



Do body-worn cameras improve community perceptions of the police? Results from a controlled experimental evaluation

Anthony A. Braga¹ · John M. MacDonald¹ · Lisa M. Barao²

Accepted: 4 June 2021 / Published online: 22 November 2021

© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

Objectives Outfitting police officers with body-worn cameras (BWCs) has been suggested to improve police-community relations. This study evaluates whether the deployment of BWCs on NYPD officers impacted resident perceptions of the police.

Methods A cluster randomized controlled trial design was used to test the influence of BWCs on resident perceptions of the NYPD in treatment precincts relative to control precincts. Dual-frame randomly selected telephone surveys were used to collect pre-intervention and post-intervention resident perception data.

Results We find no statistically significant differences between BWC treatment and control precincts in general perceptions of the NYPD or the average assessment of police officer behavior among those who have had recent encounters with the NYPD.

Conclusion Strong community support for BWC adoption and citizen expectations for videos of controversial policing events suggests the continued use of this technology. However, BWCs should be implemented with other evidence-based programs to enhance police-community relations.

Keywords Body-worn camera · Randomized experiment · Police-community relations · NYPD · Telephone survey · Citizen perceptions of police

✉ Anthony A. Braga
abraga@upenn.edu

¹ Department of Criminology, University of Pennsylvania, 558 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104-6286, USA

² Westfield State University, Westfield, MA 01086, USA

Introduction

Body-worn cameras (BWCs) have been suggested as a technological intervention that could address police-community relation problems by enhancing police legitimacy through improved transparency, accountability, civility, and lawfulness of police-citizen encounters (Stanley, 2015; President's Task Force on 21st Century Policing, 2015). Surveys generally document that citizens support the adoption of BWCs by police departments (see, e.g., Culhane et al., 2016; Crow et al., 2017; White et al., 2018a, b). However, studies suggest that BWCs have not produced consistent impacts on how citizens view the police (Lum et al., 2019). There is some suggestive evidence that the presence of BWCs may improve perceptions of procedural justice and legitimacy among citizens who experience a police encounter (McCluskey et al., 2019; Demir et al., 2020). However, it remains unclear whether BWC adoption produces meaningful changes in broader community perceptions of the police.

The New York City Police Department (NYPD) adopted BWCs as part of a series of reforms emanating from litigation over their stop, question, and frisk practices. During the late 2000s and early 2010s, the City of New York vigorously defended the use of stop, question, and frisk practices as an effective crime prevention strategy (Bloomberg, 2013; Kelly, 2015). While there is some evidence that increased stops produced modest crime control impacts (MacDonald et al., 2016; Weisburd et al., 2016), NYPD stop, question, and frisk efforts were criticized for producing unconstitutional racial disparities in stops, frisks, and searches of citizens (Fagan & Davies, 2000; Gelman et al., 2007), and federal class action lawsuits were filed against the City of New York. In 2013, the City of New York reached a court-approved settlement with the plaintiffs after a federal judge ruled that the NYPD was liable for a pattern and practice of unconstitutional stops of citizens that violated the Fourth and Fourteenth Amendments. The litigation settlement mandated a series of reforms including changes to NYPD stop policies, training, and auditing; modifications to the processing of civilian complaints and officer discipline procedures; the adoption of new measures to evaluate organizational performance; and the establishment and evaluation of a pilot BWC program. The settlement also required the appointment of an independent monitor to evaluate NYPD's progress in making the agreed-upon reforms.

In the remedial order, then US District Judge Shira Scheindlin ordered the pilot BWC program to determine whether the technology created objective records of police stops, encouraged lawful and respectful police-citizen interactions, offered a way to help determine the validity of accusations of police misconduct, and improved distrust between the NYPD and the public.¹ While the NYPD and the court-appointed monitor team was planning the BWC pilot program, the New York University (NYU) School of Law Policing Project (2016) surveyed New York City residents and found most respondents were overwhelmingly in favor of BWCs, and 82% expressed the hope that the technology would improve fractured police-community relations. At the April 2017 press conference announcing the BWC program launch, New York City government leaders suggested the placement of BWCs on NYPD officers would increase police accountability and transparency with Mayor Bill DeBlasio further noting that

¹ See page 27 of the remedial order; <http://nypdmonitor.org/wp-content/uploads/2015/09/Floyd-Remedy-Opinion-8-12-13.pdf> (accessed January 31, 2021).

“body worn cameras will bring police and communities closer together” and then New York State Senator Marisol Alcantara observing “body worn cameras will be a fabulous tool to increase trust between the community and the police.”²

This article presents the results of the court-ordered evaluation of the impact of BWCs on community perceptions of the NYPD based on telephone surveys of representative samples of residents in 20 treatment and 20 control precincts. The article begins by briefly reviewing the existing literature on public perceptions of the police and studies evaluating BWC impacts on community perceptions of the police and procedural justice during police-citizen encounters. We then summarize the NYPD BWC pilot program implementation and describe the cluster randomized controlled trial design, survey data collection methods, and the empirical models that estimate the effect of the BWC program on perceptions of the police. We find that the deployment of BWCs on officers did not produce any meaningful changes in overall community perceptions of the NYPD or officer behavior among those who reported recent encounters with the NYPD. The policy implications are considered in the concluding section.

Literature review

Public perceptions of the police

Most Americans have positive views of their local police departments. Gallup poll data has found that 86% of whites, 80% of Hispanics, and 73% of blacks were either “very satisfied” or “somewhat satisfied” with the police department in the respondents’ city (Weitzer & Tuch, 2006: 41). At the same time, there exists a racial hierarchy in the degree of satisfaction with the police, with nearly half of all whites reporting they were “very satisfied” as compared to 36% of Hispanics and only 22% of blacks responding to the survey. Differences between blacks and whites were shaped by varying perceptions of their treatment during prior encounters with the police, feelings of safety in their own neighborhood, crime control efficacy and police use of community policing strategies, and exposure to police misconduct. Unfortunately, when compared to white subjects, black respondents were more inclined to believe that the police abuse citizens, treat minorities more harshly relative to whites, and were not held accountable for misconduct (Weitzer & Tuch, 2006). Numerous studies have consistently found that blacks express lower levels of trust in and satisfaction with the police (e.g., Taylor et al., 2001; Hurst & Frank, 2000; Leiber et al., 1998; MacDonald & Stokes, 2006; MacDonald et al., 2007).

The May 25, 2020, death of George Floyd during an arrest by Minneapolis Police Department officers sparked nationwide protests and renewed a national conversation on race and policing, stemming from persistent concerns over the deaths of unarmed black men during encounters with the police and racial disparities generated by police enforcement efforts. These events have had negative impacts on public perceptions of the police. Gallup public opinion poll data shows that confidence in the police has fallen in recent years, and for the first time in three decades, less than 50% of Americans expressed “a great deal” or “quite

² Statements made at April 27, 2017, press conference; see <https://www.youtube.com/watch?v=tsAJkDNy1-4> (Accessed January 31, 2021).

a lot” of confidence in the police in 2020.³ The decline in confidence follows the same hierarchy, with the black-white gap in confidence in the police being the largest recorded since 1993.⁴ The City of New York does not conduct ongoing public opinion polls of the NYPD. However, the ad hoc studies that do exist suggest that community residents in general hold favorable opinions of the NYPD, non-white residents have less positive perceptions of the NYPD when compared to white residents, and young non-white males that experience high levels of police contacts have very negative perceptions of the NYPD.

The Vera Institute of Justice surveyed nearly 2,000 residents of five NYPD precincts on their opinions of and experiences with the NYPD (Miller et al., 2004). From the survey data, the Vera researchers used factor analyses to develop underlying latent variables measuring respondent perceptions of police effectiveness and misconduct. Overall, respondents ranked the NYPD high on their perceptions of effectiveness and low on their perceptions of misconduct (Miller et al., 2004). However, these assessments showed a racial hierarchy with white residents (.77 effectiveness, .31 misconduct) reporting more positive assessments than those of Hispanic (.70 effectiveness, .40 misconduct) and black (.65 effectiveness, .56 misconduct) residents. The Vera study noted that negative prior personal and vicarious encounters (i.e., the experiences of a friend or family member) with NYPD were highly predictive of lower effectiveness and higher misconduct ratings (Miller et al., 2004). Respondents with negative direct and vicarious encounters with the NYPD tended to be minority residents living in economically disadvantaged police precincts. Consistent with other studies noting the asymmetric impact of negative encounters with the police on public perceptions (e.g., Skogan, 2006), positive and neutral interactions with officers did not enhance respondents’ assessments of the NYPD.

During the late 1990s and continuing through the 2000s, the NYPD engaged in crime control strategies that dramatically increased the number of misdemeanor arrests and stop-question-and-frisks of citizens (Kelling & Coles, 1996; Fagan et al., 1998). Studies generally found that these strategies generated concern regarding racial disparities in who was being stopped, frisked, and arrested by the NYPD (Fagan & Davies, 2000; Gelman et al., 2007; Harcourt & Ludwig, 2006; but see Ridgeway, 2007). Fagan et al. (2010) estimated that 80% of black residents between the ages of 16 and 17 had been stopped one or more times by the NYPD in 2006. This intensive law enforcement scrutiny of minority residents in NYC appears to have increased negative perceptions of the NYPD. A recent Vera Institute of Justice survey of 500 NYC residents between the ages of 18 and 25 who had been stopped at least once in high-crime neighborhoods found that 88% of these individuals believed that neighborhood residents did not trust the police (Fratello et al., 2013). Additionally, a representative sample of interviews with 1,261 NYC residents between 18 and 26 found that nearly half had been stopped in a car or on the street and, the more experiences respondents had with the police, the less likely they were to indicate that they were treated fairly by the police and that the encounters were lawful (Tyler et al., 2014).

BWCs and citizen perceptions of the police

BWCs have been rapidly adopted by US police departments, in part, as a response to ongoing police-community relation problems and persistent concerns over police

³ <https://news.gallup.com/poll/1597/confidence-institutions.aspx> (accessed September 18, 2020).

⁴ <https://news.gallup.com/poll/317114/black-white-adults-confidence-diverges-police.aspx> (accessed September 18, 2020).

shootings of unarmed black citizens (Lum et al., 2015; White, 2014). The President's Task Force on 21st Century Policing (2015) suggested BWCs might be an important technology to enhance police legitimacy and community trust in the police. Under the Obama Administration, the US Department of Justice (2015) launched a \$20 million funding initiative to support local police departments in outfitting their officers with BWCs. The BWC grant program continued under the Trump Administration. In 2016, the US Bureau of Justice Statistics reported that about 80% of large police departments and sheriff's offices with 500 or more full-time sworn officers had acquired the BWC technology and 75% of these large law enforcement agencies reported improved community perceptions as a key reason for adopting BWCs (Hyland, 2018). Other reasons for adopting BWC technology included enhanced officer safety, increased evidence quality, reduced civilian complaints, and decreased agency liability.

The available research suggests that citizens support the adoption of BWC by police departments and hold high expectations for the technology in improving accountability and enhancing citizen confidence in the police (for a summary, see Lum et al., 2019). Using data collected from interviews and focus groups in two cities, Todak et al. (2018) found that judges, prosecutors, mental health workers, city leaders, civilian oversight members, victim advocates, and other key external stakeholders were highly supportive of BWC implementation by their police departments. Detained suspects of crime also favor the deployment of BWCs on police officers (Taylor et al., 2017). However, consistent with the broader literature on public perceptions of the police, minority citizens generally view the potential benefits of BWCs with less enthusiasm and are more skeptical than whites of its efficacy in holding officers accountable for misconduct (see, e.g., Crow et al., 2017; Sousa et al., 2018; Kerrison et al., 2018).

It is worth noting here that BWC advocates do not articulate a well-developed "theory of change" that identifies the specific mechanisms through which outfitting officers with the technology will improve community perceptions of the police. Governmental officials, civil liberty organizations, and community activists tend to describe more proximate outcomes of BWC adoptions, such as increased transparency, enhanced accountability, and more civil police-citizen interactions, that will influence public perceptions and, in turn, result in improved police-community relations. Very few studies have attempted to measure the impacts of BWCs on community-wide citizen perceptions of the police before and after the technology was implemented. And only a handful of evaluations have examined the effects of the technology on the perceptions of citizens who have had interactions with BWC officers.

Ellis et al. (2015) conducted a mail-in survey of Isle of Wight (UK) residents before and after BWCs were placed on police officers. Public approval of the police was very high before and after BWC adoption, and public confidence in the police changed very little after the technology was implemented. Ariel (2016a) conducted a quasi-experimental evaluation of the influence of BWCs on citizen crime reporting behaviors in Denver (CO). Based on an analysis of calls for service data in treatment and control areas, Ariel (2016a) found that BWCs increased the number of calls to the police in low-crime residential street segments but did not influence the number of calls in high-crime street segments. He concluded that the greater willingness of citizens from treatment low-crime residential places to report crimes to the police was a likely result of improved police-community relations stimulated by the placement of BWCs on officers.

Other research has examined the influence of BWCs on citizen perceptions after specific encounters with the police, tending to focus on police legitimacy and procedural justice issues (Lum et al., 2019). Deterrence theory suggests that the presence of BWCs during police-citizen encounters increases the risk that inappropriate behavior will be detected; public self-awareness theories suggest BWCs will cause police officers and citizens alike to be more aware of their behaviors during interactions and stimulate more respectful and courteous behavior (Ariel et al., 2015; but, more generally, see Duval & Wicklund, 1972; Zimring & Hawkins, 1973). Ariel (2016b) suggests that the purported behavioral changes associated with the deployment of BWCs on police officers are consistent with Tyler's (2003) process-based model of police legitimacy: allowing voice, making neutral decisions, showing trustworthiness, and treating citizens with respect and dignity.⁵ Studies provide some indirect evidence of behavioral changes that supports the link between BWC presence and enhanced procedural justice. Randomized experiments generally find that BWCs improve the civility of police-citizen encounters as evidenced by reduced citizen complaints against treatment officers relative to control officers (e.g., Ariel et al., 2015; Braga et al., 2018; Jennings et al., 2015). The impact of BWC on police use of force remains unclear (Lum et al., 2019; White et al., 2018a, b).

The available evidence on the impacts of BWCs on citizen perceptions of the police following encounters with officers wearing the cameras is mixed. In Spokane (WA), telephone interviews with 249 citizens who recently had an encounter with the police found that their perceptions of procedural justice during the encounter improved when they were aware that the officer was wearing a BWC (only 28% of the subjects reported being aware of the BWC; see White et al., 2017). However, a randomized controlled trial of BWCs in Arlington (TX) found no differences in perceptions of legitimacy, satisfaction, and police professionalism by citizens who recently had an encounter with officers who were equipped with cameras compared to control officers who did not have cameras (Goodison & Wilson, 2017). In Anaheim (CA), a randomized controlled trial that surveyed respondents after encounters with the police reported that the presence of a BWC combined with the use of a procedural justice script to guide officer behaviors during encounters generated larger impacts on citizen satisfaction relative to the presence of the BWC alone (McClure et al., 2017).

Two studies suggest that outfitting officers with BWCs may stimulate procedurally just behaviors during encounters with citizens. In Los Angeles (CA), McCluskey et al. (2019) used systematic social observations of police-citizen encounters to conduct a pre-post analysis of the effects of BWCs on officer behaviors. Officers equipped with BWCs were observed engaging in significantly more displays of procedurally just behaviors during their interactions with citizens compared to officers without BWCs. In the Eskisehir province of Turkey, a quasi-experimental evaluation concluded that drivers stopped by police wearing BWCs reported improved perceptions of procedural justice during the stop and enhanced perceptions of the legitimacy of the traffic officers and the police in general relative to drivers stopped by non-BWC comparison officers (Demir et al., 2020).

⁵ A number of nonexperimental studies have supported the main elements of the process-based model of police legitimacy (see, e.g., Mastrofski et al., 1996; Paternoster et al., 1997; Sunshine & Tyler, 2003; Tyler & Fagan, 2008; Tyler & Wakslak, 2004), and certain field experiments suggest citizen perceptions are positively impacted by officers who engage procedurally just behaviors (Mazerolle et al., 2013; Sahin et al., 2017). However, it is important to note that the existing evidence base is not strong enough to establish causation in terms of the influence of procedural justice on citizen compliance with the law (Nagin & Telep, 2017; Worden & McLean, 2017; National Academies 2018).

This study intends to advance our knowledge on an important gap in the existing research literature – whether the deployment of BWCs improves broader community perceptions of the police. New York City is an important research site to test the relationship between BWCs and police-community relations given the well-known concerns expressed by minority community members over the NYPD’s excessive use of stop, question, and frisk encounters, the federal court ordered pilot testing of BWCs to improve the constitutionality of police-citizen interactions, and the hope of politicians that BWCs would help “bring communities and police together.”

Methods

A cluster randomized controlled trial (RCT) was used to evaluate the effects of BWCs on a series of police work activity, civility, lawfulness, and community perception outcomes (Braga et al., 2021). Forty precincts with the highest numbers of Citizen Complaint Review Board (CCRB) complaints against NYPD officers were identified and then matched into 20 pairs based on demographics, socioeconomic characteristics, crime, and police activity (Appendix 1, Table 5). Within each pair, a computer algorithm was used to randomly select one precinct to have BWCs deployed on its officers (the treatment precinct) and the other to be without BWCs (the control precinct). Over the course of the 1-year intervention period, 3,889 NYPD uniformed officers working the third platoon (3 PM to midnight shift) and plainclothes officers working anti-crime unit assignments were equipped with BWCs. Subsequent analyses confirmed that officers in the treatment and control groups were on average similar in terms of demographics, length of service, rank, work activities, and number of citizen complaints (Braga et al., 2021; see Appendix 2, Table 6).

All five boroughs had at least one precinct included in the treatment group.⁶ The resident populations of precincts are akin to small to midsize US cities with the smallest treatment precinct serving some 47,418 residents (25th Precinct in Manhattan, mostly located in East Harlem) and the largest treatment precinct serving some 188,666 residents (105th Precinct in Queens in the easternmost part of the borough). Within specifically matched pairs, there were treatment and control precincts that both had larger ambient populations of work commuters and tourists generated by business districts and tourist attractions. While the BWC intervention could plausibly benefit residential and ambient populations alike, the key community perception data collection effort was limited to residents of the treatment and control precincts included in the experiment.

The complexities of adopting BWCs in a very large police department that provides services to citizens in a diverse set of communities across a sprawling metropolitan area required the NYPD to develop and execute a careful deployment plan. Core implementation activities included ensuring the BWC treatment precincts had the appropriate information technology resources to handle uploading videos to cloud data storage as

⁶ Brooklyn had 6 treatment precincts, Bronx had 5 treatment precincts, Manhattan had 5 treatment precincts, Queens had 3 treatment precincts, and Staten Island had 1 treatment precinct. In general, the most appropriate precinct matches were found within boroughs. There were two exceptions. One Bronx precinct was matched to a precinct in Brooklyn, and one precinct in Queens was matched to a precinct in Staten Island.

well as training participating line-level officers and supervisors in BWC policies and technological operations.⁷ To manage these requirements appropriately, BWC implementation in the 20 treatment precincts was staggered over 7 months beginning in April 2017 and ending in November 2017 (see Fig. 1). Despite varying start dates, all treatment precincts used the BWCs for one full year with the experimental intervention period ending in November 2018.

It is important to note that this experiment was executed during a department-wide effort to outfit all NYPD uniformed patrol officers and selected specialized unit officers with BWCs between December 2017 and August 2019.⁸ To avoid threats to the integrity of the experimental design, the NYPD first deployed BWCs on uniformed officers in the 37 precincts that were not included in the evaluation. As each pair of the 40 experimental precincts with the most CCRB complaints completed the 1-year intervention period, control precinct officers then became eligible to wear BWCs as part of the citywide implementation after the experiment ended. As a result, this RCT tests the effects of BWC deployment on citizen perceptions in precincts with the most contentious relationships with the NYPD. BWC deployment on specialized unit officers started in March 2019, well after the last matched pair of precincts completed the 1-year treatment period (November 2018).

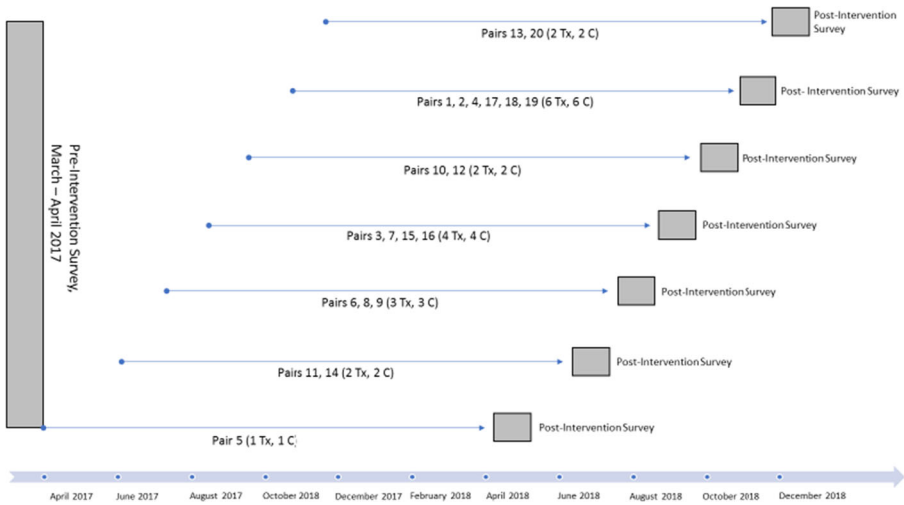
The cluster RCT found that the placement of BWCs on NYPD officers generated two statistically significant impacts on the included outcome measures (Braga et al., 2021). CCRB citizen complaints against BWC treatment officers were reduced by 21% when compared to non-BWC control officers. However, stop reports completed by BWC treatment officers increased by almost 39% relative to the number of stop reports completed by non-BWC control officers. Supplemental analyses suggested that the presence of BWCs during stop encounters made NYPD officers more likely to comply with departmental policies mandating the completion of stop reports relative to control officers who did not have their stop encounters recorded on video.

Telephone survey data collection methods

The evaluation of the impact of the body-worn camera program on NYC resident perceptions of the NYPD primarily involved a telephone survey. Hart Research Associates, an independent professional survey research firm, was selected by the NYPD Monitor to administer telephone surveys prior to the introduction of the BWCs (pre-intervention) and after cameras had been in use for a 1-year intervention period (post-intervention). The survey questionnaire measured the demographics of respondents and a series of 26 close-ended items, including questions on specific experiences of being stopped by police or other interactions with police occurring in the prior 12

⁷ NYPD BWC policy specifies a variety of situations where the BWC must be activated including crime-in-progress assignments, interior patrols of NYCHA buildings, pedestrian stops, vehicle stops, personal interactions that escalate, interactions with emotionally disturbed persons, arrests, and other law enforcement duties. The NYPD policy instructs officers to notify members of the public that an interaction is being recorded as soon as it is “reasonably practical.” See https://www1.nyc.gov/assets/nypd/downloads/pdf/public_information/body-worn-cameras-patrol-guide.pdf (accessed July 19, 2020).

⁸ <https://www1.nyc.gov/site/nypd/about/about-nypd/equipment-tech/body-worn-cameras.page> (accessed July 12, 2020).



Note: Pre-intervention surveys of precinct residents in Pair 5 was completed before the treatment officers were equipped with BWCs.

Fig. 1 Implementation of BWC treatment in 20 matched pairs included in cluster randomized controlled trial

months. The survey included outcome questions borrowed from prior surveys on individual experiences during police encounters and attitudes towards the police more generally (e.g., see Tyler & Huo, 2002; Reisig et al., 2007; Braga et al., 2014; Sahin et al., 2017). In addition to using pre-validated questions, the survey was piloted to ensure that the questions adequately captured the information needed.⁹

The pre-intervention and post-intervention samples were designed to be representative of the adult populations in the control and treatment precincts, including young minority males who are harder to reach in telephone surveys. A dual-frame random selection approach, involving both landlines and cellphones, was used (see American Association for Public Opinion Research [AAPOR], 2010). TargetSmart, a political data surveying firm with 6.2 million subjects in its New York City database, was provided shapefiles of the 40 experimental precincts to develop a sampling frame of residents and their phone numbers. The US Census Bureau's 2010 American Community Survey (ACS) was used in conjunction with the shapefile data to set allocations by age, gender, and race for the control and treatment samples. Additional telephone numbers, including contract and prepaid mobile phone numbers, were supplied by a second vendor, Link2Tek. Prospective interview subjects were randomly selected from the lists of phone numbers of residents in the 40 experimental precincts provided by

⁹ In March 2017, 190 preliminary telephone interviews were completed, and these responses were analyzed before proceeding with the full survey. One question was adjusted to ensure respondents were reliably reporting whether they had a close friend or family member who was an NYPD officer. The average length of the pretest survey completion time was just under 10 minutes which allowed two additional outcome questions to be added to the final instrument. Hart Research Associates estimated that the final survey would not take longer than 12 minutes to complete but did not include a time length measure in the final database provided to the NYPD monitor team.

TargetSmart and Link2Tek.¹⁰ No incentives were offered to prospective interview subjects.

The interviewing firm American Directions fielded the pre-intervention survey in English and Spanish (by respondent choice) during March 2017 and April 2017 (Fig. 1). Pre-intervention surveying was completed in matched pair 5 before the treatment precinct officers were outfitted with BWCs in April 2017. The post-intervention surveys were conducted as the 1-year treatment period expired in each of the matched pairs of treatment and control precincts, beginning in May 2018 and continuing through December 2018. As such, the time between pre-intervention and post-intervention surveys varied across pairs (ranging from 12 to 19 months) but did not vary for the treatment and control precincts included in each matched pair. Since the time between surveys was the same in each matched pair, our experimental analyses were not biased by the varying time periods between pre-intervention and post-intervention surveys across the matched pairs. During pre-intervention and post-intervention survey data collection periods, eligible numbers were repeatedly called at varying times of the day until prospective subjects answered the phone and either agreed or declined to participate; this continued until survey sample quotas were reached.

Eighteen to 34-year-old men in both the treatment and control precincts were oversampled to obtain a sufficient number of respondents who had interactions with the NYPD. However, the treatment and control samples were weighted separately by age and gender to reflect the population level demographics of these precincts¹¹ and minimize any nonresponse bias that may have been present in the survey data collection. The demographic distributions in the weighted survey samples closely matched those reported for adults in the 40 precincts by the 2010 ACS (see Appendix 3, Table 7 for pre-intervention comparisons).

The pre-intervention survey involved live telephone interviews with a total of 3,000 residents in the 20 treatment precincts (25.8% response rate¹²) and 3,000 residents in the 20 control precincts (26.6% response rate). The post-intervention survey involved live telephone interviews with 3,037 residents in the 20 treatment precincts (28.9% response rate) and 3,020 residents in the 20 control precincts (29.3% response rate). The response rates did not significantly differ between the treatment and control groups during the pre-intervention (two-sample difference of proportions $|z| = 0.7046$, $p =$

¹⁰ TargetSmart develops its subject list and their addresses and phone numbers from a proprietary set of data sources including voter registration data, telephone company databases (such as Telco Repository, which is the “Big Five” telecommunications companies’ near real-time database of every phone number in the United States), and other sources. TargetSmart reported that 74% of the 6.2 million listed residents were associated with telephone numbers: 19% cellphone only, 22% cellphone and landline, and 33% landline only. Following AAPOR’s (2010) recommendations, a screened approach was used when considering eligible residents in the 40 experimental precincts to remove the overlap of the dual frame.

¹¹ For instance, in the pre-intervention survey, men ages 18–24 were weighted according to their 2010 ACS proportional representation: 18% of control precinct residents and 17% of treatment precinct residents. Weight calculations for the pre-intervention and post-intervention landline samples and cell phone samples in the treatment and control precincts were estimated separately and then combined via stratified sampling techniques. The weighting used for landline samples and cellphone samples accounted for differential nonresponse by telephone usage group.

¹² We used AAPOR Response Rate 5 (RR5) to calculate our survey response rates: this is the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-off plus noncontacts plus others). https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/Standard_Definitions_07_08_Final.pdf (accessed February 17, 2021).

0.4810) and post-intervention periods (two-sample difference of proportions $|z| = 0.344$, $p = 0.733$), suggesting that any existing nonresponse bias would be similar in the treatment and control groups and not impact the ability to detect differences in resident perceptions over time. These response rates are also consistent with other recent telephone surveys of citizens who have interacted with police (34.4% in Rosenbaum et al., 2015; 25.0% in Malm et al., 2016; 27.8% in White et al., 2017). The proportion of respondents who were surveyed via cellphones relative to landlines also did not differ significantly across experimental groups in the pre-intervention (treatment = 71% cellphone, 29% landline; control = 72% cellphone, 28% landline; two-sample difference of proportions $|z| = 0.858$, $p = 0.391$) and post-intervention periods (treatment = 85% cellphone, 15% landline; control = 84% cellphone, 14% landline; two-sample difference of proportions $|z| = 1.070$, $p = 0.284$).

Limiting treatment contamination of control group respondents

Research suggests that officers with BWCs influence the behavior of officers without cameras if they work simultaneously in the same area and interact with the same people (Ariel et al., 2019; Braga et al., 2020). Similarly, BWC exposure through a subset of officers in an area could influence how civilians in that area interact with the police more broadly. Such contamination undermines the ability to detect intervention effects because both treatment and control officers (and civilians) could be modifying their behaviors due to the presence of BWCs. This is known more formally as a violation of the stable unit treatment value assumption (SUTVA) that assumes that the effect of a treatment on a given observational unit is not related to the intervention assignments of other observational units (Rubin, 1980). SUTVA violations undermine the internal validity of randomized experimental designs (Shadish et al., 2002). In a recent systematic review of BWC evaluations, Lum et al. (2020) noted that many BWC studies had some form of contamination between treatment and control groups and this phenomenon was particularly difficult to avoid and manage in BWC studies.

In this study, we attempted to address the potential contamination of the control group by design. Cluster randomization is often used to prevent contamination between treatment and control groups in public health and medical trials (see, e.g., Torgerson, 2001; Magill et al., 2019). The random allocation of clusters of officers, rather than randomizing individual officers, who work in distinct precincts to the BWC treatment condition or not limits the extensiveness of the contamination problem. There was very little contamination between treatment and control officer assignments over the course of the intervention period.¹³ Nevertheless, it remained possible that BWC officers working in treatment precincts may have responded to relatively rare emergency dispatches involving officer safety concerns that mandated additional resources in adjacent control precincts. Further, it is also possible that residents of control precincts may have worked or visited treatment precincts and encountered BWC officers in public spaces. We limited all telephone survey outcome questions to respondent

¹³ As noted in Braga, MacDonald, and McCabe (2021), only 18 officers (0.5% of 3,889) appeared in both the treatment group and control group during the intervention period. The movement observed included 12 officers moving from control to treatment conditions and 6 officers moving from treatment to control conditions.

perceptions of and experiences with NYPD officers in their immediate neighborhood to diminish the risk of contamination of control respondents due to BWC exposure in treatment precincts they may have frequented.

Measurement outcomes and sample characteristics

Survey questions were intended to capture latent measures (Long, 1983) of perceptions of the NYPD in general, distrust of neighborhood police officers, procedural justice, and negative assessments of interactions with the police that took place in the past 12 months as a result of a car, pedestrian stop, or request for assistance by the NYPD. Respondents were asked to indicate on Likert scale survey items on four- and five-point scales their general feelings or level of agreement with questions about the NYPD.

Table 1 shows the covariance of the responses collected from these sets of questions that were analyzed to develop seven outcome measures that capture perceptions of the NYPD. Confirmatory factor analysis was used to test whether the telephone survey data collected for this study fit the hypothesized measurement models applied in previous research (Tyler & Huo, 2002; Reisig et al., 2007; Braga et al., 2014; Sahin et al., 2017). All outcome measures had Cronbach's alphas (Cronbach, 1951) that exceeded .70, suggesting overall good internal consistency. Confirmatory factor analysis showed that survey items representing outcome measures had strong intra-item correlations (Kim & Mueller, 1978).¹⁴ The confirmatory factor analyses of the unweighted data further suggested a very good fit between the hypothesized models and the true covariance matrices underlying the data.¹⁵ All comparative fit index (CFI) values were larger than .90, and all standardized root mean residual (SRMR) values below were below .05 (Brown, 2015). The lowest goodness-of-fit statistics were produced by the unweighted perceptions of neighborhood officers latent variable with CFI = 0.956 and SRMR = 0.035.

Table 2 shows the similarity between respondents in the BWC treatment and control group precincts on key demographic variables and pre-intervention survey items. Standardized mean differences $|d|$ between groups on measures did not generally exceed .20 (a small effect size) (Cohen, 1988).¹⁶ Treatment and control

¹⁴ The factor loading variances of all outcome variables in Table 2 were constrained to equal 1 and can be interpreted as correlations that range from 0 to 1 (see Bollen, 1989; Bollen & Long, 1993).

¹⁵ Stata 15.1 was used to execute the analyses in this study. Since the survey data were weighted to more accurately reflect precinct populations in our analyses, the "svy" series of commands were used. Unfortunately, the CFI and SRMR goodness-of-fit statistics cannot be estimated with the weighted data. In models run with the svy prefix, standard errors are reported as "linearized standard errors." These are synonymous with robust standard errors when not using survey data. As summarized by StataCorp (2019: 6), "By default, svy computes standard errors by using the linearized variance estimator— so called because it is based on a first-order Taylor series linear approximation... In the nonsurvey context, we refer to this variance estimator as the robust variance estimator, otherwise known in Stata as the Huber/White/sandwich estimator; see [P] robust." As such, we report these commonly used goodness-of-fit statistics based on confirmatory factor analyses of the unweighted data. There were no substantive differences between the unweighted and weighted data in the factor loadings when estimating the latent variables included in this study. Comparative tables are available upon request from the authors.

¹⁶ We follow convention in referring to small mean differences as those that are less than .20 standard deviations. While randomization by design should mean that treatment and control groups are on average comparable, there is always the chance that some differences will emerge. Randomization does not ensure perfect comparability in a single experiment. What matters is how different the groups are on pre-existing differences. As such, statisticians recommend using a comparison of average differences, like a standardized effect size, rather than a test statistic and p value (Imai et al., 2008).

respondents in the telephone survey had very similar perceptions of the NYPD and neighborhood safety during pre-intervention period. During the pre-intervention period, very few telephone survey respondents reported being subjected to a car stop while driving in their neighborhood (11.9% of treatment subjects, 12.9% of control subjects) or being subjected to a pedestrian stop while in their building or other public places in their neighborhood (5.9% of treatment subjects, 4.7% of control subjects) in the prior 12 months. Some 19.3% of treatment telephone survey subjects and 20.1% of control telephone survey subjects reported contacting the police for emergency assistance during the year preceding. Cohen's $|d|$ results also show that treatment and control telephone survey respondents reported the same perceptions of police behaviors and procedural justice during the pre-intervention period. The only exception ($|d| = .29$) was that a higher share of telephone survey respondents in the control precincts who were subjected to a pedestrian stop reported being "patted down on the outside of their clothing" (51%) relative to respondents in future BWC treatment precincts who were subjected to a pedestrian stop (33%).

The similarity between respondents in background and perceptions of the NYPD across treatment and control precincts suggests that the cluster randomization process was successful in creating comparable precincts prior to the BWC intervention. Survey data analyses were also weighted to ensure that the comparisons and inferences were generalizable to resident populations in the 40 precincts included in the cluster randomized controlled trial.¹⁷

Analytical approach

The impact of BWCs on resident perceptions in the treatment precincts relative to the control precincts was estimated through a difference-in-difference (DID) estimator (Card & Krueger, 1994). The DID estimator evaluated the difference in resident perceptions in the treatment precincts during the post-intervention period compared to during the pre-intervention period, relative to the same difference for residents in the control precincts.¹⁸ The general equation for our regression models was:

$$Y_{it} = \beta_0 + \beta_1 BWC_i + \beta_2 Period_t + \beta_3 (BWC_i \times Period_t) + u_i$$

In this model, Y_{it} represents our outcome measure for residents (i) during a specific observation period (t). The regressor BWC is a dummy variable identifying whether residents (i) were in the treatment precinct receiving body cameras (1) or not (0). The reference group comprises control residents in the experiment. The regressor $Period$ is

¹⁷ Survey respondents were slightly more likely to be female and, on average, in their mid to late 40s than the general residential population. More than half of respondents were high school graduates and many reported having earned a 4-year college or higher degree. Less than 20% of participants reported living in NYCHA housing, and fewer than 20% also said that they had a friend or family member currently in the NYPD.

¹⁸ It is worth noting here that the pre-intervention and post-intervention survey responses represent repeated cross-sectional data (newly sampled subjects interviewed each time over two data collection periods) rather than pure panel data (the same subjects interviewed each time over two data collection periods). In this study, the unit of experimentation is the precinct, so we consider subjects as exposed to the BWC intervention if they lived in the treatment precincts.

Table 1 Pre-Intervention Latent Variable Outcomes: Item Correlations, Cronbach's Alpha, and Factor Loadings

Latent Variables	N	Items Correlations				α	Factor Loading			
		A	B	C	D		E	F	Means	SE
Positive Perceptions of Neighborhood Officers										
A. Overall, how would you describe your feelings about the NYPD officers in your neighborhood	5083	1.000				0.856		3.74	0.796	0.025*
B. How would you rate the job that NYPD officers are doing in your neighborhood		0.646	1.000					3.74	0.762	0.024*
C. If I needed police assistance, I would feel comfortable seeking help		0.469	0.510	1.000				3.52	0.560	0.025*
D. I respect the police officers in my neighborhood		0.411	0.433	0.533	1.000			3.66	0.379	0.020*
E. Police officers in my neighborhood respect people's rights		0.509	0.548	0.568	0.492	1.000		3.19	0.681	0.020*
F. Police officers in my neighborhood treat people fairly, regardless of race or background		0.483	0.509	0.498	0.441	0.638	1.000	3.03	0.736	0.018*
Distrust of Neighborhood Officers										
A. Police officers in my neighborhood often abuse their authority	4693	A	B	C	D	0.798		2.06	0.866	0.020*
B. I sometimes feel nervous when I see police officers in my neighborhood approach me		1.000						1.92	0.622	0.020*
C. Police officers stop and frisk too many people in my neighborhood		0.489	1.000					2.04	0.734	0.025*
D. Police officers in my neighborhood use more force than necessary		0.557	0.443	1.000				2.01	0.811	0.024*
Procedural Justice During Car Stop During Last 12 Months										
A. How satisfied were you with the way officers handled the situation	931	A	B	C	D	0.822		2.66	0.910	0.054*
B. The police officers treated you with respect		1.000						3.10	0.796	0.036*
C. The police officers had a good reason for stopping you		0.647	1.000					2.35	0.802	0.066*
D. The amount of time you were stopped was reasonable		0.592	0.436	1.000				2.82	0.853	0.049*
Negative Assessment of Car Stop During Last 12 Months										
A. The police officers used language that you found threatening or abusive	909	A	B	C		0.703		1.69	0.787	0.072*
B. You felt the police officers stopped you because of your race/ethnicity		1.000						2.11	0.684	0.054*
C. The police officers used more force than necessary		0.494	1.000					1.57	0.745	0.060*
		0.579	0.457	1.000						

Table 1 (continued)

Latent Variables	N	Items Correlations				α	Factor Loading	
		A	B	C	D		Means	SE
<u>Procedural Justice in Pedestrian Stop During Last 12 Months</u>								
A. How satisfied were you with the way officers handled the situation	499	1.000				0.845	2.88	0.920 0.044*
B. The police officers treated you with respect		0.621	1.000				2.66	0.978 0.050*
C. The police officers had a good reason for stopping you		0.557	0.501	1.000			1.99	0.854 0.077*
D. The amount of time you were stopped was reasonable		0.532	0.546	0.516	1.000		2.53	0.858 0.066*
<u>Negative Assessment of Pedestrian Stop During Last 12 Months</u>								
A. The police officers used language that you found threatening or abusive	497	A	B	C		0.712	2.09	0.992 0.085*
B. You felt the police officers stopped you because of your race/ethnicity		1.000					2.64	0.581 0.093*
C. The police officers used more force than necessary		0.375	1.000				1.94	0.954 0.073*
<u>Procedural Justice During Contact for Assistance During Last 12 Months</u>								
A. How satisfied were you with the way officers handled the situation	1111	A	B	C	D	0.894	3.19	0.928 0.028*
B. The police officers treated you with respect		1.000					3.16	0.572 0.039*
C. The police officers clearly explained where you could get the help or information needed		0.656	0.591	1.000			3.30	0.872 0.040*
D. The police officers took your problems or questions seriously		0.726	0.643	0.690	1.000		3.36	0.900 0.025*

*p<.05

a dummy variable for whether the resident perception outcome was measured during the time (t) of the post-intervention (1) or during the pre-intervention (0) period. Our primary interest is in coefficient β_3 , which represents the DID estimate of the product of the BWC group and the post-intervention period. Standard errors were clustered by matched pairs to guard against unmeasured dependence within precincts biasing the estimates of BWC impact on public perceptions of the NYPD.¹⁹

Given the mild imbalances in treatment and control group subjects noted above, the DID models of the impact of BWCs on public perceptions of the NYPD were estimated with subject demographic covariates to adjust for these differences and improve precision in our estimates. Probit regression models were used to estimate the DID when outcome variables involved binary conditions (e.g., do you know any of the police officers that work in your neighborhood by name? 0 = No, 1 = Yes). Ordered probit regression models were estimated when outcomes variables involved Likert scales (e.g. capturing resident perceptions ranging from “very satisfied” to “very dissatisfied”). Resident perception outcome variables measured using Likert scales were reverse-coded to facilitate interpretation of the DID estimator.²⁰ Finally, DID estimates were based on structural equation models when outcomes involved latent variables.²¹ We excluded cases with missing values on outcome variables in each regression model.²²

¹⁹ This evaluation used cluster randomization, a variation of the classic experimental design in which clusters (groups) of subjects, rather than individual subjects, are randomly allocated to treatment and control conditions (Mosteller & Boruch, 2002). Following convention on a statistical analysis of cluster randomized controlled trials, we clustered standard errors on groups (matched pairs) using STATA statistical software (Rogers, 1993). This approach is advantageous because it allows the errors to vary differently between clusters, rather than assume they are fixed. An alternative approach would be to estimate with the model with a group-level random effect. This alternative approach, however, assumes that the clusters are random draws of the population of precincts in NYC, when in fact the study was setup to provide a balanced comparison of the impact of BWCs on outcomes in precincts with the highest levels of interactions between the NYPD and civilians (see Campbell et al., 2007).

²⁰ For instance, reverse coding involved switching the scale to run from “very satisfied” to “very dissatisfied” to run from “very dissatisfied” to “very satisfied.” As such, negative coefficients on the DID estimator would then be interpreted as treatment conditions that generated a negative effect on the selected outcome relative to control conditions over the pre-intervention and post-intervention periods.

²¹ In addition to the generalized structural equation models (GSEM), we also used factor score regressions with a Bartlett predictor and two-step GSEMs to calculate the DID estimators for the latent variable outcomes. While there are limitations to each approach, the findings did not differ substantively across the three modeling strategies. Some statisticians suggest that factor score regressions produce biased estimates (Hoshino & Bentler, 2013). Other analysts suggest that GSEMs diminish bias and better fit the data relative to factor score regressions (Devlieger et al., 2015). However, in a simultaneous model, which estimates both the measurement part of the model (production of the latent variable) and the structural part of the model (the additional predictors) at the same time, the measurement part of the model can be affected by the inclusion of covariates (Devlieger & Rosseel, 2017). Two-step GSEMs avoid the problems associated with simultaneous estimation by predicting the measurement model and then constraining the parameters of those variables to their original values when running the full model. However, doing this ignores some of the errors produced in the measurement piece, so standard errors may be underestimated (Bakk & Kuha, 2017).

²² The exclusion rates varied across the regression models, ranging from 1.5% to 8.0% with an average of 6.5%. Sub-questions generated higher rates of exclusions given the smaller number of respondents who answered affirmatively to the larger question. For instance, a smaller number of respondents reported being stopped by the police and, as such, missing responses on sub-questions on their stop experiences generated a larger percentage of missing values. To determine whether any statistically significant differences between included and excluded cases existed for the covariates used in our main models, simple t test comparisons were used. No statistically significant differences were noted between the missing and included cases. As such, these analyses suggested that the data were missing at random and listwise case deletions were appropriate to address these modest missing data problems.

Table 2 Pre Intervention Characteristics and Respondent Outcome Measures: Treatment v. Control

Characteristics	Treatment, N=3,000		Control, N=3,000		d
	Percent / Mean		Percent / Mean		
Asian	9.2%		8.0%		0.042
Black	21.0%		30.0%		0.022
Hispanic	33.0%		30.0%		0.066
White	27.0%		22.0%		0.117
Other	4.2%		4.2%		0.001
Male	45.7%		46.0%		0.007
Age	44.5		44.3		0.006
NYCHA Housing	9.8%		11.8%		0.066
Family/Friend NYPD Officer	17.7%		17.7%		0.002
Some high school or less	8.7%		7.3%		0.052
High school graduate	55.0%		55.6%		0.012
4-yrs college degree or higher	36.3%		37.1%		0.016
Questions	N	Percent / Mean	N	Percent / Mean	d
How safe do you feel in your neighborhood (Likert scale 1-4)	2987	3.30	2978	3.33	0.05
Fair and objective investigation of complaints against officers (Likert scale 1-4)	2786	2.76	2778	2.77	0.01
Car stopped by police in your neighborhood in last 12 months (Yes/No)	2999	11.9%	2999	12.9%	0.03
Stop reason explained (Yes/No)	355	83.4%	385	79.2%	0.10
Physical force used during stop (Yes/No)	353	5.1%	385	5.4%	0.01
Patted down outside clothing (Yes/No)	356	14.5%	384	16.8%	0.06
Search inside clothing, your bag, and/or personal belongings (Yes/No)	356	11.7%	386	15.5%	0.10
Ticket or summons issued (Yes/No)	353	38.9%	384	44.2%	0.10
Arrested (Yes/No)	355	3.5%	386	5.5%	0.09
Stopped by police while walking in neighborhood in last 12 months (Yes/No)	2999	5.9%	2999	4.7%	0.05
Stop reasons explained (Yes/No)	176	59.9%	140	57.8%	0.03
Police gave information card / receipt (Yes/No)	173	11.5%	141	17.9%	0.14
Patted down outside clothing (Yes/No)	177	33.1%	140	51.1%	0.29
Search inside clothing, your bag and/or personal belonging (Yes/No)	177	38.8%	141	46.2%	0.12
Physical force used during stop (Yes/No)	176	14.7%	137	15.2%	0.01
Arrested or summons issued (Yes/No)	177	18.4%	141	24.1%	0.11
Contacted the police for help in your neighborhood in last 12 months (Yes/No)	2986	19.3%	2990	20.1%	0.02
Support for BWC use by police officers (Likert scale 1-4)	2679	3.68	2709	3.66	0.03
Has anyone in your household been stopped by police in last 12 months (Yes/No)	2924	9.6	2931	8.6%	0.03

Table 2 (continued)

Characteristics	Treatment, N=3,000		Control, N=3,000		d
	Percent	Mean	Percent	Mean	
Do you know your neighborhood police officers by name (Yes/No)	2990	12.8%	2929	11.3%	0.05
How often do you go out in neighborhood at night (Likert scale 1-4)	2891	3.14	2883	3.15	0.01
Latent Variables					
Positive Perceptions of Neighborhood Officers	2932	-0.05	2922	-0.03	0.02
Disrust of Neighborhood Officers	2493	0.01	2465	-0.02	0.04
Procedural Justice During Car Stop in Last 12 Months	356	0.00	386	-0.01	0.01
Negative Assessments of Car Stop in Last 12 Months	354	-0.04	376	0.02	0.05
Procedural Justice During Pedestrian Stop in Last 12 Months	177	0.07	139	-0.06	0.11
Negative Assessments of Pedestrian Stop in Last 12 Months	172	-0.06	137	0.05	0.10
Procedural Justice During Contact for Assistance in Last 12 Months	576	-0.06	597	0.00	0.06

The statistical power of an experimental design indicates the likelihood that the null hypothesis will be rejected by a statistical test when a particular alternative hypothesis is in fact true (Lipsey, 1990). The probability of making a type II error of failing to reject a true null hypothesis of “no difference” between treatment and control groups decreases as statistical power increases (estimates range from 0 to 1). In general, a statistical power of .80 is recognized as a desirable level to detect a small effect size (Cohen’s $|d| = .20$). Statistical power in cluster randomized controlled trials is determined by the number of clusters, sample size of clusters, and the intraclass correlation coefficient measuring the degree to which outcomes are correlated within clusters (see Hemming et al., 2017). In this study, the telephone survey had 12,057 observations clustered in 20 precinct pairs. Our power analyses suggested sufficient statistical power to detect small differences between the treatment and control groups in pre-intervention and post-intervention outcomes in the full sample and somewhat diminished power in the smaller subsamples that experienced police contact. This design had statistical power of .80 (alpha = .05) to detect small effect sizes in the telephone survey outcomes for the full sample ($|d| = .035$ to $.139$)²³ and small to medium

²³ Statistical power was calculated from the computed effect sample size that reweights the total sample size (n) by the number of observations per cluster (m=2), the number of matched pairs (k=20), and the intraclass correlations between clusters and pairs (ICC=r1, r2) according to the following formula: effective sample= (n) / (1 + (m-1)*r1+(k-m)*r2). We set the ICC for clusters (r1) to be 0 since the same people were not purposively interviewed in the pre- and post-BWC periods. As a result, the formula for effective sample size is equivalent to the following: (n) / (1 +(k-m)*r2). If the ICC (r2) is equal to 1, then the effective sample becomes the average number of observations for a matched pair. If the ICC is equal to 0, the effective sample size becomes the total sample size (n) (see Kerry & Bland, 1998). For the ICC per pair (k), we rely on a mixed effect regression that estimates the correlations within pairs.

effect sizes in the subsamples of respondents who contacted the NYPD for assistance stop ($N=1,968$; $|d|= .08$ to $.39$;) or experienced a car stop ($N=1,727$; $|d|= .07$ to $.26$) or pedestrian stop ($N=992$; $|d|= .10$ to $.32$).

Statistical analyses that involve multiple comparisons run the risk of reporting “false discoveries” as multiple simultaneous statistical tests are conducted (Miller, 1981). As the number of comparisons increases, it becomes increasingly likely that the two groups being compared will differ on some particular outcome. When compared to analyses that involve only a single outcome as a comparison, confidence in analyses that involve multiple comparison outcomes is generally weaker. Using a single comparison and a conventional two-tailed $p < .05$ statistical significance level, there is only a 5% chance of incorrectly rejecting the null hypothesis when it in fact is true (also known as a “false positive” or type I error). In this study, however, there are 26 simultaneous comparisons made between treatment and control respondents. At the $p < .05$ level, we would expect at least one false-positive test result ($26 * .05 = 1.3$) by chance alone. This relatively large number of comparisons was required to test the multitude of ways that BWCs could impact general citizen perceptions of the NYPD and citizen perceptions of specific NYPD officer behaviors during encounters that included car stops, pedestrian stops, and requests for officer assistance in the past 12 months.

There are many techniques that can be used to correct multiple comparison problems by recalculating probabilities obtained from a statistical test which was repeated multiple times. The traditional Bonferroni method and other family-wise error rate approaches to correct for multiple comparisons have been suggested to be too conservative (Benjamini, 2010). These methods risk missing many true findings by imposing stringent safeguards which control the probability of making at least one type I error. In this analysis, we used the false discovery rate (FDR) approach (Benjamini & Hochberg, 1995). FDR procedures control the expected proportion of false discoveries (incorrectly rejected null hypotheses). The FDR method generates an adjusted p value known as the q value that assesses false positive rates and allows for an interpretation of risk levels when rejecting null hypotheses (Newsom, 2010).²⁴ Like p values, q values range from 0 to 1, with $q < .05$ suggesting a bona fide statistically significant difference between treatment and control groups. A q value = 1 suggests the result is not statistically significant under any circumstances. For all outcome measures, the FDR procedure was used to determine whether any significant results calculated through traditional p values generated by the DID estimators in our models were actually “false discoveries.”

Results

Table 3 presents the results of the regressions comparing differences in survey subject responses over the course of the pre-intervention and post-intervention time periods in the BWC treatment precincts relative to the control precincts. The q values associated with the DID estimates show that presence of BWCs in the treatment precincts did not generate any statistically significant changes in telephone survey subject responses between the pre-

²⁴ For instance, a p value = .05 for a particular statistical comparison suggests that 5% of all tests will result in false positives (e.g., proportion of all tested subjects who do not have a disease who will be identified as having the disease). In contrast, a q value = .05 for a particular statistical comparison suggests that 5% of statistically significant results will be false discoveries (e.g., the proportion of all subjects identified as having the disease who do not actually have the disease).

Table 3 Probit, ordered probit, and generalized structural equation models of outcomes on differences-in-differences and control variables

	Weighted model	DID			Pseudo R ²	N
		Coef. (RSE)	<i>p</i>	<i>q</i>		
Outcome variables						
How safe do you feel in your neighborhood (Likert scale 1–4)	Ordered probit	0.047 (0.063)	0.452	1	0.039	11004
Fair and objective investigations of complaints against officers (Likert scale 1–4)	Ordered probit	0.069 (0.064)	0.277	1	0.021	10383
Car stopped by police in your neighborhood in last 12 months (yes/no)	Probit	0.026 (0.093)	0.777	1	0.043	11082
Stop reasons explained (yes/no)	Probit	0.162 (0.164)	0.323	1	0.038	1590
Physical force used during stop (yes/no)	Probit	−0.593 (0.328)	0.071	1	0.134	1606
Patted down outside clothing (yes/no)	Probit	−0.309 (0.233)	0.185	1	0.135	1608
Searched inside clothing, your bag, and/or personal belongings (yes/no)	Probit	−0.068 (0.207)	0.745	1	0.121	1605
Ticket or summons issued (yes/no)	Probit	0.213 (0.225)	0.344	1	0.010	1603
Arrested (yes/no)	Probit	−0.273 (0.370)	0.460	1	0.062	1610
Stopped by police while walking in neighborhood in last 12 months (yes/no)	Probit	−0.201 (0.114)	0.078	1	0.059	11082
Stop reasons explained (yes/no)	Probit	−0.283 (0.173)	0.100	1	0.017	924
Police gave information card / receipt (yes/no)	Probit	0.050 (0.264)	0.850	1	0.047	917
Patted down outside clothing (yes/no)	Probit	0.268 (0.259)	0.301	1	0.082	930
Searched inside clothing, your bag, and/or personal belongings (yes/no)	Probit	0.482 (0.268)	0.072	1	0.035	928
Physical force used during stop (yes/no)	Probit	0.070 (0.291)	0.810	1	0.037	924
Arrested or summons issued (yes/no)	Probit	−0.282 (0.319)	0.376	1	0.062	930
Contacted the police for help in your neighborhood in last 12 months (yes/no)	Probit	0.039 (0.076)	0.607	1	0.028	11023
Support for BWC use by police officers (Likert scale 1–4)	Ordered Probit	−0.070 (0.071)	0.326	1	0.004	10526
Has anyone in your household been stopped by poke in last 12 months (yes/no)	Probit	−0.241 (0.118)	0.040 *	1	0.032	10718
Do you know our neighborhood police officers by name (yes/no)	Probit	−0.066 (0.115)	0.568	1	0.040	11047
Latent variables						
Positive Perceptions of Neighborhood Officers	GSEM	0.063 (0.063)	0.328	1		9394
Distrust of Neighborhood Officers	GSEM	−0.162 (0.069)	0.031 *	1		8681
Procedural Justice During CarStop in the Last 12 Months	GSEM	−0.057 (0.128)	0.662	1		1543
Negative Assessments of Car Stop in the Last 12 Months	GSEM	−0.153 (0.180)	0.405	1		1511
Procedural Justice During Pedestrian Stop in the Last 12 Months	GSEM	−0.292 (0.189)	0.139	1		883
Negative Assessments of Pedestrian Stop in the Last 12 Months	GSEM	0.176 (0.202)	0.396	1		886
Procedural Justice During Contact for Assistance in the Last 12 Months	GSEM	0.046 (0.111)	0.685	1		1685

Table 4 Probits, ordered probits, and generalized structural equation models by respondent race

	White respondents (N= 3103)				Black respondents (N = 3595)				Hispanic respondents (N = 3884)			
	DID		N	q	DID		N	q	DID		N	q
	Coef (RSE)	p			Coef (RSE)	p			Coef (RSE)	p		
Neighborhood safety	0.103 (0.137)	0.455	1	2915	0.014 (0.067)	0.838	1	3376	0.069 (0.083)	0.408	1	3431
Complaint investigation	0.054 (0.111)	0.624	1	2753	0.007 (0.064)	0.916	1	3217	0.064 (0.104)	0.537	1	3206
Car stop	0.059 (0.202)	0.769	1	2929	0.126 (0.090)	0.161	1	3404	-0.085 (0.088)	0.331	1	3461
Explained reason	-0.104 (0.349)	0.766	1	287	0.060 (0.189)	0.749	1	615	4.111 (0.238)	0.641	1	501
Used force	-0.998 (0.438)	0.023 *	1	240	-0.107 (0.322)	0.739	1	619	4.175 (0.347)	0.614	1	509
Frisked	-0.314 (0.452)	0.072	1	289	-0.013 (0.232)	0.957	1	622	0.313 (0.224)	0.162	1	509
Searched	-0.657 (0.412)	0.110	1	289	0.106 (0.136)	0.569	1	622	0.266 (0.211)	0.209	1	507
Issued summons	-0.079 (0.266)	0.768	1	290	0.491 (0.261)	0.060	1	621	0.148 (0.163)	0.364	1	506
Arrested	-0.535 (0.502)	0.286	1	290	0.172 (0.353)	0.627	1	529	0.485 (0.373)	0.193	1	508
Pedestrian stop	-0.006 (0.134)	0.965	1	2929	0.010 (0.141)	0.946	1	3404	-0.304 (0.157)	0.053	1	3461
Explained reason	-0.656 (0.396)	0.093	1	152	-0.419 (0.282)	0.137	1	377	-0.176 (0.243)	0.469	1	302
Provided receipt	-0.416 (0.455)	0.360	1	151	-0.229 (0.375)	0.542	1	377	0.170 (0.438)	0.698	1	296
Frisked	-0.228 (0.478)	0.633	1	153	0.265 (0.256)	0.302	1	380	0.194 (0.275)	0.479	1	304
Searched	-0.403 (0.381)	0.290	1	152	0.672 (0.218)	0.002 **	0.403	379	0.354 (0.320)	0.269	1	305
Arrested/summons	-0.342 (0.536)	0.524	1	150	-0.316 (0.260)	0.223	1	381	-0.076 (0.393)	0.847	1	306
Contacted for assistance	0.002 (0.139)	0.988	1	2913	-0.051 (0.101)	0.615	1	3389	0.033 (0.079)	0.676	1	3441
Favor BWCs	-0.051 (0.080)	0.524	1	2758	-0.015 (0.072)	0.833	1	3222	0.027 (0.075)	0.713	1	3338
Household stopped	-0.333 (0.120)	0.005 **	0.672	2853	-0.065 (0.104)	0.533	1	3262	-0.149 (0.125)	0.234	1	3365
Know officers	-0.125 (0.137)	0.364	1	2919	0.019 (0.091)	0.836	1	3395	-0.116 (0.106)	0.276	1	3448

Table 4 (continued)

	White respondents (N= 3103)			Black respondents (N = 3595)			Hispanic respondents (N = 3884)									
	N		q	N		q	N		q							
	Coef (RSE)	p		Coef (RSE)	p		Coef (RSE)	p								
LV: Positive perceptions of neighborhood officers	0.076	(0.069)	0.269	1	2424	1	-0.032	(0.054)	0.557	1	2910	-0.028	(0.080)	0.727	1	2930
LV: Negative perceptions of neighborhood Officers	-0.168	(0.077)	0.030 *	1	2195	1	0.043	(0.083)	0.606	1	2736	-0.032	(0.094)	0.737	1	2770
LV: Procedural justice perceptions during car stop	-0.006	(0.226)	0.977	1	284	1	-0.053	(0.141)	0.689	1	592	-0.151	(0.180)	0.401	1	487
LV: Negative perceptions during car stop	-0.329	(0.174)	0.058	1	282	1	0.008	(0.198)	0.969	1	576	0.509	(0.183)	0.005 **	0.672	475
LV: Procedural justice perceptions during pedestrian stop	-0.192	(0.283)	0.498	1	145	1	-0.047	(0.208)	0.821	1	360	-0.313	(0.315)	0.320	1	289
LV: Negative perceptions during pedestrian stop	0.314	(0.269)	0.243	1	144	1	0.306	(0.246)	0.214	1	369	0.173	(0.268)	0.518	1	287
LV: Procedural justice perceptions during contact for assistance	0.169	(0.181)	0.350	1	485	1	-0.359	(0.162)	0.026 *	1	481	0.277	(0.116)	0.017 *	1	501

*p<0.05, **p<.01

Covariates included but not shown: race, ethnicity, gender, age, education, and NYPD friend/family

All models used robust standard errors clustered by precinct matched pair. GSEM, generalized structural equation model; LV, latent variable

intervention and post-intervention periods when compared to the control subject responses. Relative to control subjects, subjects in the BWC treatment areas did not report any differences over time in their perceptions of neighborhood safety, how complaints would be handled by the NYPD, their knowledge of NYPD officer names in their neighborhood, contacted the NYPD for assistance, and whether they personally had been stopped by the NYPD between the pre-intervention and post-intervention periods. For those subjects who had experienced a pedestrian and/or car stop, no significant differences were reported by subjects in BWC areas relative to control areas over time in terms of officer behaviors during the stop. Relative to control subjects, treatment subjects did report that members of their households (other than the respondent) were less likely to be stopped by the NYPD after the adoption of the BWCs in their precincts, but this difference had a q value of 1 indicating this difference was not statistically meaningful and a false discovery.

Table 3 also presents the results of the model estimating the differences in survey subject responses for the items comprising the outcome variables assessed as latent measure perceptions of the NYPD. The generalized structural equations suggested there were no bona fide statistically significant changes in the perceptions held by telephone survey subjects in the BWC treatment precincts relative to subjects in the control precincts. While the p value associated with the DID estimates suggested that their negative perceptions of police officer behaviors decreased with the introduction of BWCs (i.e., became more positive), the FDR analysis suggests this finding is a false discovery ($q = 1$).

Table 4 shows the results from an exploratory analysis of responses by the race/ethnicity of the respondents (white, black, and Hispanic). As the q values in Table 5 reveal, there were no genuine statistically significant differences in citizen perceptions of the NYPD generated by the deployment of BWCs on NYPD officers in the treatment relative to control precincts among varying racial groups of respondents.

Conclusion

This controlled experimental evaluation tested the impacts of outfitting NYPD officers with BWC cameras on broader community perceptions of the police and their experiences during their most recent encounter with NYPD officers. Telephone surveys were administered to representative samples of adult residents in the 40 NYPD precincts (20 matched pairs) with the highest number of civilian complaints. The results indicate that the deployment of BWCs on NYPD officers working in treatment precincts did not produce any statistically significant differences in resident perceptions of the NYPD and their experiences with NYPD officers over the course of the trial relative to the perceptions held by residents in the control precincts.

Existing studies have generally not tested whether BWC deployments change broader community perceptions of the police (Lum et al., 2019). As discussed earlier, media coverage and survey research suggest community members support the adoption of body-worn cameras by police departments and have high expectations for the technology in increasing police accountability and improving citizen confidence in the police (Crow et al., 2017; Taylor et al., 2017; Todak et al., 2018). However, the one existing study on changes in public perceptions of the police associated with BWCs employed a nonexperimental evaluation of a UK police department that had high levels of citizen support and satisfaction for the police before BWCs were adopted (Ellis et al., 2015). In NYC, residents expressed

similar enthusiasm for the adoption of BWCs by the NYPD (NYU Policing Project, 2016). During the pre-intervention period, our research found that surveyed residents in the experimental precincts held mixed opinions of the NYPD: less than 60% of telephone survey respondents reported somewhat favorable or very favorable feelings towards NYPD officers in their neighborhood. Similar to the UK study (Ellis et al., 2015), our experimental evaluation found that the BWC deployment did little to change these pre-existing resident perceptions of the NYPD and their encounters with NYPD officers.

This research is limited by the relatively low response rate by prospective telephone survey subjects. While our analyses suggest that nonresponse bias is not a problem in comparing public perceptions between respondents in BWC treatment and control precincts, we do not know whether nonrespondents differed between the treatment and control precincts in some unmeasured systematic way that could lead to different conclusions about changing perceptions of the NYPD. The relatively small number of treatment and control respondents who reported being stopped by the NYPD in the past 12 months also may not reflect the experiences of the broader population of people stopped by the NYPD during the pre-intervention and post-intervention periods. Due to New York State Criminal Procedure Law 140.50(4) prohibiting the entry of these data elements, the NYPD does not maintain computerized records of the names and dates of birth of individuals who were stopped by its officers. As such, it was not possible to design a data collection strategy that conducted follow-up surveys with the subjects of NYPD stop reports. The NYPD monitor team complemented the telephone survey with in-person community surveys of a much smaller number of respondents in five matched pairs of treatment and control precincts (one pair in each borough). These findings did not differ from the telephone survey results (and are available at <http://nypdmonitor.org/wp-content/uploads/2020/12/12th-Report.pdf>).

It is also important to note that our results may not be generalizable beyond the unique context of the 40 precincts in NYC. Our findings diverge from existing research suggesting that the presence of BWCs enhances citizen perceptions of procedural justice during their encounters with the police. We think these differences may, in part, be due to varying methodological approaches. Since NYPD policy requires officers to notify citizens that encounters are recorded, we did not explicitly ask whether treatment respondents noticed BWCs on officers during their encounters. As a result, we are not able to do a subgroup analysis of respondents who did notice BWC presence during their encounter.²⁵ Demir et al. (2020) conducted in-person interviews with stopped motorists immediately following their encounter with traffic police officers. Consistent with the US Bureau of Justice Statistics (2018) Police-Public Contact Survey methodology, our research asked respondents if they experienced a pedestrian stop, an automobile stop, and/or a contact for assistance within the past 12 months, and, if they affirmed, then we collected data on their perceptions of police behaviors during the encounter. While we did not ask how long ago their contacts with police occurred, it is possible that any positive perceptions of procedurally just behaviors by officers decayed in the time between the encounter and subsequent interview and did not exert much of an influence on their recollection of NYPD officer behaviors and general

²⁵ As mentioned earlier, NYPD policy requires officers to notify citizens that an interaction is being recorded “as soon as reasonably practical.” BWC officer compliance with this policy requirement was not quantifiable. The surveys conducted in this research study assumed that NYPD officers outfitted with BWCs were indeed notifying citizens that encounters were being captured on video. However, when officers are not required to provide notification, research suggests most citizens are not aware that the encounters are being recorded by BWCs (White et al., 2017).

assessments of the NYPD. However, other studies suggest that citizens often have detailed memories of their interactions with the police, especially when those interactions generate negative perceptions of officer behaviors (Brunson, 2007; Rios, 2011).

Research underscores the importance of examining direct and vicarious associations between police contacts and appraisals of the police (Rosenbaum et al., 2005; Weitzer & Tuch, 2005; Brunson, 2007). Like the asymmetrical effects noted in other cities (Skogan, 2006), the Vera Institute of Justice survey found that positive experiences with the police were not associated with substantially higher levels of confidence in the NYPD while negative experiences were associated with low confidence levels (Miller et al., 2004). Further, across nine monthly surveys of NYC residents, the Vera study reported that citizen perceptions of the police, whether positive or negative, were quite durable over time (Miller et al., 2004). The placement of BWCs on NYPD officers seemed to improve the civility of police-citizen encounters as evidenced by a 21% reduction in citizen complaints against BWC treatment officers relative to control officers in the larger evaluation that complements this study (Braga et al., 2021). However, citizen complaints of poor police behavior during encounters are fortunately rare events (Ariel et al., 2015; Braga et al., 2018). NYPD officers in the treatment and control groups generated, on average, only one citizen complaint for every 4 years of service. While noteworthy, the reduction in a very low base rate event may not be a powerful enough change over a long enough time period to generate a meaningful shift in durable resident perceptions of the NYPD.

As a policy intervention, the results of this experimental evaluation suggest that police department adoptions of BWCs are not a panacea to police-community relation problems. Indeed, our exploratory analysis of the data did not detect any divergences in the perceptions of the police held by white, black, and Hispanic subjects in the treatment precincts relative to control precincts. The NYPD and other police departments may be better served by doubling down on other programs that have solid scientific evidence of enhancing community attitudes towards the police. For instance, evaluations generally show that citizen perceptions of police performance, satisfaction, and legitimacy are improved by community policing programs (Gill et al., 2014; National Academies of Sciences, 2018). While the growing evidence base is not yet strong enough to support causal assertions (Nagin & Telep, 2017), studies show that citizen perceptions of procedural justice during their encounters with the police are associated with increased perceptions of police legitimacy and cooperation with the police (Tyler, 2006; National Academies of Sciences, 2018). BWCs seem to be helpful in inspiring officers to be procedurally just during encounters with citizens (McCluskey et al., 2019; Demir et al., 2020); however, police departments should be formally training their officers to embrace procedural justice principles during all interactions with the public and not assume the BWC technology will generate positive interactions.

This research suggests that BWCs are unlikely to lead to short-term changes in public perceptions of the police. However, it remains possible that the BWC technology could produce longer-term benefits. When controversial events happen, the public expects to see video of the police-citizen encounter so they can judge whether officers acted lawfully and behaved appropriately. The absence of BWC video capturing a highly controversial police-citizen encounter could result in diminished police legitimacy, generate expensive overtime costs when managing associated citizen protests, and be particularly detrimental should protests become destructive riots. At the very least, the presence of BWCs on officers suggests to community members that mechanisms exist to ensure transparency and hold officers accountable when they

misbehave. And, as a component of a broader set of evidence-based strategies to improve community perceptions, the placement of BWCs on officers could help to enhance the legitimacy of the police to the public they serve.

Appendix 1.

Table 5 Comparison of treatment and control precinct characteristics

	Treatment Mean	Control Mean	Std. Mean Difference	Combined K-S D
Population	115,557.2	102,705.1	.143	.250
NYCHA population	4,986.2	4,609.2	.031	.200
Concentrated disadvantage	.247	.409	-.077	.250
Officers	228.5	215.9	.118	.200
Major crimes	1,567.6	1,413.4	.161	.250
Arrests	4,878.1	4,820.9	.016	.300
Arrests w/force	63.6	66.9	-.046	.250
911 calls	72,730.4	70,269.5	.064	.250
CCRB	67.4	67.5	-.001	.350

N= 40 (20 treatment precincts, 20 control precincts)

Notes: The standardized mean differences are beta coefficients generated by ordinary least squares regressions of each precinct characteristic on group assignment. Meaningful differences between treatment and control groups would be noted by standardized mean differences in excess of |.20|. The distributions of treatment and control precinct characteristics were also compared using two-sample Kolmogorov-Smirnov (K-S) tests for equality of distribution functions. Values closer to 0 indicate more comparability between distributions

Source: Braga et al. (2021)

Appendix 2.

Table 6 Comparison of pre-intervention outcome measures and characteristics of NYPD officers working the third platoon and anti-crime units in treatment and control precincts

	Treatment	Control	Std. mean Difference	Combined K-S D
Officer sex				
% male	84.2%	84.0%	-.003	.002
% female	15.8%	16.0%		
Officer race/ethnicity				
% White	47.2%	47.0%	.002	.002
% Hispanic	30.7%	28.9%	.020	
% Black	13.2%	14.8%	-.024	
% Asian/other	8.9%	9.3%	-.007	
Officer assignment				
% Third platoon patrol	89.7%	90.1%	.007	.004
% Anti-crime	10.3%	9.9%		

Table 6 (continued)

	Treatment	Control	Std. mean Difference	Combined K-S D
Officer rank				
% Police officer	90.9%	90.3%	-.010	.006
% Sergeant	9.1%	9.7%		
	Mean (SD)	Mean (SD)		
Officer age	29.82 (6.55)	29.85 (6.59)	-.002	.011
Officer years on the job	4.33 (3.99)	4.41 (5.21)	-.007	.028
Stop reports	.72 (1.76)	.73 (1.81)	-.003	.006
Arrests	10.39 (11.60)	10.67 (12.37)	-.012	.021
Arrests with force	.13 (.46)	.15 (.47)	-.022	.017
Summons	8.85 (17.13)	8.18 (14.11)	.021	.033
Domestic incident reports	.32 (2.31)	.23 (1.78)	.020	.024
Crime complaint reports	34.39 (29.99)	33.36 (29.68)	.017	.024
CCRB complaints	.26 (.62)	.27 (.63)	-.011	.009

N = 3,889 (1,991 treatment officers, 1,898 control officers)

Notes: *SD*, standard deviation. The standardized mean differences are beta coefficients generated by ordinary least squares regressions of each precinct characteristic on group assignment. The pre-intervention outcome measures are based on 12-month counts. Meaningful differences between treatment and control groups would be noted by standardized mean differences in excess of $|\pm 0.20|$. The distributions of treatment and control precinct characteristics were also compared using two-sample Kolmogorov-Smirnov (K-S) tests for equality of distribution functions. Values closer to 0 indicate more comparability between distributions. (Braga et al., 2021)

Appendix 3

Table 7 Comparison of Demographics of Telephone Survey Respondents in Treatment and Control Precincts

Weighted Demographics of Adult Telephone Survey Respondents in Control Precincts	Survey sample %	Census (ACS 2010) %
Men	46	46
Women	54	54
Ages 18–34	36	37
Ages 35–49	26	27
Ages 50–64	21	22
Ages 65/older	16	14
Don't know/refused	1	-
White (non-Hispanic)	27	27
African American	30	35
Hispanic	30	30
Asian	8	9
Manhattan	14	15
Bronx	20	21

Table 7 (continued)

	Survey sample %	Census (ACS 2010) %
Queens	28	27
Brooklyn	38	37
Weighted Demographics of Adult telephone Survey Respondents in Treatment Precincts	Survey sample %	Census (ACS 2010) %
Men	46	46
Women	54	54
Ages 18–34	36	36
Ages 35–49	25	26
Ages 50–64	22	23
Ages 65/older	16	15
Don't know/refused	1	-
Whites (non-Hispanic)	22	22
African Americans	31	36
Hispanics	33	33
Asians	9	9
Manhattan	18	18
Bronx	25	26
Queens	21	22
Brooklyn	31	30
Staten Island	5	5

Acknowledgements The Braga and MacDonald currently serve on the federal monitor team pursuant to the court orders in *Floyd et al. v. City of New York, et al.*, 08 Civ. 1034 (AT); *Ligon, et al., v City of New York, et al.*, 12-CV-2274 (AT); and *Davis et al., v. City of New York, et al.*, 10-CV-00699 (AT). The analyses and opinions expressed in the article reflect those of the authors only and not any other entity. The work reported in this paper was funded by the City of New York as part of the court orders. The authors would like to thank Peter Zimroth, Richard Jerome, other monitor team members, and the NYPD for their assistance with this research.

References

- American Association for Public Opinion Research [AAPOR]. (2010) New considerations for survey researchers when planning and conducting RDD telephone surveys in the U.S. with respondents reached via cell phone numbers. <https://www.aapor.org/Education-Resources/Reports/Cell-Phone-Task-Force-Report.aspx>. Accessed 21 Jan 2017
- Ariel, B. (2016a). Increasing cooperation with the police using body worn cameras. *Police Quarterly*, *19*, 326–362.
- Ariel, B. (2016b). Police body cameras in large police departments. *Journal of Criminal Law and Criminology*, *106*, 729–768.
- Ariel, B., Farrar, W. A., & Sutherland, A. (2015). The effect of police body-worn cameras on use of force and citizens' complaints against the police: A randomized controlled trial. *Journal of Quantitative Criminology*, *31*, 509–535.

- Ariel, B., Sutherland, A., & Sherman, L. (2019). Preventing treatment spillover contamination in criminological field experiments: The case of body-worn police cameras. *Journal of Experimental Criminology*, *15*, 569–591.
- Bakk, Z., & Kuha, J. (2017). Two-step estimation of models between latent classes and external variables. *Psychometrika*, *83*, 871–892.
- Benjamini, Y. (2010). Discovering the false discovery rate. *Journal of the Royal Statistical Society, Series B*, *72*, 405–416.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*, 289–300.
- Bloomberg, M. (2013). *Stop and frisk keeps New York Safe* (p. 6). The Washington Post.
- Bollen, K. (1989). *Structural equations with latent variables*. Wiley.
- Bollen, K., & Long, J. (Eds.). (1993). *Testing structural equation models*. Sage Publications.
- Braga, A., Winship, C., Tyler, T., Fagan, J., & Meares, T. (2014). The salience of social contextual factors in appraisals of police interactions with citizens: A randomized factorial experiment. *Journal of Quantitative Criminology*, *30*, 599–627.
- Braga, A., Sousa, W., Coldren Jr., J. R., & Rodriguez, D. (2018). The effects of body-worn cameras on police activity and police-citizen encounters: A randomized controlled trial. *Journal of Criminal Law and Criminology*, *108*, 511–538.
- Braga, A., Barao, L., Zimmerman, G., Douglas, S., & Sheppard, K. (2020). Measuring the direct and spillover effects of body worn cameras on the civility of police-citizen encounters and police work activities. *Journal of Quantitative Criminology*, *36*, 851–876.
- Braga, A., MacDonald, J., & McCabe, J. (2021). *Body worn cameras, lawful police stops, and NYPD officer compliance: A cluster randomized controlled trial*. Unpublished paper available upon request from the authors.
- Brown, T. (2015). *Confirmatory factor analysis for applied research (2nd Ed.)*. The Guilford Press.
- Brunson, R. (2007). 'Police don't like black people': African American young men's accumulated police experiences. *Criminology & Public Policy*, *6*, 71–102.
- Campbell, M., Donner, A., & Klar, N. (2007). Developments in cluster randomized trials and statistics in medicine. *Statistics in Medicine*, *26*, 2–19.
- Card, D., & Krueger, A. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, *84*, 772–793.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum Associates.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.
- Crow, M., Snyder, J., Crichlow, V., & Smykla, J. O. (2017). Community perceptions of police body-worn cameras: The impact of views on fairness, fear, performance, and privacy. *Criminal Justice and Behavior*, *44*, 589–610.
- Culhane, S. E., Bowman, J. H., & Schweitzer, K. (2016). Public perceptions of the justifiability of police shootings: The role of body cameras in a pre- and post-Ferguson experiment. *Police Quarterly*, *19*, 251–274.
- Demir, M., Apel, R., Braga, A., Brunson, R., & Ariel, B. (2020). Body worn cameras, procedural justice, and police legitimacy: A controlled experimental evaluation of traffic stops. *Justice Quarterly*, *37*, 53–84.
- Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis: An alternative for SEM? *Methodology*, *13*, 31–38.
- Devlieger, I., Mayer, A., & Rosseel, Y. (2015). Hypothesis testing using factor score regression: A comparison of four methods. *Education and Psychological Measurement*, *76*, 741–770.
- Duval, T. S., & Wicklund, R. A. (1972). *A theory of objective self-awareness*. Academic.
- Ellis, T., Jenkins, C., & Smith, P. (2015). *Evaluation of the introduction of personal issue body worn video cameras (Operation Hyperion) on the Isle of Wight: Final report to Hampshire Constabulary*. University of Portsmouth.
- Fagan, J., & Davies, G. (2000). Street stops and broken windows: Terry, race, and disorder in New York City. *Fordham Urban Law Journal*, *28*, 457–504.
- Fagan, J., Zimring, F., & Kim, J. (1998). Declining homicide in New York: A tale of two trends. *Journal of Criminal Law and Criminology*, *88*, 1277–1324.
- Fagan, J., Geller, A., Davies, G., & West, V. (2010). Street stops and broken windows revisited: The demography and logic of proactive policing in a safe and changing city. In S. Rice & M. White (Eds.), *Race, ethnicity, and policing* (pp. 309–348). New York University Press.
- Fratello, J., Rengifo, A., Trone, J., & Velazquez, B. (2013). *Coming of age with stop and frisk: Experiences, perceptions, and public safety implications*. Vera Institute of Justice.
- 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*, 813–823.

- Gill, C., Weisburd, D., Telep, C., Vitter, Z., & Bennett, T. (2014). Community-oriented policing to reduce crime, disorder and fear and increase satisfaction and legitimacy among citizens: A systematic review. *Journal of Experimental Criminology*, *10*, 399–428.
- Goodison, S., & Wilson, T. (2017). *Citizen perceptions of body worn cameras: A randomized controlled trial*. Police Executive Research Forum.
- Harcourt, B., & Ludwig, J. (2006). Broken windows: New evidence from New York City and a five-city experiment. *University of Chicago Law Review*, *73*, 271–320.
- Hemming, K., Eldridge, S., Forbes, G., & Weijer, C. (2017). How to design efficient cluster randomised trials. *British Medical Journal*, *358*. <https://doi.org/10.1136/bmj.j3064>.
- Hoshino, T., & Bentler, P. (2013). Bias in factor score regression and a simple solution. In A. R. de Leon & K. C. Chough (Eds.), *Analysis of mixed data: Methods and applications* (pp. 43–61). Chapman & Hall.
- Hurst, Y. & Frank, J. (2000). How kids view cops: The nature of juvenile attitudes toward the police. *Journal of Criminal Justice*, *28*, 189–202.
- Hyland, S. (2018). *Body-worn cameras in law enforcement agencies, 2016*. US Bureau of Justice Statistics.
- Imai, K., King, G., & Stuart, E. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of The Royal Statistical Society: Series A (Statistics in Society)*, *171*, 481–502.
- Jennings, W. G., Lynch, M. D., & Fridell, L. A. (2015). Evaluating the impact of police officer body-worn cameras (BWCs) on response-to-resistance and serious external complaints: Evidence from the Orlando Police Department (OPD) experience utilizing a randomized controlled experiment. *Journal of Criminal Justice*, *43*, 480–486.
- Kelling, G., & Coles, C. (1996). *Fixing broken windows*. Simon & Schuster.
- Kelly, R. (2015). *Vigilance: My life serving America and protecting its empire city*. Hachette Books.
- Kerrison, E. M., Cobbina, J., & Bender, K. (2018). Stop-gaps, lip service, and the perceived futility of body-worn police officer cameras in Baltimore City. *Journal of Ethnic & Cultural Diversity in Social Work*, *27*, 271–288.
- Kerry, S. M., & Bland, J. (1998). The intracluster correlation coefficient in cluster randomisation. *British Medical Journal*, *316*, 1455–1460.
- Kim, J.-O., & Mueller, C. (1978). *Introduction to factor analysis. Quantitative applications in the social sciences, paper 13*. Sage Publications.
- Leiber, M., Nalla, M., & Farnsworth, M. (1998). Explaining juveniles' attitudes toward the police. *Justice Quarterly*, *15*, 151–174.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Sage Publications.
- Long, J. S. (1983). *Confirmatory factor analysis. Quantitative applications in the social sciences, paper 33*. Sage Publications.
- Lum, C. M., Koper, C. S., Merola, L. M., Scherer, A., & Reixou, A. (2015). *Existing and ongoing body worn camera research: Knowledge gaps and opportunities*. Laura and John Arnold Foundation.
- Lum, C., Stoltz, M., Koper, C. S., & Scherer, J. A. (2019). The research on body-worn cameras: What we know, what we need to know. *Criminology & Public Policy*, *18*, 93–118.
- Lum, C., Koper, C. S., Wilson, D. B., Stoltz, M., Goodier, M., Eggins, E., Higginson, A., & Mazerolle, L. (2020). Body-worn cameras' effects on police officers and citizen behavior: A systematic review. *Campbell Systematic Reviews*, *16*, 1–40.
- MacDonald, J., & Stokes, R. (2006). Race, social capital, and trust in the police. *Urban Affairs Review*, *41*, 358–375.
- MacDonald, J., Stokes, R., Ridgeway, G., & Riley, K. J. (2007). Race, neighborhood context, and perceptions of injustice by the police in Cincinnati. *Urban Studies*, *44*, 2567–2585.
- MacDonald, J., Fagan, J., & Geller, A. (2016). The effects of local police surges on crime and arrests in New York City. *PLoS ONE*, *11*, e0157223.
- Magill, N., Knight, R., McCrone, P., Ismail, K., & Landau, S. (2019). A scoping review of the problems and solutions associated with contamination in trials of complex interventions in mental health. *BMC Medical Research Methodology*, *19*, 4. <https://doi.org/10.1186/s12874-018-0646-z>.
- Malm, A., LaVigne, N., & Lawrence, D. S. (2016). *Cameras and police legitimacy*. Paper presented at the annual meeting of Western Society of Criminology.
- Mastrofski, S. D., Snipes, J. B., & Supina, A. E. (1996). Compliance on demand: The public's response to specific police requests. *Journal of Research in Crime and Delinquency*, *33*, 269–305.
- Mazerolle, L., Antrobus, E., Bennett, S., & Tyler, T. (2013). Shaping citizen perceptions of police legitimacy: A randomized field trial of procedural justice. *Criminology*, *51*, 33–63.
- McClure, D., La Vigne, N., Lynch, M., Golian, L., Lawrence, D., & Malm, A. (2017). *How body cameras affect community members' perceptions of police: Results from a randomized controlled trial of one agency's pilot*. California State University and Urban Institute.

- McCluskey, J., Uchida, C., Solomon, S. E., Wooditch, A., Connor, C., & Revier, L. (2019). Assessing the effects of body-worn cameras on procedural justice in the Los Angeles police department. *Criminology*, *57*, 208–236.
- Miller, R. (1981). *Simultaneous statistical inference* (2nd ed.). Springer Verlag.
- Miller, J., Davis, R., Henderson, N., Markovic, J., & Ortiz, C. (2004). *Public opinions of the police: The influence of friends, family, and news media*. Vera Institute of Justice.
- Mosteller, F., & Boruch, R. (Eds.). (2002). *Evidence matters: Randomized trials in education research*. Brookings Institution Press.
- Nagin, D. S., & Telep, C. W. (2017). Procedural justice and legal compliance. *Annual Review of Law and Social Science*, *13*, 5–28.
- New York University School of Law Policing Project. (2016). *Report to the NYPD summarizing public feedback on its proposed body-worn camera policy*. New York University School of Law.
- Newsom, R. (2010). Frequentist q-values for multiple-test procedures. *The Stata Journal*, *10*, 568–584.
- Paternoster, R., Brame, R., Bachman, R., & Sherman, L. (1997). Do fair procedures matter? *Law & Society Review*, *17*, 457–479.
- President's Task Force on 21st Century Policing. (2015). *Final report of the President's task force on 21st century policing*. Office of Community Oriented Policing Services.
- Reisig, M. D., Bratton, J., & Gertz, M. (2007). The construct validity and refinement of process-based policing measure. *Criminal Justice and Behavior*, *34*, 1005–1028.
- Ridgeway, G. (2007). *Analysis of racial disparities in the New York Police Department's stop, question, and frisk practices*. RAND Corporation.
- Rios, V. (2011). *Punished: Policing the lives of black and Latino boys*. New York University Press.
- Rogers, W. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin*, *13*, 19–23.
- Rosenbaum, D. P., Schuck, A., Costello, S., Hawkins, D., & Ring, M. (2005). Attitudes toward the police: The effects of direct and vicarious experience. *Police Quarterly*, *8*, 343–365.
- Rosenbaum, D. P., Lawrence, D. S., Hartnett, S., McDevitt, J., & Posick, C. (2015). Measuring procedural justice and legitimacy at the local level: The police community interaction survey. *Journal of Experimental Criminology*, *11*, 335–366.
- Rubin, D. (1980). Randomization analysis of experimental data: The fisher randomization test. *Journal of the American Statistical Association*, *75*, 591–593.
- Sahin, N., Braga, A. A., Apel, R., & Brunson, R. K. (2017). The impact of procedurally-just policing on citizen perceptions of police during traffic stops: The Adana randomized controlled trial. *Journal of Quantitative Criminology*, *33*, 701–726.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for general causal inference*. Wadsworth.
- Skogan, W. (2006). Asymmetry in the impact of encounters with police. *Policing & Society*, *16*, 99–126.
- Sousa, W. H., Miethe, T. D., & Sakiyama, M. (2018). Inconsistencies in public opinion of body-worn cameras on police: Transparency, trust, and improved police–citizen relationships. *Policing: A Journal of Policy and Practice*, *12*, 100–108.
- Stanley, J. (2015). *Police body-mounted cameras: With right policies in place, a win for all*. American Civil Liberties Union.
- StataCorp. (2019). *Stata survey data reference manual: Release 16*. Stata Press.
- Sunshine, J., & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review*, *37*, 513–548.
- Taylor, T., Turner, K., Esbensen, F. A., & Winfree, T. (2001). Coppin' an attitude: Attitudinal differences among juveniles toward the police. *Journal of Criminal Justice*, *29*, 295–305.
- Taylor, E., Lee, M., Willis, M., & Gannoni, A. (2017). Police detainee perspectives on police body-worn cameras. *Trends & Issues in Crime & Criminal Justice*, *537*, 1–14.
- Todak, N., Gaub, J., & White, M. D. (2018). The importance of external stakeholders for police body-worn camera diffusion. *Policing: An International Journal of Police Strategies & Management*, *41*, 448–464.
- Torgerson, D. (2001). Contamination in trials: Is cluster randomisation the answer? *British Medical Journal*, *322*, 355–357.
- Tyler, T. R. (2003). Procedural justice, legitimacy, and the effective rule of law. In M. Tonry (Ed.), *Crime and justice: A review of research* (Vol. 30, pp. 431–505). University of Chicago Press.
- Tyler, T. R. (2006). *Why people obey the law*. Princeton University Press.
- Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities? *Ohio State Journal of Criminology*, *6*, 231–275.
- Tyler, T. R., & Huo, Y. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. Russell Sage Foundation.

- Tyler, T. R., & Wakslak, C. (2004). Profiling and police legitimacy: Procedural justice, attributions of motive, and acceptance of police authority. *Criminology*, *42*, 253–282.
- Tyler, T. R., Fagan, J., & Geller, A. (2014). Street stops and police legitimacy: Teachable moments in young urban men's legal socialization. *Journal of Empirical Legal Studies*, *11*, 751–785.
- US Bureau of Justice Statistics. (2018). *Contacts between the police and the public, 2015*. US Bureau of Justice Statistics.
- US Department of Justice. (2015). *Fact sheet: Office of Justice Programs comprehensive body-worn camera program*. U.S. Department of Justice.
- US National Academies of Sciences, Engineering, & Medicine. (2018). *Proactive policing: Effects on crime and communities*. The National Academies Press.
- Weisburd, D., Wooditch, A., Weisburd, S., & Yang, S.-M. (2016). Do stop, question, and frisk practices deter crime? Evidence at micro units of space and time. *Criminology & Public Policy*, *15*, 31–56.
- Weitzer, R., & Tuch, S. (2005). Racially biased policing: Determinants of citizen perceptions. *Social Forces*, *83*, 1009–1030.
- Weitzer, R. & Tuch, S. (2006). *Race and policing in America: Conflict and reform*. (New York:Cambridge University Press)
- White, M. D. (2014). *Police officer body-worn cameras: Assessing the evidence*. Office of Community Oriented Policing Services.
- White, M. D., Todak, N., & Gaub, J. E. (2017). Assessing citizen perceptions of body-worn cameras after encounters with police. *Policing: An International Journal of Police Strategies & Management*, *40*, 689–703.
- White, M. D., Gaub, J. E., & Todak, N. (2018a). Exploring the potential for body-worn cameras to reduce violence in police–citizen encounters. *Policing: A Journal of Policy and Practice*, *12*, 66–76.
- White, M. D., Todak, N., & Gaub, J. E. (2018b). Examining body-worn camera integration and acceptance among police officers, citizens, and external stakeholders. *Criminology & Public Policy*, *17*, 649–677.
- Worden, R., & McLean, S. (2017). *Mirage of police reform: Procedural justice and police legitimacy*. University of California Press.
- Zimring, F., & Hawkins, G. (1973). *Deterrence: The legal threat in crime control*. University of Chicago Press.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Anthony A. Braga is the Jerry Lee Professor of Criminology and Director of the Crime and Justice Policy Lab at the University of Pennsylvania. He studies programs to enhance fairness and effectiveness in policing, prevent gun violence, and address violent crime hot spots.

John M. MacDonald is Professor of Criminology and Sociology at the University of Pennsylvania. He studies the effects of policies and programs to reduce crime, violence, and racial disparities in criminal justice. He is the co-Editor of the *Journal of Quantitative Criminology*.

Lisa M. Barao is an Assistant Professor in the Department of Criminal Justice at Westfield State University. She is also a Research Fellow in the Crime and Justice Policy Lab at the University of Pennsylvania.