area of a future crime or to identify potential perpetrators or victims (AI Now Institute, 2018). Besides spatial-temporal assessments of crime probability, algorithms can use criteria such as cost and social justice to distribute police activity across a territory (Ferguson, 2017).

Proponents of the technology argue that crime forecasting can anticipate future risk and prevent crime by allocating police resources accordingly. Proactive deterrence or interference is supposed to contribute to lower crime and higher clearance rates. (Beam, 2011) However, civil rights organisations such as the American Civil Liberties Union (ACLU) warn that the introduction of algorithms into police work does not alleviate discriminatory policing practices but on the contrary exacerbates them (Brantingham et al., 2018). Academics likewise voice their concern. A group of over 1,400 scientists even went as far as to call on their colleagues to stop collaborations with police departments on predictive algorithms (Castelvecchi, 2020).

Research on potential benefits as well as drawbacks of predictive policing is often inconclusive and lacks a systematic approach (Meijer & Wessels, 2019). Regarding the benefits, researchers focus on whether predictive algorithms could play a role in crime prevention and enhancing police effectiveness. However, the empirical evidence is mixed. While some studies found an improvement in effectivity, others could not identify a positive effect (Levine et al., 2017; Saunders et al., 2016). The suspected drawbacks similarly lack empirical evidence, to discern how exactly predictive policing methods negatively affect policing. As such both the claimed benefits and drawbacks of predictive policing lack empirical foundation and are mainly based on anecdotes and compelling argumentation (Meijer & Wessels, 2019). While previous research is mostly focused on outcomes, the paper at hand contributes to an understanding of how predictive policing could affect police processes. Specifically, regarding the aspect of perceived fairness. It aims to answer the question “How does predictive policing influence citizen perceptions of procedural justice during police encounters?”.

The focus of the research is Tyler’s (1988) theory of citizen perceptions of procedural justice in the context of policing. The basis of the analysis is four key components of procedural justice as identified by Tyler (2004). They are participation, neutrality, dignity and respect, and trust. How fair citizens consider a process to be has a strong influence on their view of police legitimacy, which in turn influences people’s willingness to cooperate with the police (Bolger & Walters, 2019; Tyler, 2004). Thus, it is valuable to assess whether citizens’ views on procedural justice are affected by predictive policing and could therefore influence police legitimacy and citizen cooperation. Several studies have shown that among the most powerful predictors of citizen perceptions of procedural justice is direct contact with law enforcement (Cao et al., 1996; Frank et al., 2005; Rosenbaum et al., 2005; Schafer et al., 2003). Thereby are negative experiences much more impactful on perceptions of fairness than positive ones (Dai et al., 2011; Weitzer et al., 2008). As such this paper operationalises citizen perceptions of procedural justice as complaints made to the NYPD’s oversight agency, the Civilian Complaint Review Board (CCRB). However, potential positive effects of predictive policing are not captured by the analysis since complaints only depict negative experiences. Thus a more qualitative approach is needed to capture the full effect of predictive policing on citizen perceptions of police procedural justice. However, because of the strong impact of negative encounters, it remains a strong indicator for diminished perceptions of procedural justice.

The research focuses on the NYPD’s Domain Awareness System (DAS) and its predictive models that were implemented across the police force in 2013. The observed period is split into two parts, before the introduction of predictive policing and after. First, a paired t-test is conducted to assess the effect on overall perceptions of procedural justice. Subsequently, a random-effects model with Generalised Least Squares (GLS) estimator is used to analyse potential differences in perception across racial groups. The goal is to capture the effect of predictive policing on the four key components of citizen perceptions of procedural justice as defined by Tyler (2004). While the assumed impact on all factors cannot be clearly defined, the anticipated negative influence on participation, as well as dignity and respect, leads to the stipulation of an overall detrimental effect. Furthermore, previous research suggests, that differing perceptions of procedural justice across racial groups could be a factor, especially concerning the aspect of neutrality (Tyler & Huo, 2002). Hence, the inclusion of an assessment of the impact depending on a person’s race.

The thesis is structured in six parts. First, a short overview of the history and the actors involved in predictive policing provides context. Second, the theoretical framework explains the concepts the research is based on and links theory on procedural justice to predictive policing. Third, the research design of the paper including the underlying hypotheses are introduced. Part of this section is the operationalisation of the concepts established in the theoretical framework and a description of the data. Fourth, the empirical findings of the regression analysis are presented. Then follows an analysis of the research results and a discussion of the findings’ limitations. Lastly, the conclusion places the results into the context of the wider body of literature and puts forward suggestions for further avenues of research as well as policymaking.

## Context: predictive policing in the United States

When predictive policing first came to the American public's attention, it was lauded by the media as the future of law enforcement, a way for police officers to predict the future and to stop crimes before they occurred. In several breathless articles, reporters regurgitated the spin, perpetuated by companies such as PredPol, that predictive policing would fundamentally improve the quality of police work and eliminate human bias. (Beam, 2011; Goode, 2011; Orr, 2012) In 2011 predictive policing even made the list of TIME's (2011) "50 Best Inventions". This initial uncritical praise can partly be attributed to a lack of independent research on the effectiveness of predictive policing. The only information available was studies conducted by the police departments themselves, usually in cooperation with the respective software maker. Accordingly, the first rollouts of predictive policing reported a 25% decrease in burglaries in Los Angeles and a drop in property crimes across California (Ferguson, 2017). However, follow-up studies were much less conclusive. One of the few independent studies, conducted by the RAND Corporation, found no significant reduction in crime rates in districts with predictive policing as opposed to those using regular hotspot mapping methods (Perry et al., 2013). As Lum and Isaac (2016) put it, predictive policing "is predicting future policing, not future crime" (p. 16).

Another aspect that contributed to the lure of predictive policing was the promise of objectivity and cost-effectiveness. After the 2008 recession police departments across the country faced budget cuts. Large federal grants were invested in developing smart policing strategies and streamlining law enforcement. The initiative was spearheaded by Police Chief William Bratton, who had pioneered data-driven policing in the NYPD and brought his acumen to the Los Angeles Police Department (LAPD). The involvement of Bratton, who enjoyed high prestige in law enforcement circles, contributed to the nationwide adoption of predictive policing technology. Professors from the University of California, who were involved in the initial trials with the LAPD, later formed PredPol (Winston, 2018). This further cemented the image of the technology as rooted in scientific advancement and quantifiable findings. A framing that PredPol and other companies heavily relied on for marketing purposes. At this early stage concerns about social justice or fairness found little acknowledgement. (Ferguson, 2017)

At first, statements made about more technologically advanced versions of predictive policing were similarly glowing as the ones about previous iterations. However, with the introduction of more personalised prediction models, public concern grew. One case that drew attention to the issue was New Orleans's cooperation with Palantir. The secrecy surrounding the use of predictive policing by the city's police force and the incredibly widespread collection of private data accompanying the efforts caused public outrage (Winston, 2018). A precursor to this and similar scandals was a generally enhanced scrutiny of the role of police. In 2014, following the shooting of Michael Brown by a Ferguson police officer, protests erupted and the Black Lives Matter (BLM) movement gained momentum (Ferguson, 2017). The ACLU and other civil rights groups pressed for more accountability in policing and public opinion shifted, not least because media began to depict a more nuanced picture of the technology (Bond-Graham, 2014; Haskins, 2019; Newcombe, 2014).

Los Angeles, one of the first cities to introduce predictive policing in the late 2000s, announced changes to its use of PredPol's software in October 2019 (Sturgill, 2020). This included the creation of a novel policing unit with the task to analyse current policing strategies and to seek feedback from community groups before introducing new software. The changes were made after an investigation by the inspector general failed to find evidence that the LAPD's predictive policing programme reduced crime rates. Confronted with the decision PredPol's founder Jeff Brantingham stated his belief that the right algorithms can serve to reduce racial bias in policing. Instead, Brantingham makes the introduction of the officers' human judgement into the process responsible for any discrimination that might occur. (Sturgill, 2020)

Initially predictive policing allowed policy-makers to avoid difficult questions around bias in policing. It was used as a tool to legitimise police work and to deflect criticism for discriminatory practices (Ferguson, 2017). For politicians, the new technology offered a chance to proclaim a new, more just era in policing. Once the narrative was no longer solely controlled by software developers and doubts about the effectiveness of the technology emerged, a path to regulation opened. In some cases, such as in Los Angeles in 2019, the police departments themselves expressed doubt about the use of the software. It did not depict the promised revolution in crime-fighting and the envisioned cost reductions failed to materialise (Puente, 2019). This facilitated the work of activists, such as the ACLU's (2020) Community Control Over Police Surveillance (CCOPS) campaign. The goal of the ACLU's efforts is to pass laws that give residents veto power over the technology that police use in their community. So far 15 American jurisdictions have passed CCOPS laws. In 2020 New York City adopted a CCOPS law, forcing the US' largest police force to disclose the type of surveillance technology they are using and the data they are collecting from citizens (Feiner, 2020).

## Theoretical framework

In this section concepts by Ferguson (2018), Perry et al. (2013) and others are used to give an overview of the different approaches to defining predictive policing. Then the capabilities of the NYPD's predictive tools are explained using Griffard (2019) and Levine et al. (2017). Finally, a broad definition introduced by Meijer and Wessels (2019) is deemed most fitting for this paper. In the second section, different approaches to citizen perceptions procedural justice are described based on Thibaut and Walker (1975), Leventhal (1980) and Tyler (1988) and put into the context of policing using Tyler (2004). Lastly, the different aspects of citizen perceptions of procedural justice are applied to predictive policing. To evaluate the potential benefits and drawbacks of predictive policing in regard to procedural justice articles by Harcourt (2007), Lum and Isaac (2016), Selbst (2017) and Ferguson (2018) and others are essential. Central to the assessment of the NYPD's predictive algorithms and their potential consequences are articles by Levine et al. (2017), Chohlas-Wood and Levine (2019) and Griffard (2019).

### Predictive policing

Predictive policing lacks a uniform definition throughout the literature. The most salient distinction can be made between models that focus on spatial-temporal predictions and those aiming for individual predictions. While the former is concerned with quantitative analysis to predict the time and place of possible future criminal activity (Norton, 2013; Uchida, 2009), the latter has the goal to identify specific offenders (Chan & Bennett Moses, 2016; Perry et al., 2013; Tayebi & Glässer, 2016).

According to Ferguson (2017), this distinction developed gradually due to a more in-depth level of analysis facilitated by technological advancement and new ways to harness data. He divides predictive policing into three stages that represent an evolution in technology and intervention level. First emerged predictive policing 1.0, inspired by the spatial models used to forecast earthquakes. It identifies areas in which certain crimes, namely property-based offences such as burglaries, are most likely to occur. The software uses historical crime data, based on the empirical finding of a near-repeat effect. Accordingly, crime in a certain area triggers more crime in the same neighbourhood. The next generation, predictive policing 2.0, is still focused on place, but instead of property crimes, it targets violent crimes such as robberies or shootings. The logic remains the same, geographic factors make some areas more vulnerable and therefore require increased policing. The most recent version is predictive policing 3.0. Instead of hot spots, technology is now used for pattern-matching to identify individual suspects. The software creates predictive profiles of individuals that are deemed likely to engage in criminal activity based on various factors, including their social network.

To define the NYPD's predictive algorithm, it is necessary to understand its various features. One of the NYPD's first forays into analytics was CompStat (COMPare STATistics), which was introduced in 1994 and is still in use. The programme compiles information on the time, place, and victims of crimes across New York City. Nowadays CompStat enables analysts to plot crime reports and to identify geographic hotspots to which to send officers (Griffard, 2019). In addition, the NYPD launched its Domain Awareness System (DAS) in 2008, a cooperation with Microsoft. According to the NYPD, the system comprises "a network of sensors, databases, devices, software, and infrastructure" (Levine et al., 2017, p. 1). At first, the system was only used by the NYPD's counterterrorism bureau and accessed sensor data such as security cameras, license plate recognition (LPR) or radiation. In 2010 geocoded NYPD records, e. g. arrests, warrants and emergency calls were added to the DAS database. Three years later, in 2013, DAS was made available for general policing and thus every precinct in New York City. The city-wide rollout of DAS included the incorporation of predictive algorithms to help police with resource allocation. (Levine et al., 2017) From June 2015 to April 2016 DAS mobile was implemented across the police force, enabling every officer to access the system from their phone or tablet. This allows law enforcement to access analytics in real-time and makes DAS the largest digital surveillance system in the world. (Ungerleider, 2012). The most recent addition to the NYPD's predictive tools is Patternizr, which was added in December 2016. It analyses law enforcement data to establish criminal patterns by correlating data points across crimes or offenders. (Griffard, 2019)

The NYPD's predictive software developed over time and new features were added piecemeal. Therefore, a definition is needed that captures the wide range of the NYPD's algorithmic capabilities. To date, Meijer and Wessels (2019) introduced the most expansive description of the technology. Based on a literature review of 37 publications they propose the following: "Predictive policing is the collection and analysis of data about previous crimes for identification and statistical prediction of individuals or geospatial areas with an increased probability of criminal activity to help developing policing intervention and prevention strategies and tactics." (Meijer & Wessels, 2019, p. 1033). Because this definition captures both spatial-temporal and individual-level analysis, it is used throughout the paper.

## Citizen perceptions of procedural justice

The term procedural justice to describe the fairness of a process was first coined by Thibaut and Walker (1975). Previous research mainly focused on the fairness of outcomes, i.e. distributive justice. They examined two types of control that a person can exert within the court system: process control and decision control. Thibaut and Walker (1975) found a positive relationship between disputant control and procedural fairness and a negative relationship between third-party control and procedural fairness. Meaning, the more control a person has over the court proceedings the fairer they perceived the trial to be. Leventhal (1980) built on Thibaut and Walker's (1975) research to evaluate how citizens judge the fairness of decision-making processes. For this purpose, he developed six criteria: consistency, absence of bias, decision quality, correctability, representation, and ethicality.

Tyler (1988) expanded on Thibaut and Walker's (1975) as well as Leventhal's (1980) research and examined citizen judgements on procedural justice in the context of policing. He considered the criteria citizens use to evaluate fair proceedings and whether perceived procedural fairness influences citizen satisfaction with the outcomes. Together Lind and Tyler (1988) found, that citizens care about the fairness of a process regardless of its outcomes. The significance of procedural justice for policing is best illustrated by Tyler's (1990) research on what leads citizens to obey the law and to cooperate with police. He asserts that citizen perceptions of procedural justice are the main driver for institutional legitimacy, which in turn fosters compliance. If citizens believe that the police act moral and lawful, they are more likely to comply with orders by law enforcement. Subsequent research found overwhelming support for a strong connection between procedural justice and the legitimacy of authorities (Bolger & Walters, 2019). According to Tyler and Huo (2002), procedural justice is equally important to White, Black and Hispanic citizens. It is therefore imperative that police adhere to procedural justice to foster legitimacy among the public.

Across studies, a variety of criteria has been used to capture people's judgements on procedural justice. However, Tyler's work is central to this research because of his strong focus on policing. Thus citizen perceptions of procedural justice are defined following Tyler (2004) as "public views about the appropriateness of the manner in which the police exercise their authority" (p. 91). Tyler (2004) identifies four key elements that shape citizen perceptions of procedural justice. He describes them as participation, neutrality, dignity and respect, and trust. The first element, participation, suggests that police should consider the input of the civilians involved when deciding on a course of action. This does not necessarily mean that citizens need to have a say in the outcome, but rather that they need to feel like their input is being solicited and their concerns are valued. The second factor is neutrality. Citizens consider a police decision to be made fairly if the process appears to be unbiased, which means it is based on objective criteria rather than the personal judgements of the officers. The policing process needs to be conducted as transparently as possible, to allow civilians to assess the neutrality of law enforcement.

The third aspect describes people's desire to be treated with respect and dignity. They want police to acknowledge their rights as an individual and to treat them politely. The quality of treatment affects a person's standing within the community. Especially in situations where people feel unsure or demeaned in their status. For example, during a public questioning by police or after they have become the victim of a crime. Lastly, procedures are perceived as fair when people trust that police act with good intentions. Civilians usually lack specialised knowledge about the work of law enforcement. Therefore, they must trust that decisions are made in good faith and that the actions taken are reasonable. Police can generate trust by explaining their actions and justifying their conduct in a way that makes clear that they are concerned about a person's needs. Tyler (1990) asserts that people generally want to trust authorities and to perceive them as caring. Police-citizen encounters are most impactful in shaping this aspect of procedural justice, i.e. whether a person is treated in the way they expect to be treated.

*Table 1: Key elements of citizen perceptions of procedural justice*

| **Participation** | **Neutrality** | **Dignity and respect** | **Trust** |
|---|---|---|---|
| • Satisfaction with a procedure is higher if people are given space to share their perspective with the authorities.<br>• People want their perspective to be considered during the decision-making process and for authorities to take their input seriously. | • Procedures are seen as fairer if authorities make decisions according to objective criteria.<br>• People want evidence that outcomes are the result of a fair and unbiased process.<br>• Transparent procedures allow people to assess the neutrality and factuality of a process. | • The quality of the personal interaction is a distinct element from the quality of the overall process.<br>• People want authorities to treat them in a polite, dignified, and respectful manner.<br>• How people are being treated impacts their self-worth, especially in situations where they feel unsure or demeaned, e.g. police stops. | • Procedures are perceived as fairer if people trust the motives of decision-makers, i.e. when authorities take people's needs well-being into account.<br>• People rely on their impression of a decision-maker's intentions to judge the validity of their actions.<br>• Trustworthiness is enhanced if authorities explain their decisions and justify their conduct. |

Source: Own depiction based on Tyler (2004)

In conclusion, for an action to be perceived as fair, decision-makers must enable the individuals affected by the decision to offer information and to comment on evidence they are confronted with. In addition, authorities must act absent from personal bias and self-interest, and they should openly declare the reasons for the actions they are taking. Lastly, to gain trust, decision-makers need to make sure that they act with consistency, meaning that they adhere to rules and explain their behaviour.

## Predictive policing and citizen perceptions of procedural justice

In this section, the four main components of procedural justice as defined by Tyler (2004) are put in the context of predictive policing. The goal is to assess the possible effects of the practice on perceptions of procedural justice. Throughout this section, the concept of bias is used. On the one hand, in the case of statistical models, bias does not refer to intentional discrimination, but rather to changes in outcome due to a factor that was not accounted for. For example, the use of datasets in which minority citizens are heavily overrepresented might lead to shifts in the model’s recommendations. (Silva & Kenney, 2018) On the other hand, a person’s individual bias describes either conscious or unconscious differences in the treatment of varying groups of people. While conscious bias refers to overt negative behaviour that a person engages in with intent, unconscious biases are behaviours outside of a person’s awareness and their self-declared values (Selbst, 2017).

### *Participation*

First, the element of participation could be negatively impacted by automation bias. In the context of policing automation bias describes the phenomenon that officers are likely to accept recommendations made by predictive software as correct and to disregard contradictory information (Silva & Kenney, 2018). This could lead officers to ignore input they receive from witnesses or suspects and to purely rely on the algorithm's recommendations. Citizens might feel like their concerns are not being valued and that they have no way of participating in the policing process (Ferguson, 2017). In the case of the NYPD's algorithms, Griffard (2019) argues that automation bias is likely to play a role. The software matches crime patterns and suspects and officers on the ground can subsequently access this information. Due to a lack of transparency of outcomes officers get little opportunity to verify the instructions they receive and have to accept the algorithm's recommendations as factual (Griffard, 2019). This makes law enforcement potentially less susceptible to input from citizens and people feeling more removed from the decision-making process (Silva & Kenney, 2018).

### *Neutrality*

Second, perceptions of the neutrality of police procedures could be impacted positively or negatively, depending on a person's demographic group. Making policing more data-based and objective is one of the main goals of predictive policing. It is supposed to eliminate both conscious and unconscious biases and thus lead to fairer procedures and outcomes (Perry et al., 2013). For this reason, automation bias could lead the public to view predictive policing software as free from human biases (Piotrowicz, 2019). However, if predictive policing reinforces discriminatory patterns of policing, individuals in e.g. racial minority communities are likely to perceive police procedures as insufficiently neutral (Casady, 2011). Because predictive software is usually created in cooperation with the private sector, as is the case with the NYPD and Microsoft, considerations of fairness can fall to the wayside in favour of economic incentives (Byrne & Marx, 2011). For example, it is not in the companies' interest to publicly disclose the algorithms they sell, lest they risk losing their competitive advantage. This structural problem makes it hard for independent researchers to have a say in the design of the software and to assess whether it reinforces biases in policing (Charles, 2019).

Researchers have long pointed out that data-based tools are likely to reproduce or even aggravate existing discrimination, and depending on their design could introduce additional algorithmic biases (Richardson et al., 2019; Selbst, 2017). Lum and Isaac (2016) conducted a simulation of a predictive policing algorithm used to analyse drug crime. Their findings show that "allowing a predictive policing algorithm to allocate police resources would result in the disproportionate policing of low-income communities and communities of colour" (Lum & Isaac, 2016, p. 18). Ferguson (2017) points to bad data as one of the main reasons why predictive policing does not necessarily guarantee a more fair process. Bad data can occur for two reasons. First, it can be generated by human error, i.e. the flawed input of data points, for example concerning a crime scene. These errors then compound and become hard to trace, especially if several databases are combined, as is the case with the NYPD's DAS (Levine et al., 2017). The second factor is statistical bias. The intense policing of certain areas or racial minorities leads to their overrepresentation in the database, which gives the impression of higher crime rates and in turn justifies more police presence. These dynamics can create what Harcourt (2007) describes as a "ratchet effect of targeted enforcement on profiled populations" (p. 37), resulting in higher perceived crime rates and further over-policing.

In their article on the NYPD's predictive software, the developers Chohlas-Wood and Levine (2019) explain that they are aware of concerns about the disparate impact of the technology on over-policed groups. Therefore, they claim to have designed their tool in a way that minimises bias. For one the algorithm was blinded to individual characteristics of suspects, such as race and gender. In addition, "potential proxy variables for sensitive information" (Chohlas-Wood & Levine, 2019, p. 7) were kept purposefully vague. Lastly, the developers stress the need for human review to minimise the risk that the software's recommendations could lead to a wrongful arrest. In her analysis of the NYPD's predictive software Griffard (2019) concludes, that the available information does not allow judgement on whether or not the developers succeeded in their goal to eliminate algorithmic bias from their software. She points out that previous iterations of predictive software likewise did not include demographic variables in their design, but still “produce racially-disparate results that intensify policing of already over-policed communities” (p. 83). Griffard (2019) expects similar results from the NYPD’s tool, because of the disproportionate representation of racial minorities in the police database. In addition, the investigators who review the decisions that the algorithm produces are susceptible to confirmation bias. This means they might be more likely to accept recommendations that reaffirm their prior assumptions about a suspect and to discard information that does not fit their preconceived notions (Griffard, 2019).

However, academic fears about the detrimental effect of predictive policing on procedural neutrality do not necessarily have to reflect public perception. Certain segments of the population might view policing decisions made based on algorithmic recommendations as more neutral than decisions made by the officers themselves. Thus, leading people to have higher confidence in the fairness of policing decisions because they view them as free from human bias (Piotrowicz, 2019). Especially, if they do not live in an area that is targeted by over-policing. Nonetheless, according to Piotrowicz’s (2019) survey among French citizens, 61% of respondents were still concerned about unreliable data. A possible indicator for how research on algorithmic biases affects the public perception of police neutrality is increased media attention on the detrimental effects of predictive policing. A search in the database Factiva using the term “predictive policing” in combination with “bias” or “racism” shows a drastic increase in US coverage in recent years. From 29 hits in 2017 and 32 in 2018 up to 66 in 2019 and 117 in 2020. Headlines such as “Predictive policing algorithms are racist. They need to be dismantled.” (Heaven, 2020) or “Amazon ‘Stands in Solidarity’ While Selling Racist Tech To Police” (Biddle, 2020) are exemplary for the exceedingly negative reporting in recent years (Gilbertson, 2020).

In addition, without sufficient transparency, people are unable to judge if a process is truly neutral. Civilian insight into the process of policing is severely restricted by the secrecy surrounding predictive algorithms. The type and scope of the predictive technologies that are being used are usually not known to the public (Haskins, 2019). In his survey Piotrowicz (2019) found that a large majority of civilians (85%) feel that they are not sufficiently informed about what predictive policing entails and many want more information (73%). This lack of insight makes it impossible for citizens to judge how policing decisions are made and whether biases play a role. In New York City transparency increased only very recently with the passage of the ACLU’s CCOPS laws in June 2020, forcing police to disclose the type of surveillance technology they employ (Feiner, 2020).

### *Dignity and respect*

The third element of procedural justice is respectful, dignified treatment. Ferguson (2017) states that the focus on certain areas or patterns can overwrite regular policing strategy and lead officers to perceive those areas more negatively. In their analysis Ensign et al. (2018) demonstrated this effect, showing that the continuous feeding of newly discovered crime data into the model makes predictive policing systems susceptible to feedback loops. This leads police to repeatedly return to the same area, notwithstanding the actual crime rate. For this reason, if a neighbourhood is assessed as dangerous, it could lead officers to interact differently with the residents. They might stop individuals in those areas more frequently and be more likely to suspect them of criminal behaviour (Ferguson, 2017). Therefore, it is to be expected that especially communities that are overrepresented in the police database (e.g. racial minorities) would perceive a decline in the quality of the interpersonal treatment. Additionally, if patterns are matched across large databases, as is the case with the NYPD’s tools, false positives might result in citizens being suspected of crimes they did not commit (Griffard, 2019). If an individual is identified as a potential subject by the predictive software police might regard all their actions as suspicious, giving officers reasons to interfere with a person’s privacy and making them more likely to treat citizens as guilty (Strikwerda, 2020). Consequently, officers could feel encouraged to push aside moral and ethical concerns, lowering the threshold for physical confrontations which civilians perceive as threatening or violent (Ferguson, 2017).

### *Trust*

Lastly, trust is an aspect of procedural justice where the impact of predictive policing is not ambiguous. Although Piotrowicz (2019) found that 61% of respondents did not think their trust in police would be reduced by the use of predictive tools. His survey also states that citizens are concerned about potential misuse by the police force (71%). It is also not clear if trust could vary across demographics. This indicates that trust in police motives could erode if citizens have the perception that officers abuse their power. The literature makes clear that citizens have a strong desire to see police actions as fair. In a survey among US citizens, Tyler (1990) found that 90% of respondents assumed they would receive just treatment by law enforcement. These views are most strongly impacted during police-citizen encounters. Meaning that if predictive policing does indeed lead to more negative encounters because of over-policing, trust in police motives might decrease over time (Lum & Isaac, 2016). Another potential issue that is relevant to trust is the opaqueness of the algorithms decision-making process (Griffard, 2019). This could make it more difficult for street-level officers to explain their actions to civilians and thus to justify their conduct to citizens.

*Table 2: Expected effect of predictive policing on citizen perceptions of procedural justice*

| Participation | Neutrality | Dignity and respect | Trust |
|---|---|---|---|
| • The NYPD's algorithm matches crime patterns and sends the information to officers on the ground, who have little possibility to review the software's findings and have to trust their accuracy (Griffard, 2019).<br>• Automation bias could make police officers more likely to unequivocally accept the predictive software's recommendations, disregarding input from civilians (Silva & Kenney, 2018).<br>• If officers accept the instructions they receive as factual, they might not look for additional sources of information (Silva & Kenney, 2018).<br>• In this is the case, citizens might feel like their concerns are not being valued and that they cannot participate in the policing process (Ferguson, 2017). | • If predictive policing reinforces discriminatory patterns of policing, individuals in already over-policed areas, e. g. racial minorities are likely to perceive a lack of neutrality (Casady, 2011; Ferguson, 2017; Harcourt, 2007).<br>• Public perception might be influenced by increasingly negative coverage of predictive policing in recent years (Gilbertson, 2020).<br>• However, segments of the public view policing decisions based on algorithmic recommendations as more neutral and free from human bias (Piotrowicz, 2019).<br>• People feel that they are not sufficiently knowledgeable about what predictive policing entails and want more information (Piotrowicz, 2019).<br>• The lack of transparency about the nature of the software being used means civilians are not able to accurately assess the neutrality of the policing process (ACLU, 2016; Haskins, 2019) | • The focus on certain areas or patterns can overwrite regular policing strategy and lead officers to view those areas more negatively, making citizens feel devalued during interactions (Ferguson, 2017).<br>• Officers could feel encouraged to push aside moral and ethical concerns, resulting in physical confrontations which individuals perceive as threatening or violent (Ferguson, 2017).<br>• If the software identifies a suspect, police might regard that person's actions as generally suspicious. This could lead individuals to feel like police invade their privacy and disregard the presumption of innocence. (Strikwerda, 2020). | • A majority of respondents (61%) state their trust in policing would not be negatively impacted by predictive software (Piotrowicz, 2019).<br>• The large majority of people (71%) are concerned about the potential misuse of the software by the police force (Piotrowicz, 2019).<br>• If predictive tools reinforce problematic patterns of policing, then an increase in negative police-citizen encounters could lead citizens to have a less favourable view of police motives (Lum & Isaac, 2016; Tyler, 2004).<br>• The way the software develops recommendations is opaque to street-level officers, making it harder for them to explain their decision-making process to civilians (Griffard, 2019). |
| **Strong negative effect**, due to a high potential for automation bias and a lack of insight into the process. | **Both positive effect** due to ascribed neutrality **and negative effect** in over-policed communities and because of missing transparency. | **Negative effect**, individuals or groups might feel like police disregard their rights and treat them aggressively. | **Effect unclear**, while respondents stated that they did not expect their trust to decrease, they also voiced concerns of misuse. |

## Research design

The proposed research uses a quantitative design to investigate the research question “How does predictive policing influence citizen perceptions of procedural justice during police encounters?”. The case chosen for the analysis is the rollout of predictive policing in New York City. The NYPD is of special interest because it is the biggest police department in the US and one of the first to conduct a broad rollout of predictive policing tools in 2013. It is also one of the few police departments worldwide that have released detailed information about its use of predictive software. Firstly, a paired t-test is conducted to assess differences in the overall number of complaints before and after the introduction of predictive policing. Afterwards, individual-level panel data is used in a random-effects model to determine the impact of predictive policing on perceptions of procedural justice depending on a person’s race. Following the literature review, predictive policing is expected to have a negative effect on perceptions of procedural justice. While the impact on some of the key components is not clear, the anticipated influence on participation as well as dignity and respect leads to the stipulation of an overall negative effect. This results in the formulation of hypothesis one. Furthermore, previous research on predictive policing suggests, that perceptions of neutrality of the software could differ across racial groups. Hence hypothesis two tests for a disparate effect.

**H1:** Predictive policing has a negative effect on citizen perceptions of procedural justice during police encounters.

**H2:** Predictive policing has a stronger negative effect on perceptions of procedural justice during police encounters on Black, Hispanic, and Asian citizens than on White citizens.

## Operationalisation

The NYPD's predictive software DAS was first launched in 2008. However, at this point, the system was only used by the police force's counterterrorism bureau. The software's capabilities include spatial-temporal, individual and pattern analysis (Griffard, 2019; Levine et al., 2017). The focus of the research is DAS' city-wide launch in 2013. For this reason, the period prior to the introduction of predictive policing in New York City is operationalised as the years 2006 to 2012, and the period with predictive policing is the years from 2013 onwards.

To operationalise the concept of procedural justice this paper draws from several studies demonstrating that direct contact with law enforcement strongly influences perceptions of procedural justice (Cao et al., 1996; Frank et al., 2005; Rosenbaum et al., 2005; Schafer et al., 2003). Negative experiences thereby have a far greater impact than positive ones (Dai et al., 2011; Weitzer et al., 2008). For this reason, citizen perceptions of procedural justice are operationalised using complaints about police misconduct. It is important to note, that the available data do not allow conclusions on the potential positive perceptions of predictive policing among citizens, e.g. because they see the algorithm as more neutral. However, it is assumed that the strong effect of negative encounters would cancel out potential positive perceptions. The data used are complaints made about NYPD officers to the independent Civilian Complaint Review Board (CCRB). The CCRB gathers evidence about all allegations and interviews witnesses to compile a report. The board then reviews and votes on the report. The reason civilian complaints against police misconduct are chosen above other indicators such as reports on the use of force or practices such as "Stop and Frisk", is that misconduct complaints depict a wide range of actions that are reflected in Tyler's (2004) assessment of procedural justice in policing. Complaints are segmented into four types of allegations: Force, Abuse of authority, Discourtesy, and Offensive language (FADO). Subsequently, these categories will be attributed to the key elements of procedural justice as defined by Tyler (2004).

The CCRB (2021b) defines an allegation of force as “the use of excessive or unnecessary force” including shoving or punching and up to deadly force. In these cases, all aspects of procedural justice suffer, but the potential for detrimental effect on perceptions of dignity and respect as well as trust is especially high. Since these are exceedingly negative encounters that directly contradict peoples’ inclination to view police as fair and to trust their motives (Tyler, 2004). An abuse of authority occurs, when officers “intimidate or mistreat a civilian”, for example by refusing to provide identifying information or by conducting improper stops. In these cases, the principles of dignity and respect, participation as well as trust are harmed above all. Since the offending officer abuses their position of power over a civilian, disregarding their constitutional rights and their view of the situation. The third category is discourtesy and “refers to cursing and using other foul language or gestures”. Lastly, offensive language is “slurs and derogatory remarks or gestures”, that is based on discriminatory factors such as race, gender, or sexual orientation. (CCRB, 2021a) Allegations of the former category can point to the fact that citizens perceive a lack of dignity and respect in their interactions with police. The latter category indicates that citizens do not see their interactions with police as neutral, since they involve the reinforcement of discriminatory stereotypes.

*Table 3: Matching FADO categories to citizen perceptions of procedural justice*

| **Participation** | **Neutrality** | **Dignity and respect** | **Trust** |
|---|---|---|---|
| • abuse of authority, force | • offensive language, force | • force, abuse of authority, discourtesy | • force, abuse of authority |

## Data

The basis of the research is two different datasets of civilian complaints against police misconduct. The first includes an overview of the total number of complaints, while the second offers a detailed picture of the complainants and the types of allegations made. Civilian complaints are made to the CCRB, which became an independent entity from the police department on 5 July 1993 (Patterson, 2006). The CCRB defines a complaint as any reported incident that violates Section 440 of the New York City Charter (American Legal, 2021). However, the CCRB is not responsible for the relatively small number of criminal complaints. Those are investigated by state and federal prosecutors in cooperation with the NYPD's Internal Affairs Bureau or the Federal Bureau of Investigation (FBI) and thus not part of the aforementioned datasets (ProPublica, 2020).

Each complaint that the CCRB receives comprises one or more allegations that fall into the four FADO categories. After submission complaints are investigated by the CCRB by gathering witnesses and evidence. Finally, a report on the incident is prepared and turned over for a vote to the CCRB Board. (CCRB, 2021b) Based on the available evidence an allegation can be deemed either substantiated, unsubstantiated, unfounded (evidence suggests it did not occur), exonerated (the event occurred but was not deemed improper or closed (the offending officer could not be identified) (CCRB, 2021a). The CCRB only substantiates a small number of cases each year, meaning that the offending officer faces disciplinary action. Investigations are hampered by the CCRBS's lack of resources and the NYPD's insufficient cooperation, for example, the refusal to comply with requests to turn over body camera footage. (Umansky & Simon, 2020) Therefore, the following analysis incorporates claims regardless of finding, including those that were dismissed due to an apparent lack of evidence. The NYPD's lack of compliance with the CCRB's investigations means claims that are either deemed unsubstantiated or closed without conclusion, cannot automatically be treated as unfounded. Likewise, cases in which the complainant did not cooperate with the investigation, might point to institutional hurdles, rather than the absence of police misconduct. Regardless, the validity of the claim is not vital for the analysis since the focus of the research is citizen perception rather than actual police procedural justice.

The first dataset used in the analysis stems directly from the CCRB (2021b) and provides the basis for the paired t-test. It includes all complaints made from January 2006 to December 2019, regardless of the status of the investigation. There was a total of 81,547 citizen complaints recorded in the respective time. However, the dataset only includes complaint totals and no information on individual allegations or complainant demographics. Thus, more detailed data is needed for an in-depth analysis of complaints and allegations on an individual level.

For this purpose, the second dataset ProPublica's (2020) "Civilian Complaints Against New York City Police Officers" was selected. It was released in July 2020 following a successful Freedom of Information Law (FOIL) request. The dataset spans complaints from 1985 to January 2020 across all the NYPD's 77 precincts. The data does not include allegations that were deemed unfounded by the CCRB. However, it does contain allegations that were deemed unsubstantiated. This means that the CCRB confirmed that the incident took place and fully investigated it but concluded that it did not violate the NYPD's rules. Each record in the data lists the name, shield number, rank, and precinct of the officer in question as of July 2020 and at the time of the incident, the age, race and gender of the complainant and the officer. Furthermore, there is a category describing the alleged misconduct; and whether the CCRB concluded that the officer violated NYPD rules. The database is limited to officers that had at least one substantiated complaint against them. This pertains to roughly 4,000 of the 36,000 officers in the NYPD's force. (ProPublica, 2020)

To narrow ProPublica's (2020) data a relevant timeframe was chosen first. The decision was based on the fact that in 2019 an investigation took the CCRB an average of nine months (Mayor's Office of Operations, 2020). This means that many of the complaints made in 2019 and 2020 are likely still pending and thus not a part of the data at hand. Therefore, both years were excluded from the analysis as to not distort the findings. In 2013 the NYPD conducted the first broad roll-out of predictive policing. Thus, 1 January 2013, was chosen as the breakpoint of the analysis. Subsequently, a time frame of 12 years from 1 January 2007 to 31 December 2018, was selected. Then the dataset was cleared by deleting all allegations outside the relevant timeframe and those missing a date (N = 9,228). In addition, any allegations where the race of the complainant was unknown, or the complainant refused to state their race, were excluded (N = 3,222). To enhance the interoperability of the data demographic categories used by the U.S. Census Bureau (2011) were introduced. American Indian complainants were added to the other race category. In the category gender transwomen and -men, as well as gender-nonconforming individuals, were grouped as the category other gender. This leads to a total of 7,522 complaints that comprise 20,908 allegations. Meaning that each complaint contains on average 2.78 allegations.

The dataset comprises yearly panel data, where observations are made about different cross-sections over time. In this case, the individuals observed are those falling under the NYPD's purview. The underlying data was formatted using race as the panel variable covering five groups: White = 5, Black = 4, Hispanic = 3, Asian = 2 and other race = 1. Subsequently, dummy variables for the five panel categories were added as main independent variables. Additionally, the dummy variable PredPol was included. It turns on if the time variable is greater than or equal to 2013. The dependent variables are the total number of complaints and allegations as well as the allegations across the four FADO categories (see appx. 1, 2). All values are measured quarterly, to ensure an adequate number of observations per period. Thus, the total number of observations is N = 240 in periods T = 48 making it a strongly balanced, short panel dataset (N > T). Five control variables were introduced, all measured at yearly intervals (see appx. 3). They are the number of NYPD officers per citizen, the amount of media coverage of policing and the total number of crimes, arrests as well as police stops. The variable media attention was determined by using the database Factiva and filtering for publications in New York City that contain the keywords police or policing in either headline or lead paragraph. Additional control variables like the demographic makeup of the city and the police force were considered (see appx. 4, 5). However, their use was discarded due to a lack of variation across the two time periods.

## Methods

Firstly, the paired-samples t-test is used to assess whether the mean difference between the dependent variable in period t_1 and period t_2 is statistically different from zero (Backhaus et al., 2018, p. 87). Several assumptions are underlying the paired t-test. First, the dependent variable must be measured at a continuous level and the independent variable must consist of two categorical related groups. In this case, the independent variable is periodical, i.e. the period before and after the introduction of predictive policing, the paired-samples t-test is preferable to a simple linear regression. The use of an independent samples t-test was discarded since both measurements stem from the same sample, namely the complainants. Further assumptions tested in the empirical section of this paper are the absence of significant outliers in between-group differences and approximate normal distribution of the dependent variable.

Secondly, a random-effects GLS regression was chosen based on several factors. Panel data can be analysed using both random and fixed effects models. To ascertain which is suitable for the data at hand a Breusch and Pagan Lagrangian multiplier test was conducted (see appx. 6). The null hypothesis of the test states that variances across entities equal zero. That means there is no significant difference across sections and therefore no panel effect. Running the test with race as the panel variable returns a p-value of $p = 0.0000$. This leads to a rejection of the null hypothesis and the assumption that a random-effects model is appropriate for the planned research design. That means the variance of errors of the regression is dependent on the independent variable and heteroskedasticity is present (Greene, 2008, p. 183). Therefore, the dependent variable is influenced by differences across the different categories. Indicating that the number of complaints is influenced by a person’s race. This further solidifies the case for using a random-effects model. An advantage of random effects is the possibility to include invariant variables such as gender that can serve as explanatory variables. This is relevant for several of the control variables included in the panel data. However, there is a possibility of omitted variable bias if not all variables that influence the predictor variable are included in the model. The random-effects model is:

$$Y_{it} = \beta X_{it} + \alpha + u_{it} + \varepsilon_{it}$$

With the dependent variable $Y$, the explanatory variable $X$, the regression coefficient $\beta$ and the intercept with the y-axis $\alpha$. The between entity $u_{it}$ is added together with the within entity error $\varepsilon_{it}$. It allows generalising interferences beyond the sample used in the model. The specific method used is a GLS regression. The GLS technique is commonly used to estimate unknown parameters in a linear regression when there is a degree of correlation between the model's residuals. If this is the case a regular Ordinary Least Squares (OLS) regression can be statistically inefficient or result in misleading interferences. In contrast to OLS regression, the GLS model is robust against outliers, heteroskedasticity and bias in data. It produces the best linear unbiased prediction (Greene, 2008, p. 60). To prepare the analysis, tests for normality and stationarity are conducted. Normality is evaluated by assessing the levels of skewness and kurtosis of the dependent variables. Then an Augmented Dickey-Fuller (ADF) test is conducted to test for stationarity. It assesses whether a unit root present in the data sample, i.e. that it exhibits a systematic pattern that cannot be predicted.

## Empirical findings

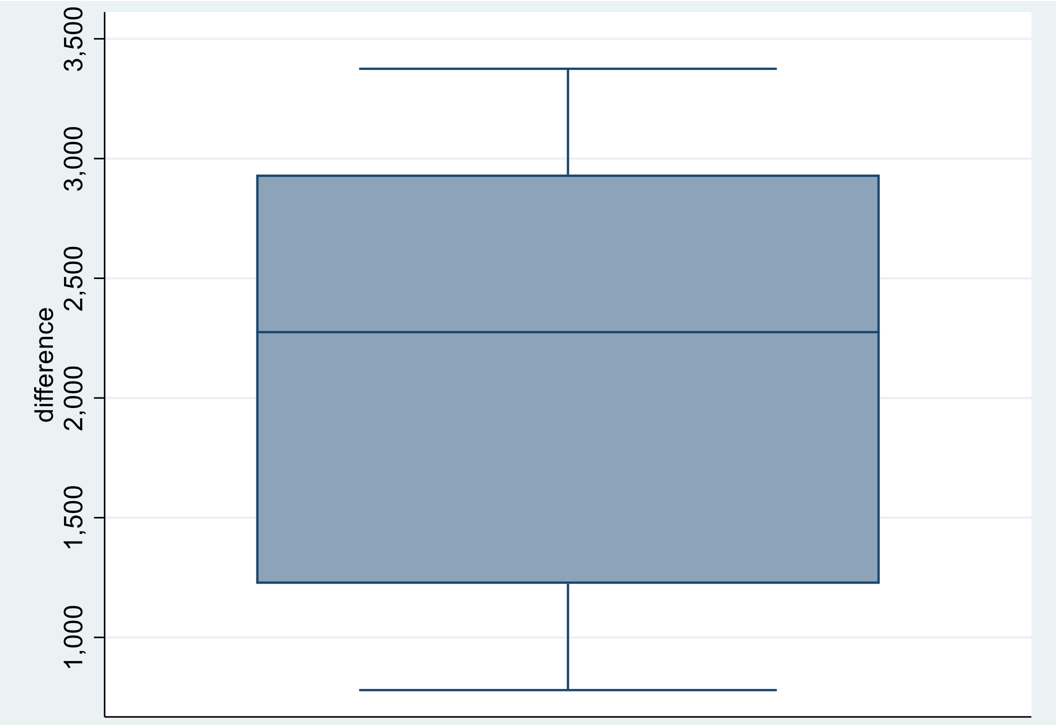
This section details the steps taken and the findings of the paired t-test and the GLS regression. First preliminary assumptions of the tests are discussed and then the main empirical findings are presented, including descriptive statistics and regression tables. While the t-test is used to test the validity of H1, the GLS regression allows for a more detailed exploration of the differences between the two periods and as such is used to examine H2 and to a lesser extent H1. This includes running random effects regressions on six different dependent variables. To cover changes in both complaints and the allegations contained within them, as well as the different categories of allegations.

### Results of the paired-samples t-test

The paired-samples t-test has four main underlying assumptions (Backhaus et al., 2018, p. 87). Firstly, the dependent variable should be measured at either interval or ratio level. In this case, the dependent variable complaints is measured at a ratio level, as it can logically never fall below zero. Secondly, the independent variable should be categorical. Here the periods before and after the introduction of predictive policing are treated as two separate categories. Both categories draw from the same underlying sample of individuals the groups are not independent of each other. As such they are related groups because the city-level panel data draws from complaints against police misconduct made within New York City.

To validate the results of the paired t-test two further assumptions have to be met. These are the absence of significant outliers and the normal distribution of the differences of the dependent variable between both groups (Backhaus et al., 2018, p. 87). Outliers are data points that do not follow the usual pattern of the dataset. If present they can distort the differences between the related groups, for example because the number of complaints in one year is extraordinarily high. This reduces the accuracy of the results and affects the statistical significance of the test. To detect outliers the difference scores of the data points are graphed in a boxplot. Any difference scores that are more than 1.5 box lengths from the edge of the box are classified as outlier values. The inspection of the boxplot shown below shows no significant outliers.

*Figure 1: Boxplot of total complaints*




Source: Own depiction

Subsequently, an assessment of the distribution of the differences in the dependent variable was conducted. Because the t-test is relatively robust to violations of normality it only needs approximately normal data to provide valid results (Backhaus et al., 2018, p. 87). To test for normality a Shapiro-Wilk test was conducted. The null hypothesis of the test is that the data is normally distributed. Therefore, if the p-value is below the chosen alpha level of 5% the null is rejected, and the data is assumed to not follow a normal distribution. The result below shows that the dependent variable is normally distributed in both t_1 and t_2, thus the null hypothesis is not rejected. This is demonstrated by the rounded p-value of 0.07 for the pre-predictive policing period and 0.69 for the post-predictive policing period.

*Figure 2: Shapiro-Wilk test for normality of complaints*

| Variable | Obs | W | V | z | Prob>z |
|---|---|---|---|---|---|
| C_1 | 7 | 0.82185 | 2.340 | 1.499 | 0.06691 |
| C_2 | 7 | 0.94575 | 0.713 | -0.498 | 0.69089 |

Source: Stata results table

The results of the paired t-test are depicted below. Numbers are reported rounded to integers in the text, to account for the fact that partial complaints are not possible. The focus of the analysis is the mean and the Standard Deviation (SD). The Standard Error (SE) of the mean will be neglected for the t-test since it is considered flawed in many cases as discussed by Carter (2013). First, the mean of complaints in t_1 is 6,920 and for t_2 it is 4,729. As such the average number of complaints in t_2 was 2,191 points lower than that in t_1. However, SD in t_1 was much higher at 839 than in t_2 with 369. The mean difference or change between the two periods is 2,191 with an SD of 937. Because of the order of the variables, a possible mean difference indicates that the values in t_1 were on average 2,191 points higher than in t_2. With 95% confidence, the true mean change in complaints lies between 1,325 and 3,058. This change in the number of complaints between the two time periods is statistically significant at p = 0.0004. It suggests that the mean change between the two groups in the population does not equal zero.

*Figure 3: Paired-samples t-test for total complaints*

| Variable | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| C_1 | 7 | 6920.429 | 317.018 | 838.7508 | 6144.713 | 7696.144 |
| C_2 | 7 | 4729.143 | 139.606 | 369.3627 | 4387.539 | 5070.746 |
| diff | 7 | 2191.286 | 354.1444 | 936.9781 | 1324.726 | 3057.846 |

```
     mean(diff) = mean(C_1 - C_2)                               t =   6.1875
 Ho: mean(diff) = 0                           degrees of freedom =        6

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.9996         Pr(|T| > |t|) = 0.0008          Pr(T > t) = 0.0004
```

Source: Stata results table

Because there was a statistically significant difference between means ($p < 0.05$) the null hypothesis can be rejected. The alternative hypothesis is accepted with 95% confidence meaning that it is unlikely that the difference between the two means occurred by chance. The size of the effect is calculated using Cohen's d ($d = \frac{M}{SD}$). For the reported mean and SD, the effect size and as such the practical significance of the result amounts to 0.38. A Cohen's d of 0.2 is considered small, a value of 0.5 medium and 0.8 high (Lakens, 2013). Thus, the observed effect can be described as small.

## Results of the GLS regression

Before conducting the GLS regression first assumptions of normality and stationarity need to be validated (Backhaus et al., 2018, p. 35). To validate the first requirement skewness and kurtosis tests were run for the dependent variables: complaints and allegations (see appx. 7). Skewness indicates whether a variable is symmetrically distributed, and kurtosis assesses whether a variable's distribution is too narrow, i.e. peaked. According to Hair et al. (2017, p. 61), a variable's distribution is considered normal if the values of both measurements equal zero. However, this is a very uncommon occurrence. The general assumption is that if the values of skewness or kurtosis are higher than +1 or lower than -1 distribution is considered nonnormal (Hair et al., 2017, p. 61). For the variable complaints, skewness is 0.0000 and kurtosis 0.3046, while for allegations skewness is 0.000 and kurtosis 0.3345. Therefore, both show a normal although slightly peaked distribution. As such the assumption of normality is fulfilled.

Subsequently, a Fisher-type unit root test based on ADF tests for stationarity was conducted (see appx. 8, 9). The test includes three lags, one less than the time variable, which is measured at an interval of four quarters per year. The null hypothesis of the ADF test states that a unit root exists. Thus, if the null is rejected no unit root, i.e. unpredictable pattern, is part of the sample and the data is stationary (Backhaus et al., 2018, p. 96). For the dependent variable complaints, the unit-root test shows p-values < 0 and the same holds for the values of allegations. Thus, both variables are considered stationary, and no autocorrelation can be detected.

The descriptive statistics for both t_1 and t_2 can be found on the next page. They include, besides mean and SD, the Coefficient of Variation (CV) as a measure for the extent of variability ($CV = \frac{\sigma}{\mu}$). All measures are expressed in yearly periods because the data for the independent variables are only available in these intervals. If reported quarterly the results regarding their variation would be distorted. Controls are the ratio of police officers to citizens in New York City, media coverage, recorded crimes as well as total stops and arrests. Those variables were chosen based on similar research to enhance comparability. Specifically the NYU Public Safety Lab's (2021) report on police misconduct in New York City. The dependent variables are complaints and allegations among the demographic groups White (5), Black (4), Hispanic (3) and Asian (2). The variable other race was dropped due to collinearity with the ethnic category Asian.

Firstly, descriptive statistics for the variable Complaints are examined. For complaints made by White people, they show a slight increase to around 74 complaints in t_2 up from 64 in t_1. The dispersion of the data points is 6.86% lower in t_2, meaning there was less variation across measurements in that period. Mean complaints by Black New Yorkers decreased from t_1 to t_2 by roughly 64 points, from 414 to 351. Variation was relatively low in both periods but decreased slightly from 5.42% to 4.72% in t_2. For Hispanic complainants, the overall number of complaints decreased slightly by around 6 points to 142 in t_2 and variation was 4.95% lower at 4.36%. For Asians, the mean number of complaints almost doubled, while remaining comparatively low, from 10 in t_1 to 17 in t_2. In t_1 variation was highest in this category with 34.06%. However, it decreased significantly in t_2 to 8.72%.

The findings show that Black people made up over 60.43% of complainants in the dataset in t_2 compared to 65.04% in t_1 while representing only roughly 22.80% of the population according to the U.S. Census Bureau (2011). The number of Hispanic complainants in the sample is slightly lower than their share in the population of 28.60% and 24.48% of the mean complaints in t_2, up from 23.34% in t_1. Meanwhile, 33.30% of New Yorkers are White but the mean of their complaints in t_2 only makes up only roughly 12.67% of the overall mean complaints, slightly up from 10% in t_1. Asian complainants are the group with the lowest number of complaints in both periods, despite an increase in t_2 their share of the complaints is only 2.93% up from 1.57% in t_1. This is significantly lower than their share in the population of 12.60%. Overall, the change in the CV from period t_1 to t_2 shows a decrease of variation across the groups. Therefore, there is less variability in the data in relation to the mean, i.e. the number of complaints does not change as much from year to year compared to the mean.

Secondly, allegations by White people increased from around 142 to 210, while variation is 1.27% lower. For Black people mean allegations increased from 1,031 to 1,109 with variation increasing slightly by 0.21%. Hispanics have the second highest mean of allegations in both periods, with an increase from 375 to 454 points and an extremely small increase in variation by 0.1%. For Asians mean allegations increased from 22 to 50 points, with variation up by 2.49%. While mean complaints by Black and Hispanic people decreased in t_2, the average number of allegations increased for both groups. The allegation to complaint ratio for Black people increased from 2.49 in t_1 to 3.16 in t_2. For White complainants, the ratio increased from 2.21 to 2.86, for Hispanic people from 2.52 to 3.19 and Asians from 2.15 to 2.92. Therefore, complaints on average contained more allegations in t_2 across all demographic groups.

Lastly, the control variables are inspected. The city's population mean increased from an average of 8,273,767 in t_1 to 8,521,224 in t_2. Likewise, the mean number of police officers increased by over 500 officers, from 35,025 to 35,590 in t_2. The media coverage increased by roughly 2,476 publications to an average of 8,770 articles per year in t_2. In contrast, the average of crimes committed decreased from 193,707 to 180,080 and mean arrests declined from 335,945 to 255,718. The number of stops exhibits the highest change, with an average of 568,914 in t_1 a contrast to merely 49,207 in t_2. This amounts to merely around 8.64% of the mean number of stops during the previous period. The expectation for the independent variables is that the higher the number of police officers per citizen, the more likely are negative police-citizen interactions. This officer-to-citizen ratio declined from t_1 to t_2, with an average of 423 officers per 100.000 citizens in t_1 compared to 418 in t_2. Furthermore, lower crime rates, fewer stops by the police and arrests should lead to fewer police-citizen interactions, thus decreasing the likelihood of complaints. However, an increase in reporting on policing could sensitise citizens to unfair treatment by police, possible leading to more widespread knowledge of reporting mechanisms.

*Table 4: Descriptive statistics for period t_1*

| | **Mean** | **SD** | **CV** |
|---|---|---|---|
| **Dependent variables** | | | |
| Complaints 5 (White) | 64.00 | 7.24 | 11.31% |
| Complaints 4 (Black) | 414.33 | 22.44 | 5.42% |
| Complaints 3 (Hispanic) | 148.67 | 13.84 | 9.31% |
| Complaints 2 (Asian) | 10.00 | 3.41 | 34.06% |
| Allegations 5 (White) | 141.67 | 9.01 | 6.36% |
| Allegations 4 (Black) | 1031.33 | 31.26 | 3.03% |
| Allegations 3 (Hispanic) | 374.50 | 14.10 | 3.77% |
| Allegations 2 (Asian) | 21.50 | 2.63 | 12.23% |
| **Independent variables** | | | |
| Population | 8,273,766.83 | 87,090.08 | 1.05% |
| Police officers | 35,025.00 | 486.81 | 1.39% |
| Media coverage | 6,294.67 | 616.69 | 9.80% |
| Crimes | 193,706.67 | 5,087.90 | 2.63% |
| Stops | 568,914.33 | 72,568.87 | 12.76% |
| Arrests | 335,945.33 | 6,494.78 | 1.93% |

*Table 5: Descriptive statistics for period t_2*

| | **Mean** | **SD** | **CV** |
|---|---|---|---|
| **Dependent variables** | | | |
| Complaints 5 (White) | 73.50 | 3.27 | 4.45% |
| Complaints 4 (Black) | 350.50 | 16.53 | 4.72% |
| Complaints 3 (Hispanic) | 142.33 | 6.21 | 4.36% |
| Complaints 2 (Asian) | 17.00 | 1.48 | 8.72% |
| Allegations 5 (White) | 210.33 | 10.71 | 5.09% |
| Allegations 4 (Black) | 1,108.50 | 35.93 | 3.24% |
| Allegations 3 (Hispanic) | 453.67 | 17.58 | 3.87% |
| Allegations 2 (Asian) | 49.67 | 7.31 | 14.72% |
| **Independent variables** | | | |
| Population | 8,521,224.33 | 31,570.66 | 0.37% |
| Police officers | 35,589.67 | 314.63 | 0.88% |
| Media coverage | 8,770.67 | 779.07 | 9.00% |
| Crimes | 180,079.50 | 3,598.30 | 2.00% |
| Stops | 49,207.17 | 29,037.08 | 59.01% |
| Arrests | 255,717.83 | 17,497.41 | 6.84% |

For the GLS regression, the data is used in the form of strongly balanced panel data, at a delta of one quarter and four quarters per year. With an overall N = 240 each separate period contains 120 observations in five groups of N = 24. Although data for the control variables are only available in yearly intervals, it should not influence the outcome since GLS regressions are robust to measurements at different intervals. In addition to running a regression on the overall number of complaints and allegations, one was also conducted for each FADO category. The regressions are reported with the p-value of the overall model Prob > $chi^2$ and the coefficient of determination $R^2$, i.e. the share of the dependent variable's variance that can be predicted from the independent variable (Backhaus et al., 2018, p. 77). The coefficients show the average effect of the independent variable over the dependent variable when the independent variable changes across time and demographic group by one unit. When interpreting the coefficients, it needs to be remembered that they include both within-entity and between-entity effects. Thus an exact interpretation is quite difficult. The coefficients can merely give a sense of the magnitude of the effect and not its exact size.

The regression of complaints returns significant results for the independent variable race for White, Black, and Hispanic complainants in both periods. For Asian complainants, the p-value equals 0.735 in t_1 and 0.754 in t_2, reaching significance neither at the 95% nor the 90% confidence level. Thus, the demographic category Asian does not have a significant effect on the number of complaints. For t_1 the panel variable race in the category White shows an effect of 12.708 points on the number of complaints. For Black people, the effect was almost 10 times higher, with roughly 100.292 points and for Hispanics, it is an increase of 33.875 points. All values exhibit a SE of 2.339, i.e. their maximum distance from the regression line. Showing that all sample means are relatively close to the population mean, although the error has a potentially greater effect the lower the coefficient. That means it has a greater impact on the independent variable White. In t_2 the effect of the category White increases slightly to 13.333, for Black people it decreases by almost 20 points to 82.583 and a slight decrease to 30.542 can be observed for Hispanic individuals. The SE increases slightly to 2.527 points but remains low.

A similar picture in terms of statistical significance emerges for the dependent variable allegations. The categories White, Black and Hispanic are all associated with highly significant results, while the effect of the racial group Asian shows no statistical significance at p = 0.690 and p = 0.736 in t_1 and t_2, respectively. The independent variable White is connected to an increase by 27.417 points and the variable Black is associated with a 249.833 increase in allegations. Signifying a more than nine times higher for Black people than a White person. For the category Hispanic the effect is an 85.625-point increase. The SE is slightly higher than that for Complaints at 6.592. However, because the values of the coefficients are higher as well this does not necessarily lead to a wider proportional spread. In t_2 the effect of the different racial groups on allegations increased across the board. For the category White, the effect was roughly 10 points higher and for Black as well as Hispanic complainants an increase of around 12 points can be observed. The 95% confidence interval of the values lies at ± 8.773.

*Table 6: GLS regression for complaints and allegations*

| **Complaints** | **t_1** (Prob > $chi^2$ = 0.0000; $R^2$ = 0.9594) | | **t_2** (Prob > $chi^2$ = 0.0000; $R^2$ = 0.9317) | |
|---|---|---|---|---|
| | **Coefficient** | **SE** | **Coefficient** | **SE** |
| **5 (White)** | 12.708 *** (0.000) | 2.339 | 13.333 *** (0.000) | 2.527 |
| **4 (Black)** | 100.292 *** (0.000) | 2.339 | 82.583 *** (0.000) | 2.527 |
| **3 (Hispanic)** | 33.875 *** (0.000) | 2.339 | 30.542 *** (0.000) | 2.527 |
| **2 (Asian)** | -.792 (0.735) | 2.339 | -.792 (0.754) | 2.527 |
| **Officers/citizen** | -66,516 * (0.075) | 37,318.03 | -32,065.48 (0.694) | 81,507.81 |
| **Media coverage** | .0005 (0.608) | .0009 | .002 (0.297) | .002 |
| **Crimes** | .001 * (0.067) | .0006 | .00001 (0.989) | .0009 |
| **Stops** | -.00001 (0.561) | .00002 | -.0001 (0.531) | .0002 |
| **Arrests** | .0009 * (0.054) | .0005 | -.0002 (0.345) | .0002 |
| **Constant** | -224.904 (0.102) | 137.595 | 161.752 (0.757) | 522.372 |
| **Allegations** | **t_1** (Prob > $chi^2$ = 0.0000; $R^2$ = 0.9473) | | **t_2** (Prob > $chi^2$ = 0.0000; $R^2$ = 0.9198) | |
| **5 (White)** | 27.417 *** (0.000) | 6.592 | 37.208 *** (0.000) | 8.773 |
| **4 (Black)** | 249.833 *** (0.000) | 6.592 | 261.75 *** (0.000) | 8.773 |
| **3 (Hispanic)** | 85.625 *** (0.000) | 6.592 | 98.042 *** (0.000) | 8.773 |
| **2 (Asian)** | -2.625 (0.690) | 6.592 | -2.958 (0.736) | 8.773 |
| **Officers/citizen** | -164,388.5 (0.118) | 105,163.4 | -34,949.55 (0.902) | 283,003.1 |
| **Media coverage** | .005 * (0.053) | .003 | .005 (0.463) | .006 |
| **Crimes** | .002 (0.154) | .007 | -.00009 (0.979) | .003 |
| **Stops** | -.00004 (0.413) | .00005 | .0001 (0.795) | .0005 |
| **Arrests** | -.002 * (0.096) | .001 | -.0003 (0.649) | .0006 |
| **Constant** | -502.294 (0.195) | 387.748 | 199.507 (0.912) | 1,813.726 |

p-values in brackets with * = p < 0.1; ** = p < 0.05; *** = p < 0.000

Subsequently, the effect of the independent variables on different categories of allegations is tested. For allegations of force, the categories White, Black and Hispanic are significant for both periods, while Asian is not statistically significant at $p = 0.642$ and $p = 0.690 > 0.05$. For t_1 the demographic category White leads to an increase of 5.667. However, for Black people, the effect is more than 10 times higher at 61.042 points and for Hispanic individuals, it is 23.5 points. Roughly four times higher than that of White people. In t_2 the effect of the category White on allegations of Force is slightly higher than in t_1 at 7.250. For Black complainants, the effect remains positive but decreased by roughly five points to 56.042. For the category Hispanic there is a marginal decrease of the effect of .250 points. The SE increased from t_1 to t_2 by .871 points, showing a slightly wider spread.

For allegations of abuse of authority, again only the categories White, Black and Hispanic are statistically significant. A complainant being White is associated with a 16.042 increase, but for Black people, it is more than nine times higher at 148.583 points. For Hispanics, it is connected to a positive effect of 46.667 points. The effect increased for all three categories in t_2. For White complainants by 7.625 points, Black citizens by 14.334 points and Hispanic people by 11.250 points. The SE reveals a rise by 1.798, in line with the increase in the value of the coefficients.

*Table 7: GLS regression for force and abuse of authority*

| **Force** | **t_1** (Prob > chi$^2$ = 0.0000; $R^2$ = 0.9393) | | **t_2** (Prob > chi$^2$ = 0.0000; $R^2$ = 0.8498) | |
|---|---|---|---|---|
| | **Coefficient** | **SE** | **Coefficient** | **SE** |
| **5 (White)** | 5.667 ** (0.002) | 1.8397 | 7.250 *** (0.008) | 2.711 |
| **4 (Black)** | 61.042 *** (0.000) | 1.8397 | 56.042 *** (0.000) | 2.711 |
| **3 (Hispanic)** | 23.500 *** (0.000) | 1.8397 | 23.250 *** (0.000) | 2.711 |
| **2 (Asian)** | -.833 (0.642) | 1.8397 | -1.083 (0.690) | 2.711 |
| **Officers/citizen** | -67,059.820 (0.019) | 28,579.9 | 83,616.280 (0.341) | 87,780.620 |
| **Media coverage** | .001 * (0.061) | .0007 | -.0005 (0.790) | .002 |
| **Crimes** | .001 ** (0.010) | .0004 | .001 (0.295) | .001 |
| **Stops** | -.00002 (0.180) | .00001 | -.0002 (0.372) | .0002 |
| **Arrests** | .001 ** (0.007) | .0004 | .0002 (0.420) | .0002 |
| **constant** | -262.658 ** (0.013) | 105.377 | -566.105 (0.314) | 562.574 |
| **Abuse of authority** | **t_1** (Prob > chi$^2$ = 0.000; $R^2$ = 0.9423) | | **t_2** (Prob > chi$^2$ = 0.000; $R^2$ = 0.9042) | |
| **5 (White)** | 16.042 *** (0.000) | 4.185 | 23.667 *** (0.000) | 5.983 |
| **4 (Black)** | 148.583 *** (0.000) | 4.185 | 162.917 *** (0.000) | 5.983 |
| **3 (Hispanic)** | 46.667 *** (0.000) | 4.185 | 57.917 *** (0.000) | 5.983 |
| **2 (Asian)** | -1.458 (0.728) | 4.185 | -.500 (0.933) | 5.983 |
| **Officers/citizens** | -47737.580 (0.475) | 66,772.940 | -132,465.700 (0.493) | 193,009.100 |
| **Media coverage** | .003 (0.102) | .002 | .005 (0.293) | .004 |
| **Crimes** | .0006 (0.543) | .001 | -.001 (0.507) | .002 |
| **Stops** | -.00003 (0.316) | .00003 | .0003 (0.433) | .0004 |
| **Arrests** | .0008 (0.334) | .0009 | -.0004 (0.358) | .0004 |
| **Constant** | -189.791 (0.441) | 246.198 | 873.505 (0.480) | 1,236.968 |

p-values in brackets with * = $p < 0.1$; ** = $p < 0.05$; *** = $p < 0.000$

Lastly, the effect of the different demographic groups on allegations of discourtesy and offensive language was tested. For discourtesy, the category White is connected to a 5.417 increase, for Black complainants, the effect is roughly seven times higher at 35.250 and for Hispanic ones the value rises by 13.917 points. The change from t_1 to t_2 is relatively small across the three categories that show a statistically significant effect of $p < 0.05$. For White complainants, the effect slightly decreases by .167 points to 5.25. For Black as well as Hispanic complainants the effect grows slightly. In the category Black, the increase equals the decrease for the group of White complaints, i.e. .167 points. For Hispanic people, the increase was slightly higher at 1.083 points elevating the effect to 15 points. The 95% confidence interval increased from ±1.579 in t_1 to ±1.698 in t_2.

For the dependent variable offensive language, the categories that exhibit a highly significant effect in both periods are Black and Hispanic. The category White does not show a statistically significant result at $p = 0.422 > 0.05$ in t_1. This can be attributed to the fact that the overall number of allegations by White people in this category is extremely low. For t_2, where the effect is significant at $p = 0.031 < 0.05$, the value of the coefficient is merely 1.042 points. For both Black and Hispanic complainants, the effect increased from period t_1 to t_2. For the first group, the increase is 2.417 points and for the second, it is only a slight effect at 0.333 points. With the overall coefficients being relatively low the SE is low as well at .353 in t_1 and .484 in t_2.

*Table 8: GLS regression for discourtesy and offensive language*

| **Discourtesy** | **t_1**<br>(Prob > $chi^2$ = 0.000; $R^2$ = 0.8659) | | **t_2**<br>(Prob > $chi^2$ = 0.000; $R^2$ = 0.8534) | |
|---|---|---|---|---|
| | **Coefficient** | **SE** | **Coefficient** | **SE** |
| **5 (White)** | 5.417 **<br>(0.001) | 1.579 | 5.250 **<br>(0.002) | 1.698 |
| **4 (Black)** | 35.250 ***<br>(0.000) | 1.579 | 35.417 ***<br>(0.000) | 1.698 |
| **3 (Hispanic)** | 13.917 ***<br>(0.000) | 1.579 | 15.000 ***<br>(0.000) | 1.698 |
| **2 (Asian)** | -.208<br>(0.895) | 1.579 | -1.417<br>(0.404) | 1.698 |
| **Officers/citizen** | -36,966.970<br>(0.142) | 25,188.650 | -6,056.550<br>(0.912) | 54,767.370 |
| **Media coverage** | .0007<br>(0.281) | .0006 | .0008<br>(0.502) | .001 |
| **Crimes** | .0004<br>(0.274) | .0004 | .00007<br>(0.909) | .0006 |
| **Stops** | 9.43e-06<br>(0.455) | .00001 | .00003<br>(0.766) | .0001 |
| **Arrests** | .0003<br>(0.348) | .0003 | -.00007<br>(0.560) | .0001 |
| **Constant** | -35.204<br>(0.705) | 92.873 | 24.056<br>(0.945) | 350.996 |
| **Offensive language** | **t_1**<br>(Prob > $chi^2$ = 0.000; $R^2$ = 0.7386) | | **t_2**<br>(Prob > $chi^2$ = 0.000; $R^2$ = 0.7455) | |
| **5 (White)** | .292<br>(0.422) | .353 | 1.042 **<br>(0.031) | .484 |
| **4 (Black)** | 4.958 ***<br>(0.000) | .353 | 7.375 ***<br>(0.000) | .484 |
| **3 (Hispanic)** | 1.542 ***<br>(0.000) | .353 | 1.875 ***<br>(0.000) | .484 |
| **2 (Asian)** | -.125<br>(0.723) | .353 | .0417<br>(0.932) | .484 |
| **Officers/citizen** | -12,624.140 **<br>(0.025) | 5,629.286 | 19,956.45<br>(0.203) | 15,682.170 |
| **Media coverage** | .0003<br>(0.016) ** | .0001 | -.0002<br>(0.583) | .0004 |
| **Crimes** | .0001<br>(0.176) | .00009 | .0003<br>(0.155) | .0002 |
| **Stops** | 1.77e-07<br>(0.950) | 2.82e-06 | -.00003<br>(0.316) | .00003 |
| **Arrests** | .0001 *<br>(0.072) | .00007 | .00002<br>(0.532) | .00003 |
| **Constant** | -14.640<br>(0.481) | 20.756 | -131.948<br>(0.189) | 100.505 |

p-values in brackets with * = p < 0.1; ** = p < 0.05; *** = p < 0.000

## Analysis

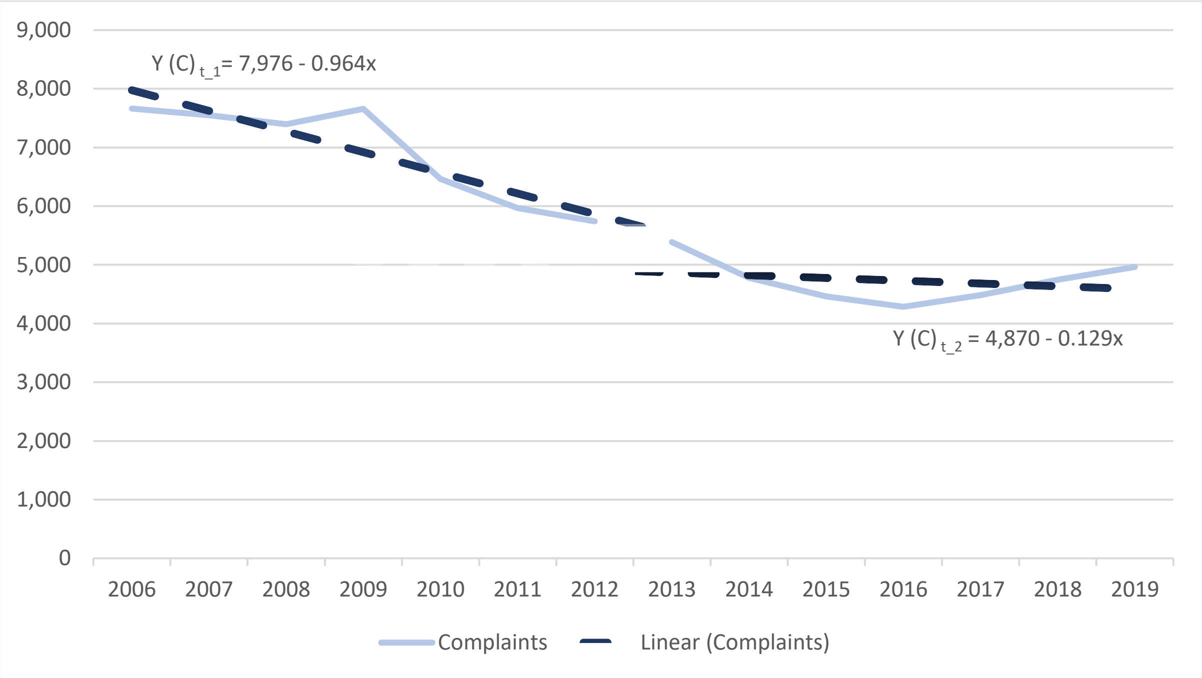
In this section, the empirical findings of the paired-samples t-test as well as the GLS regression are analysed regarding what they suggest about the validity of H1 and H2. Subsequently, possible explanations for the findings as well as the limitations of the research are discussed. Regarding H1 the findings run counter to the assumption that predictive policing would fundamentally harm citizen perceptions of procedural justice in their interactions with police. However, there is evidence that certain aspects such as participation and respectful and dignified treatment might be harmed. The findings also demonstrate that a person's race has an impact on their likelihood to perceive a lower sense of procedural justice during police encounters. However, contrary to H2 an unequivocal negative effect of predictive policing on BPoC cannot be detected.

### Analysis of the paired-samples t-test

The paired-samples t-test was used to investigate, whether there is a significant difference in the overall number of complaints from t_1 to t_2 and to test for H1. The results show that there is a statistically significant difference between the period prior to the introduction of predictive policing and the period post introduction. The mean difference between total complaints is higher in t_1 than in t_2, although the size of the effect is small. This result contradicts H1 and the assumption that total complaints would rise with predictive policing. Below a graphical analysis was conducted to investigate the difference and overall trend more thoroughly.

In addition to the total complaints per year the line chart below includes an estimation of the linear trends for both periods. The linear trendline takes the form $Y = a + bX$ and shows the predicted development of complaint numbers. For t_1 the complaints declined at a rate of 0.964. The decline was 0.835 points higher than the slope in t_2 of b = 0.129. This indicates, that while the average number of complaints was higher much higher in t_1 than in t_2, the linear trend changed from a steep decline to a much lower decrease. This is attributable to the overall number of complaints reaching a minimum in 2016 and then rising in subsequent years. Nonetheless, it is unclear if the stagnation of the rate of decrease in complaints is attributable to the introduction of predictive policing. One potential factor could be the city-wide rollout of predictive software for mobile devices from June 2015 to April 2016 (Levine et al., 2017). This could have significantly increased the impact of predictive policing on an individual officer's decision-making. However, without precise data on the influence of different iterations of predictive tools on policing, this remains speculation.

*Figure 4: Total complaints per year*




Source: Own depiction based on CCRB (2021b)

A potential influencing factor on citizen perceptions of police procedural justice is the election of Mayor Bill de Blasio in January 2014 and subsequent police reforms. First, the reduction of the city's Stop and Frisk programme and the increase of body-worn cameras on police officers. The death of Eric Garner at the hands of NYPD officers in July 2014 led to further policy changes. Such as a three-day retraining course on street-level interactions, including training for officers on how to assume a non-judgmental demeanour. (Santora, 2014) Further reforms were enacted through the Criminal Justice Reform Act, signed into law in June 2016. The law introduced civil in place of criminal penalties for low-level offences and increased transparency by requiring the NYPD to disclose both civil and criminal summonses (The City of New York, 2016).

What stands out, is the fact, that there are more police officers per citizen, higher crime rates, more stops by police and arrests in t_1 compared to t_2. All factors affect the number of police-citizen interactions and thus should influence the likelihood of negative encounters. As such, it is surprising that the rate of police complaints did not continue its sharp decline in t_2 but the overall linear trend stagnated. In this case, the introduction of predictive policing could play a role in entrenching established patterns of policing and making the system resistant to reforms designed to increase procedural justice (Ferguson, 2017). Due to its reliance on historic data, including datasets compiled during the height of Stop and Frisk, the software has the potential to reinforce outdated policing practices, despite apparent changes at the policy level (Griffard, 2019). This could explain why even the reforms implemented by Mayor de Blasio seem to have had little impact on reducing the average number of citizen complaints in the long term.

## Analysis of the results of the GLS regression

Turning to the results of the GLS regression and the corresponding descriptive statistics, a similarly nuanced picture emerges. Overall, the descriptive statistics show that the percentage of complaints by Black in the sample is almost three times higher than their share in the overall population. In comparison to the comprehensive statistics released by the CCRB (2021b), Black people are slightly overrepresented in the sample (see appx. 10, 11). While White and Hispanic complainants are slightly underrepresented in terms of their share in the total number of complaints. However, even in the CCRB (2021b) statistics Black people represent roughly 57% of complainants in t_1 and 53% in t_2, more than double their share in the population. With 13% in t_1 and 14% in t_2 White complainants make up less than half of their share of the population and the proportion of Hispanic complainants at 25% (t_1) and 26% (t_2) remains slightly lower than their overall population share. Thus, the sample does accurately represent the high share of Black complainants and its overall characteristics mirror those of the CCRB's overview. However, the effect on the Black population might be slightly overstated in the ProPublica (2020) database used for the GLS regression.

These discrepancies stem from the fact that ProPublica's data only includes complaints made against officers that had at least one substantiated claim against them. However, this could increase the likelihood that the complaints in the sample depict actual police transgressions since they are not simply a record of all complaints made regardless of validity. Potentially making the sample a better indicator of actual citizen perceptions of procedural justice. Another aspect that could affect the results depicted here is the likely differing reporting ratios between racial groups. Peck's (2015) meta-analysis of 92 articles, shows that Black people are more likely to hold a negative view of police, followed by Hispanic and then White people. Thus, BPoC might be more likely to refrain from making complaints, on account of their already low perception of police. Another hurdle to making a complaint could be the language barrier. This is especially a factor in the Hispanic community, with roughly 24.10% of New Yorkers reporting Spanish as their primary language according to the 2019 American Community Survey (ACS). Both low confidences in police and language barriers are factors that must be considered when concluding complaint statistics.

The descriptive statistics show, that while the mean number of complaints of White and Asian individuals increased, those of Black and Hispanic individuals slightly dropped after the introduction of predictive policing. However, even in t_2 Black people still make up a significant majority of complaints, while White people are underrepresented in relation to their share in the overall population. This could be partially attributed to a change in demographics. For a lack of availability of population statistics from the yet to be released 2020 US Census, the results of the previous census are consulted. Historic trends indicate that the percentage of White and Black people is likely to decline, while that of Hispanics and Asians should rise (The Furman Center, 2012). This could be a partial explanation, why total complaints by Asians went up in t_2. For the dependent variable allegations, values increased across the four racial groups after 2013. The number of allegations per complaint is highest for Hispanics, followed by Black people, then Asians and finally White citizens. This indicates that single complaints were more severe compared to those during the pre-predictive policing period. Absent the effect of predictive policing another possible explanation for this result is an increased push for police accountability. Especially the creation of the BLM movement in 2014 raised public awareness of police misconduct (AP-NORC, 2020). Consequently, individuals might be more knowledgeable about the type of police actions, they can file a complaint against.

In the descriptive overview of the control variables a drastic decline in police stops and thus potentially police-citizen interaction is especially noticeable. This can be attributed to the fact, that the NYPD's Stop and Frisk programme was at its height under Michael Bloomberg's terms as mayor of New York City from 2002 to 2013 (Southall & Gold, 2019). The use of the practice sharply declined after a series of lawsuits and the subsequent implementation of court-ordered reforms in 2013. Statistics of police stops show a similar overrepresentation of Black people as in the complaint statistics. Despite the strong decline in police stops, the racial makeup of the people stopped does not change significantly from t_1 to t_2 (NYCLU, 2019). However, running the regression including the number of stops by race did not influence the results of the regression.

The GLS regression for complaints reveals a strong effect of a person's race on the number of complaints that get filed, the only exception is Asian citizens. A possible reason is that overall complaint numbers among Asian citizens are quite low, leading to a small number of available data points. At the outset, the effect is by far strongest for Black people, roughly 10 times higher than that for White citizens. Hispanics experience the second-highest effect. Their complaint coefficient is three times higher than that of White people.

However, the change from t_1 to t_2 only partially supports the hypothesis that overall complaints will increase with the introduction of predictive policing, thus having an overall negative effect on perceptions of procedural justice. For t_2 only the effect in the category White was higher, while it decreased for both Black and Hispanic individuals. This finding also contradicts H2, according to which predictive policing will lead to increased negative police encounters with BPoC and thus disproportionately affecting their perception of police procedural justice. However, delving into the allegations that make up the individual complaints paints a slightly more nuanced picture. From t_1 to t_2 the effect of race on the number of complaints increases for all racial groups. This increase was slightly higher for Black and Hispanic citizens than for White people. This points to the previously mentioned increased severity of the complaints. It could indicate, that while the overall number of complaints might be slightly lower their severity shows that individual encounters are perceived as more negative and thus have a stronger negative effect on perceptions of procedural justice.

The examination of the effect on the dispersion of allegations across the FADO categories, reveals that the likelihood for Black and Hispanic people to file a complaint of force is slightly lower in t_2, while White people experienced an increase. However, the overall effect of the categories Black and Hispanic remains much higher than that for White people. Thus, the perception of procedural justice, especially regarding the aspects of trust in motives and dignity and respect, is likely lower among White citizens in t_2. A violent altercation, a complainant’s views on how police centre citizen wellbeing in their decision-making could be negatively impacted. Secondly, the use of force constitutes an infringement on the polite and dignified treatment that citizens expect from the police.

Abuse of authority allegations increased across all categories. However, the increase was roughly two times higher for Black citizens and one point five times for Hispanic people than for White people. This points to the fact, that following the introduction of predictive policing, BPoC were generally more likely to view their interactions with police as lacking in the aspect of quality of interpersonal treatment, i.e. respect and dignity, and to perceive officers as less responsive to their input. It could also point towards a loss of trust in the motives of police officers. For allegations of discourtesy, the effect for White citizens decreased slightly, while Black and Hispanic people experienced an increase. Again, this could point to the fact that BPoC citizens see a decline in procedural justice when it comes to the aspect of receiving dignified and respectful treatment. Lastly, the dependent variable offensive language also shows a slight increase across the categories. However, the effect for the category White remains negligible. The lack of effect for White people can be explained by the fact, that the largest share of allegations of this kind pertains to the use of offensive slurs based on race or ethnicity. Thus, it is likely that Black and Hispanic citizens view their interactions with police as less neutral in t_2 than in t_1. Indicating that they perceive police officers to act in a biased way and that they act on personal views, rather than objective guidelines.

## Limitations

The first limitation of the research is connected to the available data. To assess the overall effect of predictive policing on citizen perceptions of procedural justice both positive and negative experiences need to be considered. Using complaints limits the conclusions drawn from the research to the impact on direct police-citizen encounters that were perceived as negative. For a more holistic picture qualitative measurements, such as surveys or interviews would be more suitable. This approach could capture how civilians think about predictive policing in the context of procedural justice, regardless of an actual shift in police-citizen interactions. Furthermore, to prepare the data, complaints with missing values in the categories race and date were deleted. It is unclear whether certain types of allegations are overrepresented within this sample, e.g. if incidents involving the use of force are less likely to be fully documented. Another limitation is related to selection bias. ProPublica’s (2020) database only includes reports against officers that had at least one substantiated complaint against them. They might exhibit more aggressive behaviours than the average NYPD officer. Thus, the ratio of allegations to complaints might be overstated in the dataset.

Additional limitations arise from the lack of information that police departments release about the use of predictive software. This makes the choice of a breakpoint very difficult. The exact timeline of the NYPD's software rollout is not public knowledge and likewise it is not clear how subsequent additional features impacted the software's performance. Although the initial rollout of DAS occurred in 2013, the mobile version of the software was implemented from June 2015 to April 2016. From then on, every officer across the city was able to access the software on their mobile phone or tablet. This might have impacted the software's influence over an individual officer's decision-making. This could lead to omitted variable bias because the technology's adoption rate and impact are not included in the model. In addition, if departments were transparent about the sequence of the technology's implementation analysis on the precinct level could be carried out. With one precinct functioning as a control group and the other as a treatment group. Using two very similar precincts would mitigate the effect of policing reforms, such as those introduced in 2014, the effect of which on complaint rates is hard to estimate.

## Conclusion

"Police have entered the age of actuarial justice and (….) there is no real hope of going back." (Ferguson, 2017, p. 1189). As such it is of crucial importance, to understand how predictive algorithms not only affect the outcome of policing but also the process. Regarding the question of procedural justice, the findings at hand paint a less grim picture than previous research leads one to believe. A clear negative effect of predictive policing on New York City residents' perceptions of procedural justice during police encounters cannot be established. However, there is evidence that some of the key aspects of citizen perceptions of procedural justice defined by Tyler (2004) could be negatively impacted. Especially, the elements of participation, trust as well as respect and dignity, might be harmed.

Furthermore, the analysis demonstrates that a person's race impacts how they perceive procedural fairness during encounters with police. This confirms the findings presented in Bolger et al.'s (2021) meta-analysis. In this case, the results of the random-effects model are significant for White, Black, and Hispanic citizens. Especially Black people are much more likely to perceive low procedural fairness during police encounters than White citizens. However, the introduction of predictive policing does not seem to have had a disproportionality negative effect on perceptions of procedural justice among Black and Hispanic individuals. Despite the prevalence of so-called problematic policing practices such as Stop and Frisk, which are racially biased and embedded in the NYPD's data (Richardson et al., 2019). An interesting aspect is, that complaints seem to have increased in severity after the introduction of predictive policing. Pointing to the fact that while overall negative encounters decreased, individuals might view procedural justice in single encounters more negatively. For Black and Hispanic people this is due to a higher likelihood of allegations of discourtesy and offensive language compared to the previous period. Possibly indicating, that they perceive a lack of dignified, respectful and unbiased treatment (Tyler, 2004).

Further research should be conducted to explore the relationship between predictive policing and different aspects of procedural justice more thoroughly. First, a qualitative approach could be used, to capture a potentially positive impact on citizen perceptions of procedural justice. As Piotrowicz (2019) describes, citizens view decisions made by algorithms as more objective than those coming from the officers themselves. In addition, surveying civilians and police would reveal any differences in expectation regarding the practice's impact on procedural justice. For example, officers might feel that predictive policing enables them to employ a more targeted and neutral approach, while citizens possibly perceive a lack of transparency as harmful to procedural justice. In this context, it would also be interesting to consider, how citizens feel about different iterations of predictive technology, i.e. if the level of intrusion and data collection has any impact on their perceptions of police procedural justice. On a more quantitative level subsequent research could utilize the fact that the rollout of predictive software is often done sequentially, to compare precincts with predictive policing to those without. Similar to research Gelman et al. (2007) conducted on the practice of Stop and Frisk.

Even though suppliers of predictive software such as PredPol, HunchLab, Microsoft, or Palantir continue to stress the neutrality of their algorithms public perception has changed (Edwards, 2016). In the wake of the most recent protests of police brutality, further regulations have been adopted in the US. Extreme measures, such as a ban on predictive policing in Santa Cruz, signify that politicians have renewed interest in demonstrating to their constituents that they are attuned to their concerns (Asher-Schapiro, 2020). However, companies and police departments still mostly refuse to disclose which algorithms and datasets are being used. As a first step to enhance procedural justice in the age of predictive policing policy-makers should push for the disclosure of both the technology being used and the data being processed. The issue of transparency is strongly connected to the developer's profit motive, which leads them to be reluctant to share specifics about their software. Thus, strict regulations of police departments' private-public partnerships with software companies are necessary. A possible measure could be to require police departments to move from closed to open source solutions, allowing independent researchers to evaluate the tools. In addition, oversight bodies, such as the CCRB, need to be empowered and equipped with the necessary tools to force police departments to cooperate with investigations. Furthermore, the public should have a veto regarding the predictive tools that police use in their neighbourhoods. Software solutions should be evaluated carefully and on a case-by-case basis regarding their potential effect on data privacy and discriminatory practices in policing. In conclusion, a fundamental shift in the mindset of politicians and law enforcement is needed. The current model of “invention first, then adoption, and finally assessment only after the fact” (Ferguson, 2017, p. 1189) has the potential to harm people’s perceptions of police procedural justice and to damage public trust in law enforcement in the long term.

## References


ACLU. (2016). Statement of Concern About Predictive Policing by ACLU and 16 Civil Rights Privacy, Racial Justice, and Technology Organizations. https://www.aclu.org/other/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice

ACLU. (2020). *Community Control Over Police Surveillance.* American Civil Liberties Union. https://www.aclu.org/issues/privacy-technology/surveillance-technologies/community-control-over-police-surveillance

AI Now Institute. (2018). *Automated Decision Systems.* AI Now Institute. https://ainowinstitute.org/nycadschart.pdf

American Community Survey. (2019). *U.S. Census Bureau: New York city, New York.* U.S. Census Bureau. https://www.census.gov/quickfacts/newyorkcitynewyork

American Legal. (2021). *Chapter 18-A: Civilian Complaint Review Board*. https://codelibrary.amlegal.com/codes/newyorkcity/latest/NYCcharter/0-0-0-1642

AP-NORC. (2020). *Significant Shifts in Attitudes on Race and Policing.* AP-NORC. https://apnorc.org/projects/significant-shifts-in-attitudes-on-race-and-policing/

Asher-Schapiro, A. (2020, June 24). California city bans predictive policing in U.S. first. *Reuters Media*. https://www.reuters.com/article/us-usa-police-tech-trfn-idUSKBN23V2XC

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2018). *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung* (15., vollständig überarbeitete Auflage). Springer. https://doi.org/10.1007/978-3-662-56655-8

Beam, C. (2011, January 25). Predictive policing LAPD: Can police really predict crime before it happens? *Slate*. https://slate.com/news-and-politics/2011/01/predictive-policing-lapd-can-police-really-predict-crime-before-it-happens.html

Biddle, S. (2020, June 3). Amazon “Stands in Solidarity” Against Police Racism While Selling Racist Tech to Police. *The Intercept*. https://theintercept.com/2020/06/03/amazon-police-racism-tech-black-lives-matter/

Bolger, C. P., & Walters, G. D. (2019). The relationship between police procedural justice, police legitimacy, and people's willingness to cooperate with law enforcement: A meta-analysis. *Journal of Criminal Justice*, *60*, 93–99. https://doi.org/10.1016/j.jcrimjus.2019.01.001

Bolger, M. A., Lytle, D. J., & Bolger, C. P. (2021). What matters in citizen satisfaction with police: A meta-analysis. *Journal of Criminal Justice*, *72*, 101760. https://doi.org/10.1016/j.jcrimjus.2020.101760

Bond-Graham, D. (2014, February 27). Forget the NSA, the LAPD Spies on Millions of Innocent Folks. *LA Weekly*. http://www.laweekly.com/forget-the-nsa-the-lapd-spies-on-millions-of-innocent-folks/

Brantingham, P. J., Valasik, M., & Mohler, G. (2018). Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial. *Statistics and Public Policy*, *5*(1), 1–6. https://doi.org/10.1080/2330443X.2018.1438940

Byrne, J., & Marx, G. (2011). Technological Innovations in Crime Prevention and Policing: A Review of the Research on Implementation and Impact. *Cahiers Politiestudies*, *3*(20), 17–40. https://www.ncjrs.gov/pdffiles1/nij/238011.pdf

Cao, L., Frank, J., & Cullen, F. T. (1996). Race, community context and confidence in the police. *American Journal of Police*, *15*(1), 3–22. https://doi.org/10.1108/07358549610116536

Carter, R. E. (2013). A standard error: Distinguishing standard deviation from standard error. *Diabetes*, *62*(8), e15. https://doi.org/10.2337/db13-0692

Casady, T. (2011). Police Legitimacy and Predictive Policing. *Geography & Public Safety*, *2*(4), 1–2. https://cops.usdoj.gov/RIC/Publications/cops-w0598-pub.pdf

Castelvecchi, D. (2020, June 19). Mathematicians urge colleagues to boycott police work in wake of killings. *Nature*. https://www.nature.com/articles/d41586-020-01874-9

CCRB. (2021a). *Allegations.* CCRB. https://www1.nyc.gov/site/ccrb/policy/data-transparency-initiative-allegations.page

CCRB. (2021b). *Complaints*. https://www1.nyc.gov/site/ccrb/policy/data-transparency-initiative-complaints.page

Chan, J., & Bennett Moses, L. (2016). Is Big Data challenging criminology? *Theoretical Criminology*, *20*(1), 21–39. https://doi.org/10.1177/1362480615586614

Charles, B. J. (2019, March 26). NYPD's Big Artificial-Intelligence Reveal. *Governing*. https://www.governing.com/archive/gov-new-york-police-nypd-data-artificial-intelligence-patternizr.html

Chohlas-Wood, A., & Levine, E. S. (2019). A Recommendation Engine to Aid in Identifying Crime Patterns. *INFORMS Journal on Applied Analytics*, *49*(2), 1–11. https://doi.org/10.1287/inte.2019.0985

The City of New York. (2016). *Mayor de Blasio Signs the Criminal Justice Reform Act.* The City of New York. https://www1.nyc.gov/office-of-the-mayor/news/530-16/mayor-de-blasio-signs-criminal-justice-reform-act

Dai, M., Frank, J., & Sun, I. (2011). Procedural justice during police-citizen encounters: The effects of process-based policing on citizen compliance and demeanor. *Journal of Criminal Justice*, *39*(2), 159–168. https://doi.org/10.1016/j.jcrimjus.2011.01.004

Edwards, E. (2016). *Predictive Policing Software Is More Accurate at Predicting Policing Than Predicting Crime.* ACLU. https://www.aclu.org/blog/criminal-law-reform/reforming-police/predictive-policing-software-more-accurate-predicting

Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018). Runaway Feedback Loops in Predictive Policing. *Conference on Fairness, Accountability and Transparency*, 160–171. http://proceedings.mlr.press/v81/ensign18a.html

Feiner, L. (2020, June 18). NYC lawmakers pass bill requiring police to disclose surveillance technology. *CNBC*. https://www.cnbc.com/2020/06/18/nyc-passes-bill-requiring-police-to-disclose-surveillance-technology.html

Ferguson, A. G. (2017). Policing Predictive Policing. *Washington University Law Review*, *94*(5), 1109–1189. https://openscholarship.wustl.edu/lawlawreview/vol94/iss5/5

Ferguson, A. G. (2018). *The High-Definition, Artificially Intelligent, All-Seeing Future of Big Data Policing.* American Civil Liberties Union. https://www.aclu.org/issues/privacy-technology/surveillance-technologies/high-definition-artificially-intelligent-all

Frank, J., Smith, B. W., & Novak, K. J. (2005). Exploring the Basis of Citizens' Attitudes Toward the Police. *Police Quarterly*, *8*(2), 206–228. https://doi.org/10.1177/1098611103258955

The Furman Center. (2012). *The Changing Racial and Ethnic Makeup of New York City Neighborhoods.* The Furman Center. https://furmancenter.org/files/sotc/The_Changing_Racial_and_Ethnic_Makeup_of_New_York_City_Neighborhoods_11.pdf

Gelman, A., Fagan, J., & 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. https://doi.org/10.1198/016214506000001040

Gilbertson, A. (2020, August 20). Data-Informed Predictive Policing Was Heralded As Less Biased. Is It? – The Markup. *The Markup*. https://themarkup.org/ask-the-markup/2020/08/20/does-predictive-police-technology-contribute-to-bias

Goode, E. (2011, August 15). Sending the Police Before There’s a Crime. *The New York Times*. https://www.nytimes.com/2011/08/16/us/16police.html

Greene, W. H. (2008). *Econometric analysis* (6. ed.). Pearson Prentice Hall. http://worldcatlibraries.org/wcpa/oclc/137325275

Griffard, M. (2019). A Bias-Free Predictive Policing Tool? An Evaluation of the NYPD's Patternizr. *Fordham Urban Law Journal*, *47*(1). https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=2779&context=ulj

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (Second edition). SAGE Publications.

Harcourt, B. E. (2007). Against prediction: Profiling, policing, and punishing in an actuarial age. University of Chicago Press.

Haskins, C. (2019, February 6). Dozens of Cities Have Secretly Experimented With Predictive Policing Software. *Vice*. https://www.vice.com/en/article/d3m7jq/dozens-of-cities-have-secretly-experimented-with-predictive-policing-software

Heaven, W. D. (2020, July 17). Predictive policing algorithms are racist. They need to be dismantled. *MIT Technology Review*. https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

Hill, E., & Stein, R. (2020, May 31). 8 Minutes and 46 Seconds: How George Floyd Was Killed in Police Custody. *The New York Times*. https://www.nytimes.com/2020/05/31/us/george-floyd-investigation.html

Kaplan, J. (2021). Jacob Kaplan's Concatenated Files: Uniform Crime Reporting (UCR) Program Data: Arrests by Age, Sex, and Race, 1974-2018. https://doi.org/10.3886/E102263V11-74340

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Leventhal, G. S. (1980). What Should Be Done with Equity Theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27–55). Plenum Press. https://doi.org/10.1007/978-1-4613-3087-5_2

Levine, E. S., Tisch, J., Tasso, A., & Joy, M. (2017). The New York City Police Department’s Domain Awareness System. *Interfaces*, *47*(1), 1–15. https://doi.org/10.1287/inte.2016.0860

Lind, E. A., & Tyler, T. R. (Eds.). (1988). *Critical Issues in Social Justice*. *The Social Psychology of Procedural Justice*. Springer. https://doi.org/10.1007/978-1-4899-2115-4

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, *13*(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Mayor’s Office of Operations. (2020). *Mayor’s Management Report: Fiscal 2020.* Mayor’s Office of Operations. https://www1.nyc.gov/assets/operations/downloads/pdf/mmr2020/2020_mmr.pdf

Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of Benefits and Drawbacks. *International Journal of Public Administration*, *42*(12), 1031–1039. https://doi.org/10.1080/01900692.2019.1575664

Newcombe, T. (2014). *What Predictive Policing Can, and Can't, Do to Prevent Crime.* Governing. https://www.governing.com/news/headlines/what-predictive-policing-can-and-cant-do-to-prevent-crime.html

Norton, A. (2013). Predictive Policing: The Future of Law Enforcement in the Trinidad and Tobago Police Service (TTPS). *International Journal of Computer Applications*, *62*(4), 32–36. https://doi.org/10.5120/10070-4680

NYCLU. (2019). *Stop-and-Frisk Data*. https://www.nyclu.org/en/stop-and-frisk-data

NYPD. (2020). *Crime & Enforcement Activity Reports*. https://www1.nyc.gov/site/nypd/stats/reports-analysis/crime-enf.page

NYS DCJS. (2021). *Criminal Justice Reports & Statistics.* NYS DCJS. https://www.criminaljustice.ny.gov/crimnet/ojsa/stats.htm

Orr, B. (2012, December 4). LAPD computer program prevents crime by predicting it. *CBS News*. https://www.cbsnews.com/news/lapd-computer-program-prevents-crime-by-predicting-it/

Patterson, R. (2006). Resolving Civilian-Police Complaints in New York City: Reflections on Mediation in the Real World. *Scholarly Works*, Article 493. http://scholars.law.unlv.edu/facpub/493

Peck, J. H. (2015). Minority perceptions of the police: a state-of-the-art review. *Policing: An International Journal of Police Strategies & Management*, *38*(1), 173–203. https://doi.org/10.1108/PIJPSM-01-2015-0001

Perry, W. L., McInnis, B., Price, C. C., Smith, S., & Hollywood, J. S. (Eds.). (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation.

Piotrowicz, C. (2019). Predictive Policing. *European Law Enforcement Research Bulletin*(4), 107–111. https://bulletin.cepol.europa.eu/index.php/bulletin/article/view/374

ProPublica. (2020). *Civilian Complaints Against New York City Police Officers*. https://www.propublica.org/datastore/dataset/civilian-complaints-against-new-york-city-police-officers

Public Safety Lab. (2021, February 21). *Analysis of NYPD Officer Misconduct Complaint Data.* NYU. https://github.com/publicsafetylab/public-psl-ccrb

Puente, M. (2019, March 7). LAPD pioneered predicting crime with data. Many police don't think it works. *Los Angeles Times*. https://www.latimes.com/local/lanow/la-me-lapd-precision-policing-data-20190703-story.html

Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review*, *94*(192), 192–233.

Rosenbaum, D. P., Schuck, A. M., Costello, S. K., Hawkins, D. F., & Ring, M. K. (2005). Attitudes Toward the Police: The Effects of Direct and Vicarious Experience. *Police Quarterly*, *8*(3), 343–365. https://doi.org/10.1177/1098611104271085

Santora, M. (2014, December 4). Mayor de Blasio Announces Retraining of New York Police. *The New York Times*. https://www.nytimes.com/2014/12/05/nyregion/mayor-bill-de-blasio-retraining-new-york-police-dept-eric-garner.html

Saunders, J., Hunt, P., & Hollywood, J. S. (2016). Predictions put into practice: A quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology*, *12*(3), 347–371. https://doi.org/10.1007/s11292-016-9272-0

Schafer, J. A., Huebner, B. M., & Bynum, T. S. (2003). Citizen Perceptions of Police Services: Race, Neighborhood Context, and Community Policing. *Police Quarterly*, *6*(4), 440–468. https://doi.org/10.1177/1098611102250459

Selbst, A. D. (2017). Disparate Impact in Big Data Policing. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2819182

Silva, S., & Kenney, M. (2018). Algorithms, Platforms, and Ethnic Bias: An Integrative Essay. *Phylon (1960-)*, *55*(1 & 2), 9–37. https://www.jstor.org/stable/10.2307/26545017

Southall, A., & Gold, M. (2019, November 17). Why 'Stop-and-Frisk' Inflamed Black and Hispanic Neighborhoods. *The New York Times*. https://www.nytimes.com/2019/11/17/nyregion/bloomberg-stop-and-frisk-new-york.html

Strikwerda, L. (2020). Predictive policing: The risks associated with risk assessment. *The Police Journal: Theory, Practice and Principles*, 0032258X2094774. https://doi.org/10.1177/0032258X20947749

Sturgill, K. (2020, June 26). Santa Cruz becomes the first U.S. city to ban predictive policing. *Los Angeles Times*. https://www.latimes.com/california/story/2020-06-26/santa-cruz-becomes-first-u-s-city-to-ban-predictive-policing

Tayebi, M. A., & Glässer, U. (2016). Social Network Analysis in Predictive Policing: Concepts, Models and Methods. Lecture Notes in Social Networks. Springer. https://doi.org/10.1007/978-3-319-41492-8

Thibaut, J. W., & Walker, L. (1975). Procedural justice: A psychological analysis. https://lib.ugent.be/catalog/rug01:001876843

TIME (2011, November 18). The 50 Best Inventions. http://content.time.com/time/magazine/article/0,9171,2099708,00.html

Tyler, T. R. (1988). What is Procedural Justice? Criteria used by Citizens to Assess the Fairness of Legal Procedures. *Law & Society Review*, *22*(1), 103. https://doi.org/10.2307/3053563

Tyler, T. R. (1990). *Why people obey the law*. Princeton University Press. http://www.loc.gov/catdir/enhancements/fy0654/2005938383-d.html

Tyler, T. R. (2004). Enhancing Police Legitimacy. *The ANNALS of the American Academy of Political and Social Science*, *593*(1), 84–99. https://doi.org/10.1177/0002716203262627

Tyler, T. R., & Huo, Y. J. (2002). Trust in the law: Encouraging public cooperation with the police and courts. The Russell Sage Foundation Series on Trust. Russell Sage Foundation. http://www.jstor.org/stable/10.7758/9781610445429

U.S. Census Bureau. (2011). *2010 Census Redistricting Data (Public Law 94-171) Summary File.* U.S. Census Bureau. https://assets.documentcloud.org/documents/87708/pl94-171-1.pdf

Uchida, C. (2009). A national discussion on predictive policing defining our terms and mapping successful implementation strategies. Washington D.C. http://worldcatlibraries.org/wcpa/oclc/665183357

Umansky, E., & Simon, M. (2020, August 17). The NYPD Is Withholding Evidence From Investigations Into Police Abuse. *ProPublica*. https://www.propublica.org/article/the-nypd-is-withholding-evidence-from-investigations-into-police-abuse

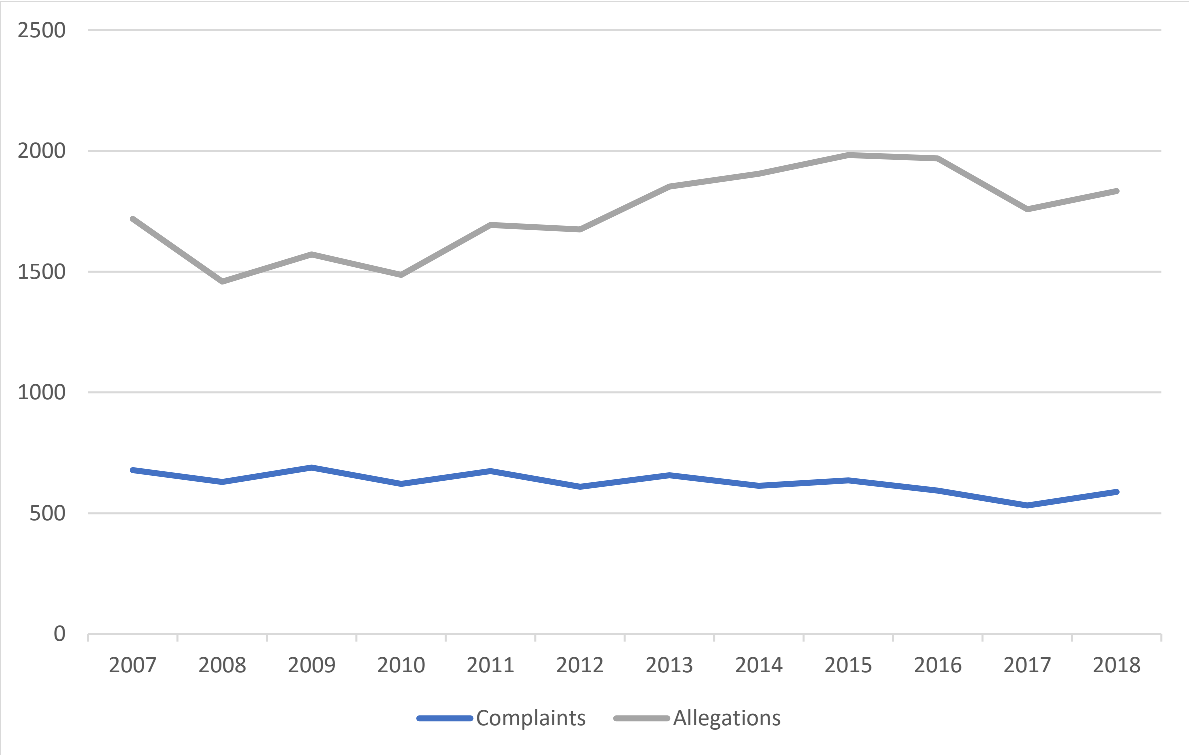
Ungerleider, N. (2012, August 8). NYPD, Microsoft Launch All-Seeing “Domain Awareness System” With Real-Time CCTV, License Plate Monitoring [Updated]. *Fast Company*. https://www.fastcompany.com/3000272/nypd-microsoft-launch-all-seeing-domain-awareness-system-real-time-cctv-license-plate-monito

Weitzer, R., Tuch, S. A., & Skogan, W. G. (2008). Police–Community Relations in a Majority-Black City. *Journal of Research in Crime and Delinquency*, *45*(4), 398–428. https://doi.org/10.1177/0022427808322617

Winston, A. (2018, February 27). Palantir has secretly been using New Orleans to test its predictive policing technology. *The Verge*. https://www.theverge.com/2018/2/27/17054740/palantir-predictive-policing-tool-new-orleans-nopd

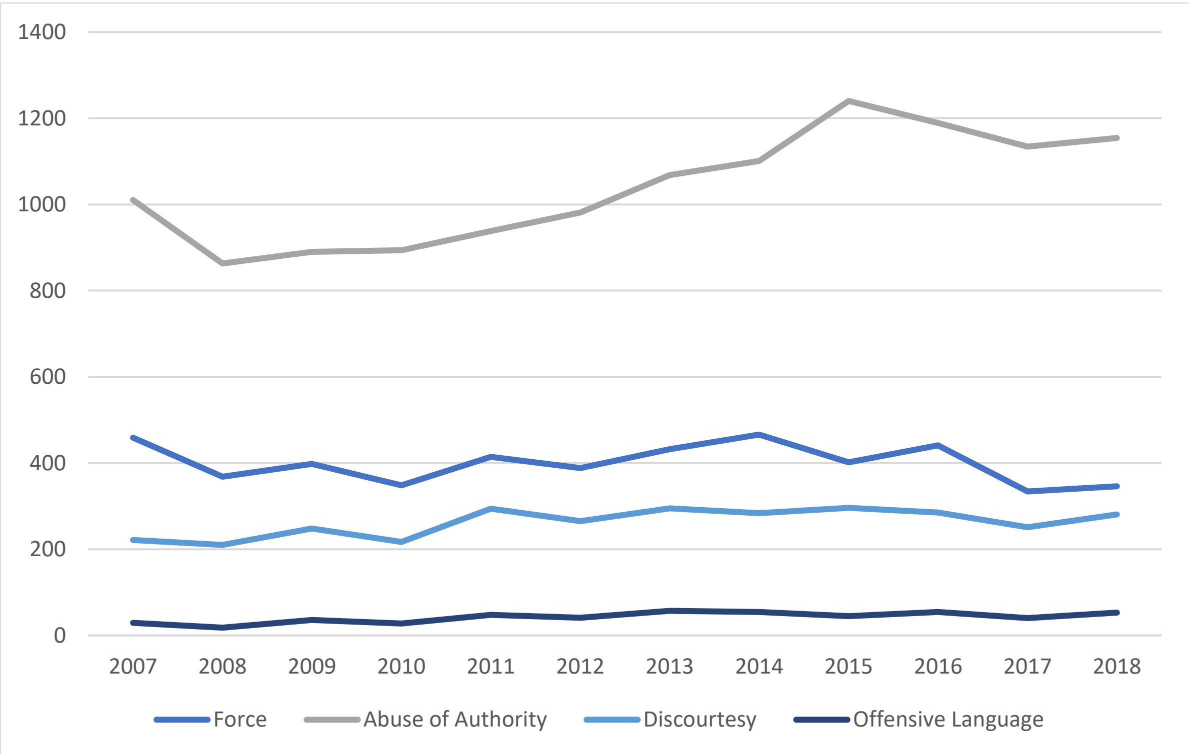
## Appendix

*Appendix 1: Development of complaints and allegations*




Source: Own depiction based on ProPublica (2020)

*Appendix 2: Development of allegations by FADO type*




Source: Own depiction based on ProPublica (2020)

*Appendix 3: Values of control variables by year*

| Year | Population | Officers | Media coverage | Crimes | Stops | Arrests |
|---|---|---|---|---|---|---|
| 2007 | 8,220,196 | 35,404 | 6,992 | 199,941 | 472,096 | 334,208 |
| 2008 | 8,345,075 | 35,761 | 5,309 | 198,419 | 540,302 | 333,428 |
| 2009 | 8,400,907 | 35,071 | 4,343 | 188,357 | 581,168 | 340,986 |
| 2010 | 8,175,133 | 34,817 | 5,516 | 188,104 | 601,285 | 343,294 |
| 2011 | 8,211,875 | 34,542 | 7,104 | 191,666 | 685,724 | 338,551 |
| 2012 | 8,289,415 | 34,555 | 8,504 | 195,753 | 532,911 | 325,205 |
| 2013 | 8,396,126 | 34,822 | 7,881 | 194,367 | 191,851 | 318,518 |
| 2014 | 8,473,938 | 34,581 | 10,321 | 186,334 | 45,787 | 286,694 |
| 2015 | 8,550,861 | 35,395 | 10,333 | 179,957 | 22,563 | 255,479 |
| 2016 | 8,566,917 | 36,228 | 10,696 | 174,407 | 12,405 | 249,459 |
| 2017 | 8,616,333 | 36,378 | 7,058 | 171,248 | 11,629 | 226,799 |
| 2018 | 8,523,171 | 36,134 | 6,335 | 174,164 | 11,008 | 197,358 |

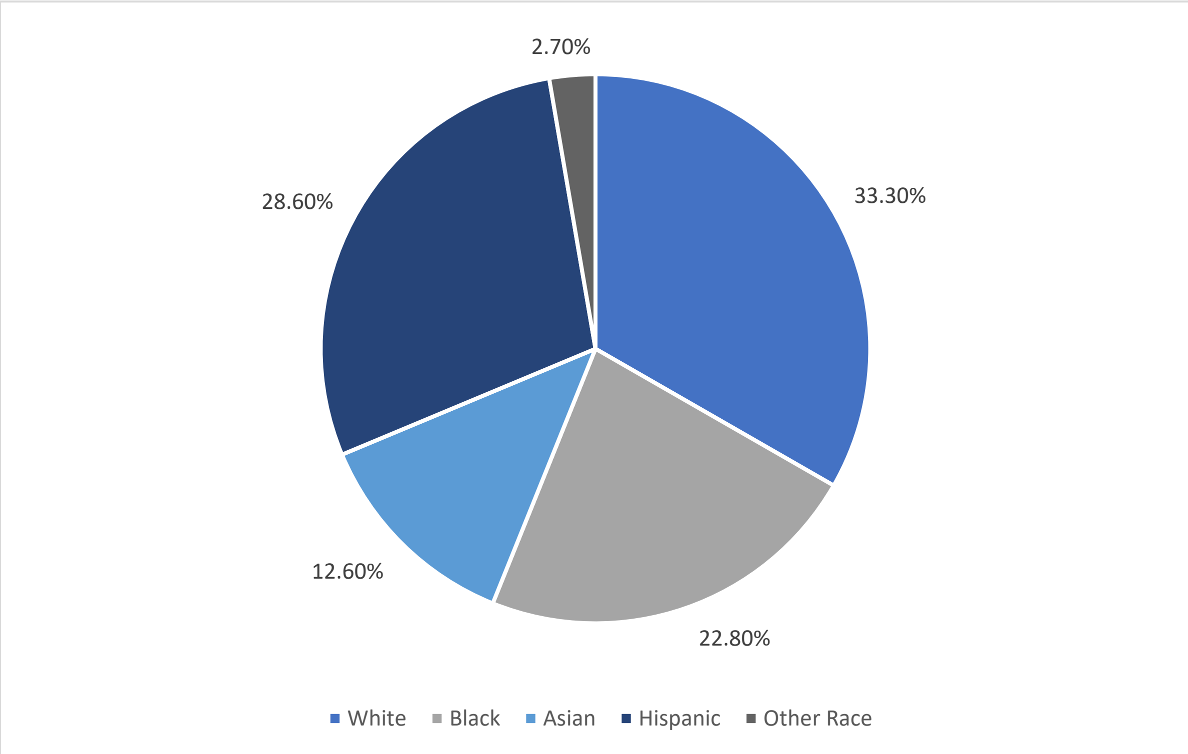
Source: Own depiction based on Kaplan (2021), NYPD (2020) and NYS DCJS (2021)

*Appendix 4: NYPD officers by race and gender in %*

| | Average t_1 | Average t_2 |
|---|---|---|
| **White** | 53.17 | 50.43 |
| **Black** | 16.27 | 15.38 |
| **Hispanic** | 25.43 | 27.18 |
| **Asian** | 4.70 | 6.95 |
| **Other Race** | 0.42 | 0.10 |
| **Male** | 83.00 | 82.67 |
| **Female** | 17.00 | 17.33 |
| **Officers per 100,000 citizens** | 423.35 | 417.63 |

Source: Own depiction based on NYPD (2020)

*Appendix 5: Population of New York City by race*




Source: Own depiction based on U.S. Census Bureau (2011)

*Appendix 6: Breusch and Pagan Lagrangian multiplier test for random effects*

```
Complaints[Race,t] = Xb + u[Race] + e[Race,t]

Estimated results:
                 |       Var     sd = sqrt(Var)
        ---------+-----------------------------
        Complai~s |   1262.284       35.52864
                e |   83.57429       9.141898
                u |   1873.437       43.28321

Test:   Var(u) = 0
                          chibar2(01) =  4885.96
                       Prob > chibar2 =   0.0000
```

Source: Stata results table

*Appendix 7: Test for skewness and kurtosis (normality)*

Skewness and kurtosis tests for normality

| Variable | Obs | Pr(skewness) | Pr(kurtosis) | Joint test Adj chi2(2) | Joint test Prob>chi2 |
|---|---|---|---|---|---|
| Complaints | **240** | **0.0000** | **0.3046** | **34.04** | **0.0000** |

Skewness and kurtosis tests for normality

| Variable | Obs | Pr(skewness) | Pr(kurtosis) | Joint test Adj chi2(2) | Joint test Prob>chi2 |
|---|---|---|---|---|---|
| Allegations | **240** | **0.0000** | **0.3345** | **33.48** | **0.0000** |

Source: Stata results table

*Appendix 8: Fisher-type unit-root test for complaints*

```
Fisher-type unit-root test for Complaints
Based on augmented Dickey-Fuller tests
-------------------------------------------
Ho: All panels contain unit roots            Number of panels  =      5
Ha: At least one panel is stationary         Number of periods =     48

AR parameter: Panel-specific                 Asymptotics: T -> Infinity
Panel means:  Included
Time trend:   Not included
Drift term:   Not included                   ADF regressions: 3 lags
-----------------------------------------------------------------------------
                                  Statistic      p-value
-----------------------------------------------------------------------------
 Inverse chi-squared(10)     P     26.4345       0.0032
 Inverse normal              Z     -2.7911       0.0026
 Inverse logit t(29)         L*    -3.0452       0.0025
 Modified inv. chi-squared Pm      3.6749       0.0001
-----------------------------------------------------------------------------
 P statistic requires number of panels to be finite.
 Other statistics are suitable for finite or infinite number of panels.
```

Source: Stata results table

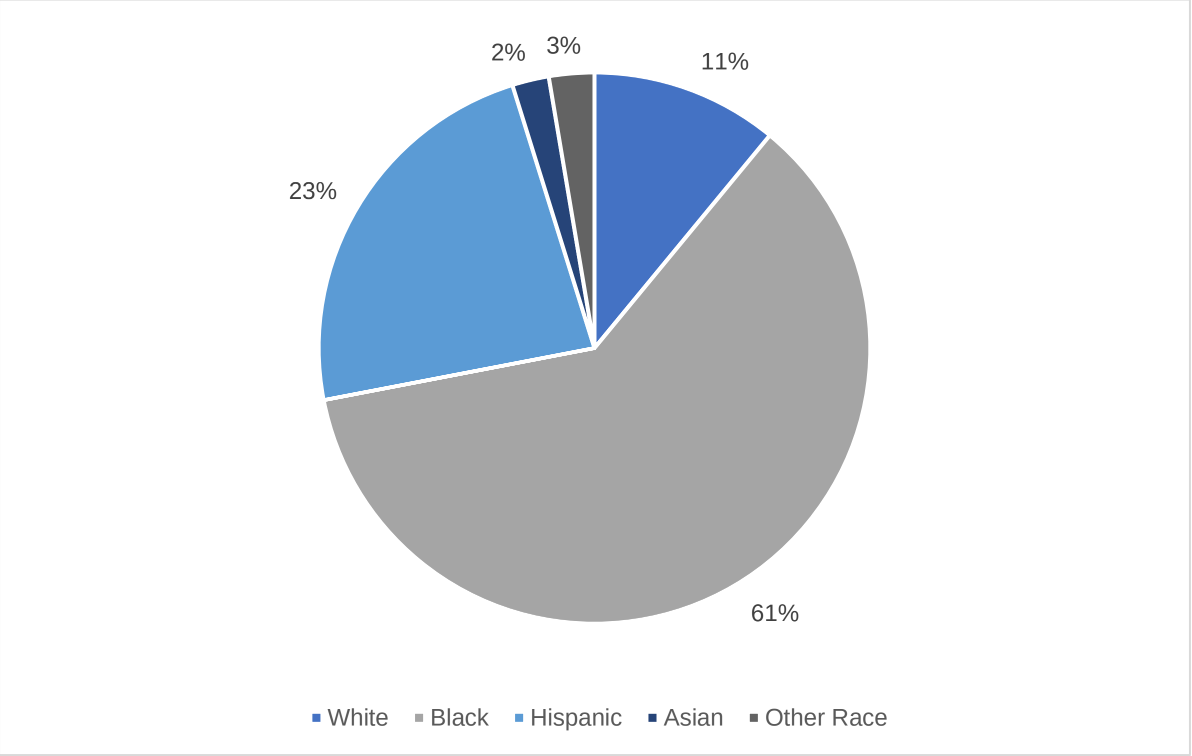
*Appendix 9: Fisher-type unit-root test for allegations*

```
Fisher-type unit-root test for Allegations
Based on augmented Dickey-Fuller tests
--------------------------------------------------------------------
Ho: All panels contain unit roots           Number of panels  =     5
Ha: At least one panel is stationary        Number of periods =    48

AR parameter: Panel-specific                Asymptotics: T -> Infinity
Panel means:  Included
Time trend:   Not included
Drift term:   Not included                  ADF regressions: 3 lags
------------------------------------------------------------------------------
                                  Statistic      p-value
------------------------------------------------------------------------------
 Inverse chi-squared(10)   P        27.7396       0.0020
 Inverse normal            Z        -2.6661       0.0038
 Inverse logit t(29)       L*       -2.9754       0.0029
 Modified inv. chi-squared Pm        3.9667       0.0000
------------------------------------------------------------------------------
 P statistic requires number of panels to be finite.
 Other statistics are suitable for finite or infinite number of panels.
```

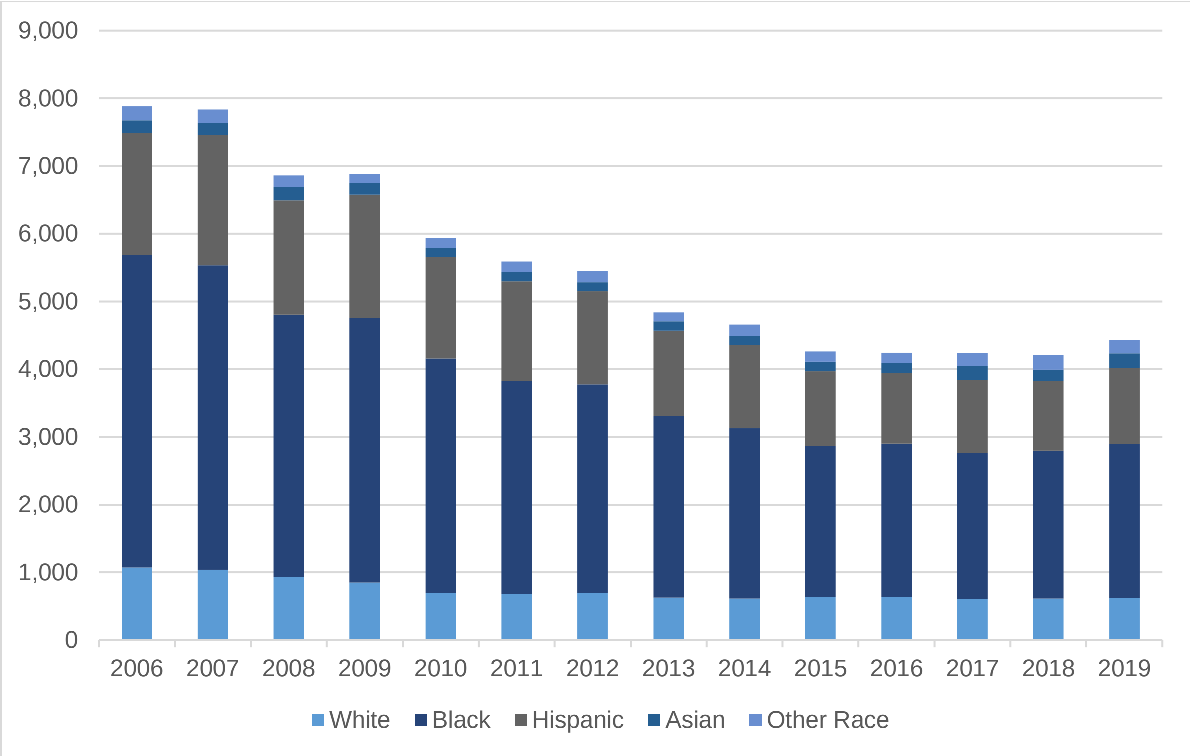
Source: Stata results table

*Appendix 10: Distribution of complainants by race*




Source: Own depiction based on ProPublica (2020)

*Appendix 11: Total complainants by race*




Source: Own depiction based on CCRB (2021b)