SCIENTIFIC COMMUNICATION

Police Uses of Force in the USA: a Wealth of Theories and a Lack of Evidence

Kyle McLean[1] · Seth W. Stoughton[2] · Geoffrey P. Alpert[2,3]

Abstract

Research Question How adequate is research in the USA for discovering best policies and practices, and best implementation strategies, for reducing loss of life and injury from police use of force.

Data This analysis examines police agency policies on the use of force regulation, evaluations of training initiatives, research on supervision, problem officers, and other dimensions of the issues and possible solutions.

Methods The analysis examines both documented correlates and the strength of causal inference about those correlates of reductions in the use of force.

Findings The analysis concludes that while many promising ideas have been offered, there are few tested strategies that have demonstrated substantial effects in reducing the use of force. There are virtually no successful strategies that have been replicated.

Conclusions The current proliferation of untested programs may divert effort from a coherent and well-funded program of research to test and select effective policies that are urgently needed.

[*Academic editor's note: this scientific communication is a revised version of a 2021 report that was commissioned by the US National Academies of Sciences, Engineering and Medicine as part of its charge to a Committee to review the* "Evidence to Advance Reform in the Global Security and Justice Sectors." *Its focus is on research in the USA, including institutions, laws, and procedures that are unique to the USA. Readers from other countries may find the analysis useful by lateral thinking about the same issues in other national contexts, with different institutions, laws, and procedures.*]

✉ Kyle McLean
  kdmclea@clemson.edu

[1] Clemson University, Clemson, SC, USA

[2] University of South Carolina, Columbia, SC, USA

[3] Griffith University, Brisbane, Australia

Springer

Introduction

The use of force by the police remains a challenging and divisive issue in the USA and around the world. Although relatively rare in police encounters and typically coming in the form of low-level shoving and empty-hand controls (Adams, 1995), these physical uses of governmental authority can have a devastating impact on how the public perceives the police and how citizens respond to officers (Mourtgos & Adams, 2020). Since the murder of George Floyd in May 2020, concern over police use of force has reached a crescendo, spurred on by the blurred information and uninformed opinions that proliferate on social media.

Motivated by entirely understandable concerns that police uses of force occur too frequently, are avoidable, excessive, or disproportionately targeted at specific populations, community members, elected officials, activists, and police leaders have made—and continue to make—a range of claims about when, why, and how much force is used. Unfortunately, many of the descriptive claims are reductive, incomplete, or simply wrong. For example, while many people are familiar with the research findings that demonstrate that African-Americans killed by police are more likely to be unarmed than white Americans killed by the police (Nix et al., 2017), the scope of the problem is often exaggerated. A recent study demonstrated a substantial portion of individuals believe that more than 1000 unarmed black Americans are killed every year by the police (McCaffree & Saide, 2021) when the number of unarmed individuals of any race killed by the police is relatively small ($n = 93$) (Nix et al., 2017). The disconnect between the perceived scope of the issue and the actual scope of the issue has important implications for how optimally to calibrate a public policy response.

Descriptive claims are not proffered in a vacuum; they are provided as the predicates to prescriptive assertions about how best to address the situation: Change agency policy or training! Hire different officers! Make internal oversight more rigorous! Create external oversight mechanisms! Unfortunately, many of the proposed solutions, like the descriptive claims that underlie them, lack supporting evidence.

In large part, this may be because there is substantially less information available about police uses of force than one might expect. Consider that, in less than 3 years since the emergence of the SARS-CoV-2 pandemic, researchers and public health professionals have collected a wealth of data about infections, hospitalizations, deaths, and a number of potential pharmaceutical interventions. In contrast, there is an embarrassing paucity of data about the frequency and severity of officers' use-of-force decisions. Even looking only at lethal force events, the data reported by official sources—including the Federal Bureau of Investigation and the Centers for Disease Control and Prevention—lag far behind the actual count (Loftin et al., 2017). The data on nonlethal incidents are more dismal yet. Only one state, New Jersey, has anything approaching standardized, robust data on (statewide) uses of force, and they exist only as a result of a prolonged effort by a private media organization to obtain and code all uses of force in the state over a 4-year period (NJ Advance Media, 2018).

Addressing appropriately important crises—whether the COVID-19 pandemic or the use of excessive force by police—requires scientific evidence that is sufficiently robust, granular, and representative to allow for a complete understanding of the issues and the development of clear communication about successful responses. The bases for public policymaking on technical matters must withstand scientific and social scrutiny and stand above political whims and ideological beliefs.

Of course, the difference between the COVID crisis, on the one hand, and policing, on the other, is painfully simple: billions of dollars. Billions of dollars have been spent to develop the scientific evidence necessary to inform the COVID policy recommendations relating to masking, physical distancing, and vaccinations. To date, no similar investment has been made to understand or resolve the issue of excessive force. Operation Warp Speed—the effort to quickly develop a COVID vaccine—had a reported budget of \$12.4 billion (Barone, 2020), while the entirety of the research and evaluation budget for the Office of Justice Programs in 2020 was significantly less than 1% of that figure at only \$79 million (Office of Justice Programs, 2021).

Despite the regrettable lack of investment, social scientists across a range of disciplines have dedicated their careers to understanding and reducing the use of excessive force by the police. Scholars in fields ranging from sociology (e.g., Legewie, 2016) to economics (e.g., Fryer, 2019) to political science (e.g., Mullinix et al., 2021) and, of course, criminology (e.g., Nix et al., 2017), have attempted to understand the scope, causes, and possible solutions to the issue of excessive force. Despite this considerable attention, only recently has the use of force been studied with rigorous research methods, and only recently has that research started to have a systemic impact on public policy.[1]

This modern effort to develop sound evidence about policing has been linked to a growing public concern, as well as an increasing acceptance of evidence-based methods and practices from police command staff, policymakers, and trainers. Research findings have formed the basis for many decisions, which were previously left to dubious theories, unsupported assumptions, and political whims. Unfortunately, the evidentiary foundation is both thin and full of gaps, incapable of meeting the needs and requests of either the public or police practitioners. If we truly consider the issue of police use of force as being as urgent and pressing as the COVID pandemic, then we must dedicate greater resources to the scientific study of policing as a distinct set of social practices rather than as a topic suited principally for philosophical discussion and theoretical musings.

This is not to say that policies should not be enacted, and practices should not be undertaken in the absence of a body of indisputable scientific evidence. That would be unrealistic. It is to say, however, that public policy should be predicated on

---

[1] For our purposes, excessive force refers to force that exceeds what the governing legal standards and/or what the public deems reasonable in a given society. When we refer to the study of the use of force, this includes excessive force, as well as reasonable uses of force. Academic studies in this area tend to focus on the use of force rather than excessive force as excessive force requires a determination of unreasonableness that is unavailable in most datasets. Accordingly, we focus on the study of the use of force with the intent of reducing the problem of excessive force.

robust evidence when available, on reliably developed findings when possible, and on sound theoretical assumptions when needed. To the detriment of police agencies, officers, and communities, it is far too common to see use-of-force policymaking based on nothing more than the historical practices that are grandiosely titled "conventional wisdom" (or, put another way, "how things have always been done"). Even recent reforms reflect good ideas, but few empirical studies of their effectiveness. Sound public policy demands more than optimism and untested assumptions.

Understanding, explaining, and regulating police uses of force and excessive force requires a multifaceted approach that incorporates a thorough understanding of policies, training, supervision, accountability, and their interactions. In this article, we will review the research findings in each of the identified areas to show the evidence, or lack thereof, that drives our knowledge and decisions concerning the uses of force. In each of these substantive areas, scholars and policymakers have generated a lot of good ideas, tested only a few of them, and replicated an even smaller number of them. This leaves us with an abundance of testable ideas, but sizeable gaps in our knowledge of their efficacy.

We can do better. We must do better. Our knowledge base must be broadened and deepened through rigorous evaluation so that the public, elected officials, and police leaders can have faith in research findings and use them to develop rational and reasonable responses to this critical issue.

Policies

Agency policies that regulate the use of force represent perhaps the clearest area in which decisions are dominated by the conventional wisdom about what seems like a good idea instead of data about what works. This is dictated, in no small part, by the need for agencies to adopt defensible policies that comply with state and federal law (Stoughton et al., 2020), reflect existing jurisprudence on the use of force (Stoughton, 2021), and garner both officer and public acceptance. Unfortunately, jurisprudence and public sentiment do not always make for an effective means of properly regulating the use of force.

What has come to be called "best practices" for use-of-force policies are guided by a small number of key organizations, each of which has released so-called "model policies." The International Association of Chiefs of Police (IACP) has released its National Consensus Policy, the Police Executive Research Foundation (PERF) has released its Guiding Principles on the Use of Force, the College of Policing in the UK has a list of general principles and three core questions for every officer to evaluate during a potential use of force encounter, and the Australia and New Zealand Policing Advisory Agency (ANZPAA) has a list of recommended use of force principles. These documents overlap in several areas, such as emphasizing that officers use force only when it is reasonable, necessary, and proportionate. They each make a number of distinct recommendations that include incorporating a statement on the sanctity of life in agency policy, adopting a range of components of "de-escalation," articulating that officers have a duty to intervene, and, in the case of PERF's Guiding Principles, advocating for agencies to provide all patrol officers with riot shields.

Regardless of their content, each of these documents aims to build upon clear legal, moral, and philosophical foundations. Notably absent from that foundation, however, is empirical evidence of effectiveness.

Empirical examinations of use-of-force policies are few and far between. Early examinations of force policies used the use of force continuum to create a force factor score for examining the prevalence of potentially problematic use of force incidents (Terrill, 2005; Alpert & Dunham, 1997, 2004). This work, while informative, provides little information on whether the use of that continuum—or certain characteristics of the various continua—reduced the uses of force. Similarly, Terrill and Paoline (2013) conducted a national survey of police departments to assess the prevalence of the use of force continua in the early 2010s. Their study demonstrated that while use of force continua are adopted in most police departments, there is substantial variation in the structure and specification of the continua. Again, however, there was no reliable indication about the substantive impact, if any, of these continua on officers' use-of-force decision-making.

A small number of studies have determined that more restrictive force policies lead to reductions in the use of force, likely beginning with work by Fyfe (1979). White (2001) used interrupted time series analyses to examine the impact of a series of policy changes on officer-involved shootings in the Philadelphia Police Department, concluding that restrictive policies can reduce the number of shootings. Similarly, Terrill and Paoline (2017) studied less-lethal force policies in eight departments across the country, concluding that more restrictive policies were associated with fewer use of force incidents.

Still, analyses in this area struggle to identify specific causes. Methodologically, studies of agency policies are plagued by endogeneity problems. For legal reasons, different policies cannot be randomly assigned to individual officers. Policies also tend to be revised substantially after major incidents or leadership changes, introducing variables beyond the policy revision itself (such as public outcry). Furthermore, policy changes are often wholesale, with the rewriting of an entire policy or set of policies, rather than with isolated changes to specific policy components.

Policy revisions—of whatever variety—cannot easily be disentangled from the modifications to training or supervisory expectations that accompany those revisions. Accordingly, analyses that examine alterations to use of force policies cannot cleanly determine whether any impact results from not only the policy change, but a host of additional factors, and if so what specific change drives which impact. Similarly, analyses that compare the use of force policies across multiple jurisdictions must contend with the issue that more restrictive force policies may reflect different agency cultures or that training within these departments that may be contributing to any detected effects. To be clear, these observations are not a criticism of the scholars who have worked in this area but rather a statement on how difficult it is to design and conduct this type of research.

In fact, the only recent peer-reviewed study that identified a specific change in use of force policy and measured the impact on reported uses of force in our review was conducted by Shjarback and colleagues (Shjarback et al., 2021). These researchers assessed the impact of a change in policy at the Dallas Police Department that required officers

to report when they pointed a firearm at a subject, finding that the policy modification resulted in a slow but lasting reduction in officer-involved shootings.

Finally, it should be noted that several journalistic studies of the use of force (e.g., the Washington Post's Fatal Force Project, Campaign Zero) have provided some recommendations on the contents of use of force policies based on data. These studies are not peer-reviewed and come with many of the same caveats noted above regarding separating the independent impact of individual policies. Overall, their recommendations are largely consistent with the findings that more restrictive policies result in fewer uses of force. Still, it is striking that efforts by advocacy organizations and the media have created more evidence-based suggestions than governmental agencies or many of the professional organizations that spend significant effort promulgating and promoting their model force policies for nationwide adoption.

Training

Compared to agency policies, police training has been the focus of considerably more research. Over the past decade, in-service training programs have been studied for their potential to reduce the use of force. One study examined the impact of a procedural justice training program in the Chicago Police Department on trainees' attitudes compared to a control group of untrained officers (Skogan et al., 2015). This study was one of the few evaluations of police training and advanced the state of the literature, but it was disadvantaged by the lack of experimental design and its failure to specifically measure the use of force. Procedural justice—the outcome of interest in the study—has been linked to citizens' compliance (Tyler, 2006; Tyler & Huo, 2002), offering a theoretical connection between procedural justice and the use of force, but the study did not look at the uses of force as a distinct outcome.

Giacomantonio and colleagues (Giacomantonio et al., 2020) later pushed this type of evaluation further by examining the impact of a verbal judo training program in the Halifax Regional Police (CAN) on officers' performance in simulated scenarios. On the one hand, their study used advanced technology—an immersive virtual reality simulator—to expose treatment and control officers to the same stimuli to test for differences in behavior. On the other hand, the study lacked an experimental design and showed no differences in the use of force between trained and untrained officers. Additionally, there is no robust evidence as to the relationship between changes in performance in simulated scenarios, especially using virtual reality simulators, and changes to performance in the field.

Following these initial studies, a number of experimental evaluations have been conducted to test for differences in administrative reports of uses of force. These include an evaluation of Polis Solutions' T3 training program by McLean and colleagues (McLean et al., 2020), an evaluation of PERF's ICAT program by Engel and colleagues (Engel et al., 2022), an evaluation of procedural justice training by Wood and colleagues (Wood et al., 2020), and an evaluation of de-escalation training by White and colleagues (White et al., 2021a, b).

McLean and colleagues' study randomized individual officers to treatment and control conditions and was strengthened by an experimental design and repetitive training exposure but was limited by potential contamination issues and agency measures of use of force (see McLean et al., 2020 for a discussion). The results showed no significant differences in the number of force incidents involving trained officers compared to the number involving non-trained officers.

Engel and colleagues' use of a stepped-wedge design limited contamination issues but left unaddressed the possibility that the observed reductions could be attributed to exogenous changes that coincided with the introduction of training, such as distinct policy changes, the occurrence of critical incidents in the jurisdiction, or de-policing (see also Worden and colleagues (Worden et al., 2020), for a stepped-wedge study impacted by external factors).

Wood and colleagues studied a procedural justice training program in the Chicago Police Department using a stepped-wedge design. The study had similar issues with the potential for external factors influencing the use of force (see Correction to Wood et al., 2020) and found a small reduction in the use of force in the Chicago Police Department.

White and colleagues (White et al., 2021a, b) recently published their evaluation of a de-escalation training program in the Tempe Police Department (TPD) that found no impact on officers' uses of force. However, as the authors note:

> [A]dministrative use of force and injury data are blunt measures that might not capture the impact of the training. In simple terms, TPD is not a high use of force department and injuries are rare. Since use of force is a rare event, the evaluation needed to take a more nuanced approach to fully capture the impact of de-escalation training. (White et al., 2021a, b:60).

These studies represent the first rigorous reviews of training programs designed to reduce the frequency and severity of uses of force. But several commonalities that dominate the studies create serious issues for providing scientific recommendations about police training. First, each study evaluated a different training program, and no known studies attempt to replicate existing evaluations of police training. Accordingly, the evidence base for any training program is limited to a single study. Even in the most promising of studies—perhaps especially with the most promising of studies—replication is important to improve confidence in the generalizability of the findings.

Second, each study evaluated an in-service training program. In-service training programs are undoubtedly important, but the focus on in-service training in experimental evaluations has highlighted the complete lack of evidence on what programs and techniques are effective in preservice (academy) training of new officers.

Third, each study (except the one that evaluated Polis Solutions' T3 program) evaluated a one-time training program provided to officers over 1 or 2 days. Best practices for adult learning in training from other fields suggest that high-repetition training over more prolonged periods is important for both retention and understanding (Ericsson, 2004; Ericsson & Ward, 2007).

To be clear, a number of promising training programs exist that may reduce officers' uses of force or use of excessive force. Programs like Georgetown University's

Active Bystandership for Law Enforcement promote peer intervention when officers perceive a colleague to be using or about to use excessive force (Active Bystandership for Law Enforcement Project, 2021). The Marietta, GA, Police Department has adopted a Brazilian Jiu-Jitsu program that they claim will increase officers' confidence in their physical abilities, making officers both more willing to communicate before employing physical control techniques and more willing to use empty-hand controls rather than resorting to intermediate weapons such as TASERS (Huddleston, 2021). Additionally, the LaGrange, GA, Police Department has developed "Shoot to Incapacitate," a controversial program that trains officers to shoot at subjects' pelvic girdles or upper legs in certain deadly force situations to reduce lethality vis-à-vis center mass (Schrade, 2021). All three programs reflect innovative ideas and are based on sound theoretical assumptions, but none of the three is backed by rigorous scientific research into their effectiveness in achieving their stated goals.

In sum, our review of studies measuring the impact of police training on the use of force is consistent with our review in each of the other areas. Promising ideas abound, but empirical research continues to lag behind the growing desire to implement evidence-based training programs. Dedicating more money, time, and resources to this critical topic is key to reducing the quantity and improving the legitimacy of police use of force incidents.

Internal Police Oversight

Police oversight focuses on the internal and external mechanisms that police departments use to ensure that officers follow both governing legal rules and their own agency's administrative policies and training. Internal oversight mechanisms are a critical component of regulation; even the best policies and training programs will be undermined if officers do not believe that the department takes them seriously enough to ensure compliance. In 1981, the U.S. Commission on Civil Rights examined the issue of police oversight in *Who's Guarding the Guardians*, a report that is a "must read" for students of police reform and still guides many of the topics that we will review in this section. Discussions of accountability and oversight in early reports on reform highlight a division of internal and external control mechanisms, which we will echo here (see Noble & Alpert, 2009 and Kane & White, 2009). However, both the 1981 report and previous reports on police misconduct and associated issues (including the 1973 National Advisory Commission: Task Force on Police, see Walker, 2012) suffer from a lack of meaningful data and evidence-based recommendations. Here, too, we see the theme: "good ideas but no data."

Supervision

Internal oversight represents efforts undertaken within a police department to ensure that its officers are following its policies, training, and relevant legal precedents. Supervision represents one of the more promising areas of research for police oversight when considering the weight of evidence about its impact on the uses of force.

Findings that effective supervision affects an officer's use of force are hardly surprising, given the role that bosses have in encouraging or even requiring particular types of behavior on their shifts. It is often stated that the success of any system of oversight starts with the frontline supervisors and their leadership abilities (see Perez, 1994). At the same time, the practical implementation of these findings is unclear, given the limited research on how to identify, promote, and train good supervisors. Practically speaking, we might know that good supervision matters without knowing how to promote it.

Early efforts at police reform focused heavily on the role that supervision played in improving policing. The 1981 *Who's Guarding the Guardians* report noted—without explaining how to fix it—that the lack of supervision allowed officers to act outside of training and policies, one consequence of which is the excessive use of force. In the late 1980s, the International Association of Chiefs of Police (1989: 57) published a study on building integrity and concluded that "Supervisors who fail to audit the activities of personnel fail to perform their jobs." Their point was that first-line supervisors were responsible for knowing their officers' strengths, weaknesses, and styles of working with the community and for socializing and training officers in their proper roles to reduce any inclination or need to use force unnecessarily. Although this necessary aspect of supervision has been recognized for decades, it remains surprising that very little research has been conducted on the impact of supervisory styles and emphases.

Early studies on the role of supervision in police use of force used surveys to correlate officer attitudes and behaviors to supervisor's attitudes and behaviors (Engel, 2000, 2001, 2002; Engel & Worden, 2003). Using these techniques, Engel (2000) classified officers with "active" supervisors as (1) having increased activity in the field; (2) those who take over situations; (3) acting with the dual role of both a street officer and a supervisor; and (4) more likely to use force than officers with less active supervisors. Unsurprisingly, modeling "active" behavior encourages street officers to be more active and engage in more uses of force.

A more recent study examined the impact of procedurally fair supervision on officers' use of force (Owens et al., 2018). In this study, officers identified as working in high-risk circumstances were randomly assigned to either a treatment or control condition. In the treatment condition, officers had a non-disciplinary meeting with their supervisor in which the supervisor modeled procedurally fair behavior, reviewed a recent incident the officer engaged in, and pointed out areas of success and potential areas for improvement. This cognitive debrief was juxtaposed to the control group who were debriefed in a "matter-of-fact" manner, without any emphasis on procedural justice. As a result of the intervention, treated officers were less likely than control officers to engage in uses of force in the 6 weeks following their supervisory meeting. In fact, the officers in the experimental group passed forward the procedurally just treatment they learned from their supervisors to the individuals with whom they subsequently interacted (Owens et al., 2017).

This study is particularly promising. It not only reinforces the hypothesis that supervisors have a meaningful impact on officers' uses of force: it also identifies a specific intervention that can promote this type of supervision. This is promising, but it suffers from many of the same caveats as the studies of police training: the

intervention only resulted in a demonstrated reduction in uses of force for 6 weeks; it was studied in only a single police department and has never been subject to replication. There is no evidence on whether the effects of supervisory meetings are purely transient, or on the optimal number, frequency, and duration of such supervisory meetings, or on the opportunity costs of any long-term, repetitive intervention such as the redistribution of supervisor's time away from other supervisory responsibilities.

Body-Worn Cameras

Another potentially fruitful area of research on supervision comes from the body-worn camera literature (Adams et al., 2021; see Lum et al., 2019 for a review of body-worn camera research). Arguments for body-worn cameras typically claim that cameras can increase supervision because supervisors can review footage of police officers performing daily duties (Stoughton, 2018). However, the reality of the use of body-worn cameras for internal supervision is likely much weaker.

Koen and Mathna (2019) found that, in a department with a specific policy that supervisors should review body-worn camera footage proactively to look for potential issues with officers, proactive reviews were rarely (if ever) actually conducted. The management of the body-worn camera program required a considerable investment of time and effort to curate footage for cases that were either proceeding through the judicial system or involved citizen complaints. Very little time was left over for reviewing footage that had not been specifically requested by someone else. This practice substantially weakens the supervision effects of body-worn camera policy as the footage is only reviewed when the department is already aware that something has happened. This effectively uses body-worn camera footage as a source of information about known incidents rather than as a mechanism for random reviews that could uncover otherwise unknown issues, such as low-level uses of force omitted from an officer's written report.

Despite these unanswered questions, the clear message from this literature is that supervision generally is a promising avenue for reducing the use of force and merits dedicating considerably more resources to evaluating and replicating promising programs.

Focusing on Problem Officers

It is frequently said in policing that 10% of the population causes 90% of the problems that officers are called upon to address (see United States Commission on Civil Rights (1981). Although the specific numbers may prove imprecise, the concept that a relatively small percentage of people are responsible for a disproportionate amount of a particular behavior is a familiar one. This concept certainly applies in policing, where police leaders and academics have for years been concerned about the small group of officers who are responsible for a large percentage of force incidents or misconduct (bad apples). Data from New Jersey, for example, shows that 10% of

officers in the state accounted for 38% of all uses of force, and a small group of just 252 officers used force at a rate five times higher than the state average (NJ Advance Media, 2018).

One internal accountability tool noted in *Who's Guarding the Guardians* is the Early Warning System, later renamed the Early Identification System (EIS). An EIS uses internally-generated data on various aspects of officers' performance—ranging from uses of force to use of sick leave—to alert supervisors to officers who are potentially problematic before they become a serious concern. While there is evidence that EIS can be effective if managed properly with moderate goals (Walker et al., 2001), a recent systematic review of EIS shows that there were only eight studies that matched the selection criteria for proper methods, six of which were single agency studies (Gullion & King, 2020). The results of the reviewed studies are encouraging, but measures of success/effectiveness are inconsistent. Here, as elsewhere in the police research literature, data drawn from single-agency studies provide only weak results.

Chalfin and Kaplan (2021) recently published a study examining the potential outcomes if EIS were properly implemented including the use of resources to identify and extricate problem officers. Concluding that removing "high-complaint 'bad apples'" from police departments would result in only small reductions in civilian complaints, Chalfin and Kaplan contend that agencies should focus their resources on broader reforms rather than committing substantial resources to EIS. Sierra-Arévalo and Papachristos (2021), in turn, challenged the analytical approach that Chalfin and Kaplan relied upon, concluding that the Chalfin and Kaplan estimates are comparable to the modest results seen in other programs. Specifically, Sierra-Arévalo and Papachristos make an argument similar to ours: the existing literature is all but devoid of empirically based recommendations that result in substantial reductions in the use of force or citizen complaints. Rather than committing resources to policies and programs with indeterminate outcomes, Sierra-Arévalo and Papachristos suggest, a modest (4–6%) but proven reduction may be preferable to unsupported optimism.

This argument is worth discussing at greater length because it reflects the conundrum replete in our findings. A policymaker or police chief who wishes to commit resources to reduce the uses of force in their agency and wants to make an evidence-based policy determination must choose among reforms that are all associated with modest reductions. These may include the development of EIS or the implementation of the procedural justice training program evaluated by Wood and colleagues (Wood et al., 2020), or programs that are theoretically sound, and which may offer a more robust effect but have yet to be evaluated, such as Active Bystandership for Law Enforcement. Even assuming the policymaker is scientifically literate and well informed, their position in making such choices is not an enviable one (see Fleming & Rhodes, 2018 and Aronie & Alpert, 2020).

In the event that a policymaker agrees with Sierra-Arévalo and Papachristos' (2021) that these "rotten apples" should be terminated, that decision may be easier to make than to implement. Investigating and terminating an officer can be difficult and time consuming, not least because of the substantive and procedural protections provided by collective bargaining agreements (Rushin, 2017, 2019) or the state Law

Enforcement Officers Bill of Rights. Additionally, departmental discipline is often overturned in arbitration (Rushin, 2021) or by civil service boards (see Grunwald & Rappaport, 2020).

Even assuming that termination is successful, doing so may provide no more than a local solution that displaces problem officers to other agencies in rather the same way that certain crime-fighting strategies may result in the non-reductive displacement of criminal behavior (Johnson et al., 2014). Robust decertification programs, which strip someone of their license to be a police officer in the state, are sparse (Goldman, 2013, 2016). Many fired officers find work with other agencies, especially during difficult recruiting periods (Grunwald & Rappaport, 2020). Hiring an experienced officer provides some cost savings to agencies, and an agency administrator must balance the immediate reality of putting an officer on the street without the lag time of training against the speculative possibility that the officer may engage in future misconduct.

That said, hiring an officer who resigned in lieu of termination or was fired or asked to resign can create significant liability. A study of such "wandering officers" in Florida found that they were significantly more likely to receive complaints of violent or sexual conduct even when controlling for age, gender, and education (Grunwald & Rappaport, 2020). Plaintiffs' attorneys highly value cases involving an officer who repeated the bad behavior that caused them to leave a prior agency; indeed, such situations can open the door to municipal liability claims that are otherwise difficult to establish because they require evidence that the agency adopted an unconstitutional policy or custom. While the total cost of lawsuits against the police is unknown, news reports and academic assessments suggest they are quite significant (see Schwartz, 2014). A 2021 report by the Marshall Project shows that New York and Chicago have paid out more than 2 billion dollars in police misconduct claims over the past 10 years (Thomson-Devaux et al., 2021).

External Police Oversight

Like internal oversight, external oversight seeks to ensure that officers follow governing legal rules, agency policies and training, and community expectations. Unlike internal oversight, external oversight goes outside of the traditional command and control mechanisms that have evolved within police agencies by involving agency outsiders in the form of civilian review boards, consent decrees, or individual litigation.

Civilian Review Boards

Civilian Review Boards are often touted as one of the essential tools for police reform. In the late 1990s, the National Institute of Justice funded a project to examine civilian oversight of the police (Finn, 2001). The resulting report explained in detail the structures and functions of nine different systems of oversight and

described the strengths and weaknesses of each. The nine systems are then aggregated into four types of programs (Finn, 2001, vii), including.

1) Citizens investigate allegations of police misconduct and recommend findings to the chief or sheriff;
2) Police officers investigate allegations and develop findings; citizens review them and recommend that the chief or sheriff approve or reject the findings;
3) Complainants may appeal findings established by the police or sheriff's department to citizens, who review them and then recommend their own findings to the chief or sheriff;
4) An auditor investigates the process by which the police or sheriff's department accepts and investigates complaints and reports on the thoroughness and fairness of the process to the department and the public.

Importantly, we should not divorce civilian oversight from other accountability mechanisms, including internal accountability. One of the major conclusions of the 2001 report was the need for a comprehensive system of accountability, as "citizen oversight alone cannot ensure that police will act responsibly" (Finn, 2001, vii). While the report noted that all oversight efforts should be evaluated (Chapter 7) and included a summary from an assessment of the Albuquerque oversight program (Luna & Walker, 1997), there was no methodological roadmap for a systematic review of oversight programs.

In 2016, De Angelis et al. (2016) surveyed almost 100 police oversight executives and reported on the key issues associated with the implementation and sustainability of civilian oversight of police. As in prior reports, they looked at trends in the implementation of these oversight bodies and discussed objectives and goals but reported no systematic data or evaluations of the various programs.

In a more recent review, Stephens et al. (2018) reviewed the results of a roundtable discussion and a survey concerning the civilian oversight programs in the major city police departments. The report (2018, p. 20) identifies and reviews multiple objectives of civilian oversight, including.

- Transparency;
- Independent investigations;
- Improving accountability;
- Improving public trust and legitimacy;
- Engaging the community; and
- Demystifying police internal affairs investigations.

While the Stephens et al. (2018) report updates Finn's 2001 review regarding the structures and functions of these boards, it concludes (p. 20) that "There is little, if any, empirical research on the effectiveness of civilian oversight of the police—nor are the programs subjected to any systematic evaluation."

Since these reviews have been published, the National Association for Civilian Oversight of Law Enforcement has provided an impressive array of civilian

oversight models (https://www.nacole.org). It does not, however, offer any empirical evidence demonstrating the effectiveness of civilian oversight. Similarly, when the Council on Criminal Justice: Task Force on Policing published a Policy Assessment on Civilian Oversight in April 2021 (CCJ Task Force on Policing, 2021), it reviewed the variety of models and the potential merits of civilian oversight and concluded (p. 1, 3):

> Rigorous empirical research on the impact of civilian oversight and the relative merits of different models does not exist. Other research has yielded mixed findings about the ability of civilian oversight to reduce excessive use of force and other forms of police misconduct. … Research to date has been largely descriptive of existing civilian oversight entities, either taking a deep look at the operations of a single entity or comparing and contrasting across the functions and structures of several entities.

Specifically, the report reviews the most influential studies and notes many methodological problems, including the lack of control over policy changes during the research period. The result is a characteristically mixed message: civilian oversight has promising aspects, it is helpful as a matter of public relations and transparency, but there is no empirical evidence that it works to reduce police misconduct or the use of force. We know that civilian oversight of the police is controversial, that there is not an accepted methodology or set of criteria to measure effectiveness; and that, as a result, there have not been any systematic studies of the effectiveness of any of the various models of civilian review. Once again, we are left with a lot of good ideas and theories but precious little data.

Consent Decrees

One approach to police use of force that is unique to the US stems from post-Civil War legislation that makes deprivation of civil rights under color of law a federal crime, as well as a tort. More recent legislation empowers the US Department of Justice to put state and local police agencies into a de facto state of "receivership" (bankruptcy) in which a court can effectively take over a police agency. This approach can take several specific forms, including Consent Decrees, Settlement Agreements, and Collaborative Reforms. These all offer another model of externally generated reform. Each category of these external reforms differs from the others in terms of degree and oversight requirements.

While consent decrees have been around for a few decades, they and their data have been shrouded in secrecy until recently (Jiao, 2021 and Alpert et al., 2017). Unfortunately, the secrecy under which the consent decrees have operated has kept valuable evaluation data out of the hands of researchers or restricted the publication of studies. Monitors' reports and agency data may be public, but the research behind them is unlikely to be made available. While there are a number of issues to consider, such as federal versus state consent decrees and lawsuits, the lack of access to these data leaves research with lost opportunities.

The research conducted in several agencies under consent decrees could assist reform efforts across the country. On the one hand, a comprehensive evaluation of the paragraphs of the "Findings Letters" and/or decrees would provide a real-time roadmap of where agencies are failing and where they need assistance. These data are available and could be used by other agencies as a framework for reviewing their own policies, training, supervision, and accountability systems in light of what they know about why a Finding Letter was issued. The data collected by the agencies subject to these external controls could also be analyzed. They could, for example, demonstrate the agencies' capacity for improvement and ability to fulfill the requirements of the consent decree or other reform agreements with the government. These data could be analyzed to develop a roadmap or toolkit for other agencies to follow. This may prove particularly fruitful, given that the process of satisfying a Consent Decree requires input and oversight from agency command staff, external monitors, and a battery of government attorneys. In this way, both the failures and successes of one agency could be leveraged by others. For the most part, though, these data have not been made readily available or used to assist other agencies.

As for the impact or effectiveness of consent decrees, it is difficult to determine whether that impact justifies the significant amount of time, money, and effort that they inevitably require. The relative merits of a consent decree are difficult to assess with any precision. Several of the best descriptive studies of consent decrees are by Goh (2020), Rushin (2015), Chanin (2015, 2017), and Harmon (2009). There is some evidence that consent decrees and other forms of structural reform litigation have been effective in reducing misconduct (Rushin, 2015), but we do not know if consent decrees were an efficient or cost-effective mechanism for reform. Chillar (2021: 7) conducted an extensive and thorough review of the evaluation research conducted on consent decrees and published an important article that concluded as follows:

> While some consent decree evaluations find compelling evidence regarding the outcomes used to track compliance and the positive impacts police and citizens experienced, it is important to note that the studies were conducted during the period in which the department was under consent decree or was recently released from federal oversight. This ensured that the departments had no time to revert to the problematic policies and practices that warranted the implementation of the consent decree. Following the exit of the monitoring team in the cities of Pittsburgh, Washington D.C., Cincinnati, Detroit, and Prince George's County in Maryland, Chanin (2014, 2015) used focus groups and administrative data to investigate the effect of consent decrees. While results indicate that consent decrees offered departments the opportunity to bring about organizational change, such change faltered and reverted to business as usual upon the absence of DOJ oversight.

Chillar's observation reflects an issue raised in several of the articles; whether the reforms that are measured in the short-term term persist over the long term. His own research examined the short-term impacts on police stops of community members by the Newark, NJ Police Department. Overall, Chillar found that the consent decree reforms likely explained improvements in data quality collected in field

interrogations but that there were unexplained decreases in reported field interrogations of blacks and Latinos. Significantly, he concluded that regardless of specific reforms and oversight efforts, cultural change in the agency was necessary to improve police behavior and actions toward black and Latino community members.

The best available evidence shows that consent decrees tend to improve departments in the short term, but that, in the absence of ongoing oversight, agencies will revert to many of the problematic behaviors that resulted in the consent decree in the first place.

Qualified Immunity

Perhaps, no aspect of the legal regulation of policing has been subject to more contentious public debate than qualified immunity: a legal doctrine that protects officers who violate the Constitution from civil liability unless it was "clearly established" at the time that their action was a constitutional violation. In other words, so long as a "reasonable officer" could have thought that a particular action did not violate the Constitution, an officer will not be liable even if they did violate the Constitution. And the unconstitutionality of an officer's actions is only "clearly established" if it is particularly egregious or if a court in the jurisdiction has decided a factually similar case. As the leading qualified immunity scholar described it, "Current Supreme Court doctrine suggests that an officer violates clearly established law only if there is a prior court of appeals or Supreme Court decision holding virtually identical facts to be unconstitutional" (Schwartz, 2021). Under this standard, qualified immunity has become a significant bar to vindicating civil rights claims.

In no small part, this is because the constitutional standards that regulate policing—such as Fourth Amendment standards of proof like "probable cause" and "reasonable suspicion"—are themselves often indistinct (Stoughton et al., 2021). The result is a series of questions about the substantive standard (i.e., whether an officer violated the constitution) and the applicability of qualified immunity (i.e., whether the violation was clearly established) that are both framed in ways that often favor and defer to officers.

In part because of these ambiguities and the barriers to accountability that they create, there are arguments both to maintain and abolish qualified immunity. On the one hand, it is argued that the doctrine helps weed out weak cases, develop acceptable practices, develop the law, and protect the purses of municipalities. On the other hand, as Professor Joanna Schwartz has noted (2017: 1):

> The United States Supreme Court appears to be on a mission to curb civil rights lawsuits against law enforcement officers, and appears to believe qualified immunity is the means of achieving its goal. The Supreme Court has long described qualified immunity doctrine as robust—protecting 'all but the plainly incompetent or those who knowingly violate the law.' (citing Mallory v Briggs, 475 U.S. 335, 341 (1986)).

Her work, along with that of Kinports (2016), explores the arguments and judicial opinions that have used qualified immunity as a shield for the police, protecting them from abusing or mistreating community members. Schwartz presents an analysis of the role of qualified immunity in constitutional litigation that leads her to conclude that qualified immunity has been seen as both unclear and unreasonably protective of government actors.

The evidence on qualified immunity suggests that the law, as it now stands, not only shields officers from liability for abuses, it also shields their employers. Private employers are liable for the actions of their employees under a legal theory known as "vicarious liability" or "respondeat superior," but public employers, including police agencies, are not. Nevertheless, police agencies virtually always cover the costs of defending their officers in litigation as well as the settlement or judgment costs, known in law as indemnification (Schwartz, 2014). Thus, when individual officers are insulated from liability, it is almost always employers, rather than the officers themselves, that benefit.

What is far less clear is whether eliminating qualified immunity—or revising municipal liability doctrines or tort reform, for that matter—would have any significant effect on officer action. Schwartz has predicted that there would be more civil rights litigation if qualified immunity were abolished, but that the average time, cost, and complexity per case would decrease and that the success rates for plaintiffs and defendants would remain roughly stable (Schwartz, 2020). Here, too, we have a sound theory but little empirical information. Schwartz's predictions are well informed, based on doctrinal analysis and interviews with attorneys, but there is no supporting analytic model. As the 2021 Marshall Project concluded its report finding that police misconduct cost New York and Chicago more than 2 billion dollars over 10 years, "If not for inconsistent and shoddy record-keeping, we might know if settlements make a difference in police misconduct" (Thomson-Devaux et al., 2021).

Discussion

The current state of the literature on police use of force can best be summarized as a wealth of ideas with a lack of support. Our review suggests that quality studies of policies, training, selection, supervision, and accountability exist in small numbers, with few of the many policies and programs implemented by police agencies having been evaluated properly. Existing studies of policy changes, training, selection or supervision programs, or accountability measures never contained more than a few agencies and were frequently limited to single departments.

Moreover, replications of existing studies are virtually nonexistent. Creating a robust evidence base requires replication. Yet replication is generally rare across social sciences. One analysis, for example, examined articles in 100 psychological journals with the highest impact scores, (Makel et al., 2012), finding that only 1% of published studies are replications.

There are a number of reasons for this replication crisis. First, journal editors who place a higher price on originality and positive results are reluctant to publish replications. In fact, a study of journal editors by Madden et al. (1995, p. 85) concluded:

A paradox of replication in the social sciences is that a researcher who operationally replicates and finds nonsupport for previous work may be accused of not being true to the original method, but if the researcher finds support for the previous work, then the argument will be made that nothing new has been learned.

The difficulty of getting replications published reduces the desirability of conducting them, although it is worth noting that the more controversial the topic the more likely a replication study will be published (Madden et al., 1995).

The second major reason for the lack of replication studies is the difficulty of obtaining funding for major applied research projects in policing. As noted in the introduction, the research and evaluation budget for the Office of Justice Programs—the major national agency that is responsible for funding these types of evaluations—is relatively small, especially considering that it covers not only policing but other areas of criminal justice (e.g., courts, corrections, and forensic sciences). The annual Research and Evaluation on Policing solicitation funds evaluations of policing programs but usually fewer than 5 per year, as in 2021. In a competitive environment where only a select number of evaluations of policing programs are funded every year, it is nearly impossible to obtain funding to replicate an evaluation of a program in a different context. Even www.CrimeSolutions.gov, a program sponsored by the National Institute of Justice (NIJ) that rates research to determine the certainty of the outcomes and effectiveness, does not use replications as an evaluation criterion. The National Institute for Justice (or another federal agency) should create a solicitation dedicated to replications of studies that show promise by prior evidence, as already exists in the private sector (Arnold Ventures).[2]

In the end, the strongest recommendation we can provide is that more resources—in the form of time, money, effort, and commitment from police researchers and agencies alike—must be dedicated to studying police behavior and uses of force. More detailed policies, different equipment, better communication skills, the extended use of mental health professionals, improved supervision, changes to accountability strategies, and other reforms could all work—or, more likely, could all marginally advance the ultimate goal as components of comprehensive reform—but it will take time and diligent research to determine how to best make them work and whether each particular effort is individually or collectively worth the time, effort and cost.

Generating an evidence base for dealing with the use of force in policing is exceedingly difficult. Reducing unnecessary and excessive force is an objective of vital importance to the public. National attention to this issue has generated the immediate need for departments to "do something." Unfortunately, from a scientific perspective, immediate action is often not supported by or conducive to generating solid research findings. Just as responsible public health officials did not immediately roll out untested SARS-CoV-2 treatments nationwide, we should resist the

---

[2] https://craftmediabucket.s3.amazonaws.com/uploads/RFP-for-RCTs-in-Criminal-Justice-CJ-EBP.pdf

urge to rush out untested programs to reduce the use of force nationwide despite the urgent need for improvement.

Indeed, the importance of the issue is precisely why we must take a deliberative approach. Getting it right slowly comes with real costs, but those costs are far less than continuing to get it wrong indefinitely. This is not to suggest that no one should take action. Rather, those actions that are taken should be, at a minimum, predicated on sound theory and then evaluated to determine their efficacy. The implementation of new programs and policies must be accompanied by evaluations. Resources should be expended to replicate those evaluations and ensure the accuracy of the findings. In this area, the USA, in particular, lags considerably behind other democratic countries, yet it never seems to learn its lesson. Each new use of force controversy is accompanied by a plethora of new ideas but little support for evaluating them.

Developing a truly evidence-based approach to the regulation of police use of force will take time and resources. But if this is—properly, in our view—regarded as an urgent social issue, then society should make funding more readily available. We should all demand that police agencies commit to not only relying on theoretically sound and evidence-based reforms but also to taking part in research that includes their own policing agencies.

References

Active Bystandership for Law Enforcement. (2021). Active bystandership for law enforcement (ABLE) project fact sheet. Retrieved from: https://drive.google.com/file/d/1PPImHwxyl8nYP8E38FI XVNKSXoQbCjf7/view

Adams, K. (1995). Measuring the prevalence of police abuse of force. In W. A. Geller & H. Toch (Eds.), *And Justice for All: Understanding and Controlling Police Abuse of Force* (pp. 61–98). Police Executive Research Forum.

Adams, I., Mourtgos, S., & Mastracci, S. (2021). High-stakes administrative discretion: What drives body-worn camera activations? *Public Administration Review, 81*, 691–703.

Alpert, G. P., & Dunham, R. G. (1997). *The force factor: Measuring police use of force relative to suspect resistance*. Police Executive Research Forum.

Alpert, G. P., & Dunham, R. G. (2004). *Understanding police use of force: Officers, suspects, and reciprocity*. Cambridge University Press.

Alpert, G., McLean, K., & Wolfe, S. (2017). Consent decrees: An approach to police accountability and reform. *Police Quarterly, 20*, 239–249.

Aronie, J, & Alpert, G. (2020). 16 to a dealer's 10: Could blackjack odds help inform police pursuit policies? PoliceOne. March 2. Retrieved from: https://www.police1.com/police-products/pursuit-manag ementtechnology/articles/16-to-a-dealers-10-could-blackjack-odds-help-inform-police-pursuit-polic ies-Y9x2Avr8SWzJG1c4/

Barone, E. (2020). The Trump administration's 'Operation Warp Speed' has spent $12.4 billion on vaccines. How much is that, really? Time. Retrieved from: https://time.com/5921360/operation-warps peed-vaccine-spending/. Accessed 11 Aug 2021

Chalfin, A., & Kaplan, J. (2021). How many complaints against police officers can be abated by incapacitating a few "bad apples?" *Criminology & Public Policy, 20*, 351–370.

Chanin, J. (2014). On the implementation of pattern or practice police reform. *Criminology, Criminal Justice, Law & Society, 15*, 38–56.

Chanin, J. (2015). Examining the sustainability of pattern or practice police misconduct reform. *Police Quarterly, 18*, 163–192.

Chanin, J. (2017). Early stages of pattern or practice police misconduct reform: An examination of the Department of Justice's investigation and negotiation processes. *Police Quarterly, 20*, 250–274.

Chillar, V. F. (2021). The racial divide at micro places: A pre/post analysis of the effects of the Newark Consent Decree on field inquiries (2015–2017). *Journal of Research in Crime and Delinquency*. https://doi.org/10.1177/00224278211030964

CCJ Task Force on Policing. (2021). Civilian oversight: Policy assessment. Retrieved from: https://assets.foleon.com/eu-west-2/uploads-7e3kk3/41697/civilian_oversight.2690411fd370.pdf. Accessed 11 Aug 2021

De Angelis, J., Rosenthal, R., & Buchner, B. (2016). Civilian oversight of law enforcement: A review of the strengths and weaknesses of various models NCJ Number 250265 Downloaded December 20, 2021). https://www.ojp.gov/ncjrs/virtual-library/abstracts/civilian-oversight-law-enforcement-assessing-evidence. Accessed 11 Aug 2021

Engel, R. S. (2000). The effects of supervisory styles on patrol officer behavior. *Police Quarterly, 3*, 262–293.

Engel, R. S. (2001). Supervisory styles of patrol sergeants and lieutenants. *Journal of Criminal Justice, 29*, 341–355.

Engel, R. S. (2002). Patrol officer supervision in the community policing era. *Journal of Criminal Justice, 30*, 51–64.

Engel, R. S., & Worden, R. E. (2003). Police officers' attitudes, behavior, and supervisory influences: An analysis of problem solving. *Criminology, 41*, 131–166.

Engel, R. S., Corsaro, N., Isaza, G. T., & McManus, H. D. (2022). Assessing the impact of de-escalation training on police behavior: Reducing police use of force in the Louisville, KY Metro Police Department. *Criminology & Public Policy, 21*, 199–233.

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*, 570–581.

Ericsson, K. A., & Ward, P. (2007). Capturing the nautrally occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. *Current Directions in Psychological Science, 16*, 346–350.

Finn, P. (2001). Citizen review of police: Approaches and implementtion. Washington, DC: National Insitute of Justice. Retrieved from: https://www.ojp.gov/pdffiles1/nij/184430.pdf

Fleming, J., & Rhodes, R. (2018). Can experience be evidence? Craft knowledge and evidence-based policing. *Policy and Politics, 46*, 3–26.

Fryer, R.G. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy. 127*, 1210–1261.

Fyfe, J. (1979). Administrative interventions on police shooting discretion: An empirical examination. *Journal of Criminal Justice, 7*, 309–323.

Giacomantonio, C., Goodwin, S., & Carmichael, G. (2020). Learning to de-escalate: Evaluating the behavioural impact of Verbal Judo training on police constables. *Police Practice and Research, 21*, 401–417.

Goldman, R. L. (2013). A model decertification law. *Saint Louis University Public Law Review, 32*, 147–156.

Goldman, R. L. (2016). The importance of state law in police reform. *Saint Louis University Law Journal, 60*, 363–390.

Goh, L. S. (2020). Going local: Do consent decrees and other forms of federal intervention in municipal police departments reduce police killings? *Justice Quarterly, 37*, 900–929.

Grunwald, B., & Rappaport, J. (2020). The wandering officer. *Yale Law Journal, 129*, 1676–1782.

Gullion, C., & King, W. (2020). Early intervention systems for police: A state-of-the-art review. *Policing: An International Journal, 43*, 643–658.

Harmon, R. (2009). Promoting civil rights through proactive policing reform. *Stanford Law Review, 62*, 1–68.

Huddleston, D. (2021). Marietta police officers learning Brazilian Jiu Jitsu to help keep the streets safe. WSB-TV 2. Retrieved from: https://www.wsbtv.com/news/local/cobb-county/marietta-police-offic

ers-learning-brazilian-jiu-jitsu-help-keep-streets-safe/CCYMTFR47FFD7B4QU5WOUQGC6Q/. Accessed 11 Aug 2021

International Association of Chiefs of Police. (1989). *Building integrity and reducing drug corruption in police departments*. IACP.

Jiao, A. (2021). Federal consent decrees: A review of policies, processes, and outcomes. *Police Practice and Research, 22*, 793–804.

Johnson, S., Guerette, R. T., & Bowers, K. (2014). Crime displacement: What we know, what we don't know, and what it means for crime reduction. *Journal of Experimental Criminology, 10*, 549–571.

Kane, R. J., & White, M. D. (2009). Bad cops: A study of career-ending misconduct among New York City police officers. *Criminology & Public Policy, 8*, 737–769.

Kinports, K. (2016). The Supreme Court's quiet expansion of qualified immunity. *Minnesota Law Review, 100*, 62–78.

Koen, M., & Mathna, B. (2019). Body-worn cameras and internal accountability at a police agency. *American Journal of Qualitative Research, 3*, 1–22.

Legewie, J. (2016). Racial profiling and use of force in police stops: How local events trigger periods of increased discrimination. *American Journal of Sociology, 122*, 379–424.

Loftin, C., McDowall, D., & Xie, M. (2017). Underreporting of homicides by police in the United States, 1976–2013. *Homicide Studies, 21*(2), 159–174.

Lum, C., Stultz, M., Koper, C. S., & Scherer, J. A. (2019). Research on body-worn cameras: What we know, what we need to know. *Criminology & Public Policy, 18*, 93–118.

Luna, E., & Walker, S. (1997). A report on the oversight mechanisms of the Albuquerque Police Department, prepared for the Albuquerque City Council. Retrieved from: https://www.cabq.gov/council/documents/Walker%20Luna%20Report.pdf

Madden, C., Easley, R., & Dunn, M. (1995). How journal editors view replication research. *Journal of Advertising, 24*, 77–87.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*, 537–542.

McCaffree, K. & Saide, A. (2021). How informed are Americans about race and policing? Skeptic Research Center, CUPES-007.

McLean, K., Wolfe, S. E., Rojek, J., Alpert, G. P., & Smith, M. R. (2020). Randomized controlled trial of social interaction police training. *Criminology & Public Policy, 19*, 805–832.

Mourtgos, S. M., & Adams, I. T. (2020). Assessing public perceptions of police use-of-force: Legal reasonableness and community standards. *Justice Quarterly, 37*, 869–899.

Mullinix, K. J., Bolsen, T., & Norris, R. J. (2021). The feedback effects of controversial police use of force. *Political Behavior, 43*, 881–898.

Nix, J., Campbell, B. A., Byers, E. H., & Alpert, G. P. (2017). A bird's eye view of civilians killed by police in 2015: Further evidence of implicit bias. *Criminology & Public Policy, 16,* 309–340.

NJ Advance Media. (2018). The force report. NJ.com. Retrieved from https://force.nj.com/. Accessed 11 Aug 2021

Noble, J., & Alpert, G. (2009). *Managing accountability systems for police conduct: Internal affairs and external oversight*. Waveland Press.

Office of Justice Programs. (2021). FY 2021 performance budget. US DOJ. Retrieved from: https://www.justice.gov/doj/page/file/1246736/download. Accessed 11 Aug 2021

Owens, E., Weisburd, D., Amendola, K. L., & Alpert, G. P. (2017). Research in brief: Passing it forward: Modeling procedural jutice in the workplace to improve public-police interactions. *Police Chief, 79*, 14–15.

Owens, E., Weisburd, D., Amendola, K. L., & Alpert, G. P. (2018). Can you build a better cop? Experimental evidence on supervision, training, and policing in the community. *Criminology & Public Policy, 17*, 41–87.

Perez, D. W. (1994). *Common Sense About Police Review*. Temple University Press.

Rushin, S. (2015). Structural reform litigation in American police departments. *Minnesota Law Review, 99*, 1343–1422.

Rushin, S. (2017). Police union contracts. *Duke Law Journal, 66*, 1191–1266.

Rushin, S. (2019). Police disciplinary appeals. *University of Pennsylvania Law Review, 167*, 545–610.

Rushin, S. (2021). Police arbitration. *Vanderbilt Law Review, 74,* 1023–1078.

Schrade, B. (2021). In Georgia agency, police train to shoot, not kill. The Atlanta Journal-Constitution. Retrieved from: https://www.ajc.com/news/crime/in-georgia-agency-police-train-to-shoot-notkill/IJNVJCHXBRHJHKPFHLEXQ672YI/

Schwartz, J. (2014). Police indemnification. *New York University Law Review, 89*, 885–1006.

Schwartz, J. (2017). How qualified immunity fails. *The Yale Law Journal, 127*, 2–76.

Schwartz, J. (2020). After qualified immunity. *Columbia Law Review, 120*, 309–388.

Schwartz, J. (2021). Qualified immunity's boldest lie. *The University of Chicago Law Review, 88*, 605–684.

Shjarback, J. A., White, M. D., & Bishopp, S. A. (2021). Can police shootings be reduced by requiring officers to document when they point firearms at citizens? *Injury Prevention.* https://doi.org/10.1136/injuryprev-2020-043932

Sierra-Arévalo, M., & Papachristos, A. (2021). Bad apples and incredible certitude. *Criminology & Public Policy, 20*, 371–381.

Skogan, W. G., Van Craen, M., & Hennessy, C. (2015). Training police for procedural justice. *Journal of Experimental Criminology, 11*, 319–334.

Stephens, D., Scrivner, E., & Cambareri, J. (2018). *Civilian oversight of the police in major cities*. Office of Community Oriented Policing Services.

Stoughton, S. W. (2018). Police body-worn cameras. *North Carolina Law Review, 96*, 1363–1423.

Stoughton, S. W. (2021). How the Fourth Amendment frustrates the regulation of police violence. *Emory Law Journal, 70*, 521–585.

Stoughton, S. W., Noble, J. J., & Alpert, G. P. (2020). *Evaluating police uses of force*. NYU Press.

Stoughton, S., McLean, K., Nix, J., & Alpert, G. (2021). Policing suspicion: Qualified immunity and "clearly established" standards of proof. *Journal of Criminal Law & Criminology, 112*, 1–42.

Terrill, W. (2005). Police use of force: A transactional approach. *Justice Quarterly, 22*, 107–138.

Terrill, W., & Paoline, E. A. (2013). Examining less lethal force policy and the force continuum: Results from a national use-of-force study. *Police Quarterly, 16*, 38–65.

Terrill, W., & Paoline, E. A. (2017). Police use of less lethal force: Does administrative policy matter? *Justice Quarterly, 34*, 193–216.

Thomson-Devaux, A., Bronner, L., & Sharma, D. (2021). Police misconduct costs cities millions every year. But that's where the accountability ends. The Marshall Project. Retrieved from: https://www.themarshallproject.org/2021/02/22/police-misconduct-costs-cities-millions-every-year-but-that-s-where-the-accountability-ends. Accessed 11 Aug 2021

Tyler, T. R. (2006). *Why people obey the law*. Princeton University Press.

Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. Russell Sage.

United States Commission on Civil Rights. (1981). *Who's Guarding the Guardians*. USGPO.

Walker, S. (2012). Institutionalizing police accountability reforms: The problem of making police reforms endure. *Saint Louis University Public Law Review, 32*, 57–92.

Walker, S., Alpert, G.P., & Kenney, D.J. (2001). Early warning systems: Responding to the problem police officer. National Institute of Justice: Research in Brief. https://www.ojp.gov/pdffiles1/nij/188565.pdf

White, M. D. (2001). Controlling police decisions to use deadly force: Reexamining the importance of administrative policy. *Crime & Delinquency, 47*, 131–151.

White, M. D., Mora, V. J., Orosco, C., & Hedberg, E. C. (2021a). Moving the needle: Can training alter officer perceptions and use of de-escalation? *Policing: An International Journal, 44*, 418–436.

White, M. D., Orosco, C., Pooley, M., & Sorenson, D. (2021b). Testing the impact of de-escalation training on officer behavior: The Tempe (AZ) Smart Policing Initiative. Retrieved from: https://www.smart-policing.com/sites/default/files/inline-files/Tempe%20SPI%20Final%20Report%2012-21.pdf. Accessed 11 Aug 2021

Wood, G., Tyler, T. R., & Papachristos, A. V. (2020). Procedural justice training reduces police use of force and complaints against officers. *Proceedings of the National Academy of Sciences of the United States of America, 117*, 9815–9821.

Worden, R., McLean, S. J., Engel, R. S., Cochran, H., Corsaro, N., Reynolds, D., Najdowski, C. J., & Isaza, G. T. (2020). The impacts of implicit bias awareness training in the NYPD. IACP/UC Center for Police Research and Policy & the John F. Finn Institute for Public Safety. Retrieved from: https://www1.nyc.gov/assets/nypd/downloads/pdf/analysis_and_planning/impacts-of-implicit-bias-awareness-training-in-%20the-nypd.pdf. Accessed 11 Aug 2021