

# Overcoming the Challenges of Experimental Research: Lessons From a Criminal Justice Case Study Involving TASER Exposure

Evaluation Review  
2018, Vol. 42(3) 358-385  
© The Author(s) 2018  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0193841X18803205  
journals.sagepub.com/home/erx



Natalie Todak<sup>1</sup>, Michael D. White<sup>2</sup>,  
Lisa M. Dario<sup>3</sup>, and Andrea R. Borrego<sup>4</sup>

## Abstract

**Objective:** To provide guidance to criminologists for conducting experiments in light of two common discouraging factors: the belief that they are overly time-consuming and the belief that they can compromise the ethical principles of human subjects' research. **Method:** A case study approach is used, based on a large-scale randomized controlled trial

---

<sup>1</sup> Department of Criminal Justice, University of Alabama at Birmingham, Birmingham, AL, USA

<sup>2</sup> School of Criminology and Criminal Justice, Arizona State University, Phoenix, AZ, USA

<sup>3</sup> School of Criminology and Criminal Justice, Florida Atlantic University, Boca Raton, FL, USA

<sup>4</sup> Department of Criminal Justice and Criminology, Metropolitan State University of Denver, Denver, CO, USA

## Corresponding Author:

Natalie Todak, Department of Criminal Justice, University of Alabama at Birmingham, 1202 University Boulevard, Birmingham, AL 35294, USA.

Email: ntod@uab.edu

experiment in which we exposed participants to a 5-s TASER shock, to describe how the authors overcame ethical, methodological, and logistical difficulties. **Results:** We derive four pieces of advice from our experiences carrying out this experimental trial: (1) know your limitations, (2) employ pilot testing, (3) remain flexible and patient, and (4) “hold the line” to maintain the integrity of the research and the safety of human subjects. **Conclusions:** Criminologists have an obligation to provide the best possible evidence regarding the impact and consequences of criminal justice practices and programs. Experiments, considered by many to be the gold standard of empirical research methodologies, should be used whenever possible in order to fulfill this obligation.

### **Keywords**

randomized controlled trial, experiment, police, TASER, attrition, criminology, harm

The randomized controlled trial (RCT) is considered the gold standard in research because its methodological design allows for isolating causal effects between an intervention and outcome (Farrington & Welsh, 2005; Sherman et al., 1998). In criminal justice, RCTs have established the effectiveness (or lack thereof) of a range of programs and practices. Further, “the promise of experimental criminology is that better evidence can help reduce harm and increase liberty” (Sherman, 2009, p. 23). Although RCTs in criminal justice have become more common over the last decade, they are still quite rare for a variety of reasons (Telep, Garner, & Visher, 2015; Weisburd, 2000; Weisburd, Telep, Hinkle, & Eck, 2010). There are concerns among some researchers that experiments are more vulnerable than less rigorous designs to ethical problems. Moreover, RCTs come with a high degree of difficulty, particularly with regard to logistical (e.g., cost) and methodological concerns (e.g., treatment integrity).

Although challenging, the RCT is the foundation of good science. In fact, advocates for experimental methods have argued that researchers have an obligation to carry out RCTs (Boruch, Snyder, & DeMoya, 2000; Farrington, 1983; Feder, Jolin, & Feyerherm, 2000; Weisburd, 2000). Weisburd (2003) stated:

There is a moral imperative for the conduct of randomized experiments in crime and justice. That imperative develops from our professional obligation to provide valid answers to questions about the effectiveness of treatments,

practices, and programs. It is supported by a statistical argument that makes randomized experiments the preferred method for ruling out alternative causes of the outcomes observed. (p. 336)

Still many find the ethical, logistical, and methodological challenges of carrying out an RCT to be overwhelming, and they often rely on less rigorous methods.<sup>1</sup> The reluctance to employ RCTs is explained, in part, by the lack of guidance to researchers on how to best overcome challenges.<sup>2</sup> This article seeks to address these concerns about experimental research designs. We adopt a case study approach, documenting our experiences in carrying out an RCT where half of the research participants received a controversial and by all accounts painful intervention: a 5-s TASER exposure (White et al., 2015). We found that participants experienced a statistically significant disruption in several dimensions of cognitive functioning following TASER exposure, but deficits did not last longer than 1 hr. In the following sections, the authors review the relevant literature on methodological rigor, RCTs, and their challenges, specifically in the context of policing research. We describe the ethical and methodological challenges encountered during the TASER cognition study, how they were overcome, and lessons learned. Although many of the challenges we faced were tied to the unique nature of the experiment, we believe the solutions applied to overcome those obstacles and the lessons learned have relevance for researchers employing experimental methods in other criminal justice settings. Overall, the article offers insights on how to manage the complexities of an experiment, even when the challenges seem daunting.

## **Experiments in Policing and Their Challenges**

The history of experiments in policing is complex, though many of the most influential studies in the 1970s, 1980s, and 1990s employed RCT designs (Farrington & Welsh, 2005). The Kansas City Preventive Patrol Study raised questions about the value of random preventive patrol and opened a dialogue about alternative methods of allocating personnel (Kelling, Pate, Dieckman, & Brown, 1974). In the Minneapolis Domestic Violence Experiment, Sherman and Berk (1984) examined the impact of varied interventions on likelihood of re-offending and found that arrest significantly reduced future domestic violence. Although replications produced mixed results (Sherman, Smith, Schmidt, & Rogan, 1992), the Minneapolis study facilitated a discussion about the proper police response to domestic violence, and more broadly the impact of arrest on recidivism. RCTs have also

documented the effectiveness of other police practices such as hot spots policing (Braga, 2005; Braga & Bond, 2008; Sherman & Weisburd, 1995) and foot patrol (Ratcliffe, Taniguchi, Groff, & Wood, 2011; Sorg, Haberman, Ratcliffe, & Groff, 2013). Since 2013, researchers have employed RCTs to test the impact of police body-worn cameras (BWCs; Ariel, Farrar, & Sutherland, 2015; Braga, Coldren, Sousa, Rodriguez, & Alper, 2017; Jennings, Lynch, & Fridell, 2015; White, Gaub, & Todak, 2018; Yokum, Ravishankar, & Coppock, 2017).

RCTs in criminal justice are difficult to implement due to ethical, political, logistical, and methodological challenges (Pettus-Davis, Howard, Dunningan, Scheyett, & Roberts-Lewis, 2016; Strang, 2012). Challenges are exacerbated by the fact that many social experiments occur in the field, rather than a controlled laboratory, and because they often involve vulnerable or protected populations (e.g., prisoners, children, the mentally ill). As a consequence, the federal government established a set of rules governing the ethics of research and three basic principles for the protection of human subjects (entitled *The Belmont Report*; National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978):

1. Respect for persons: treat persons as autonomous agents and protect those with diminished autonomy.
2. Beneficence: minimize harms and maximize benefits.
3. Justice: distribute benefits and risks of research fairly.

Most research in the criminal justice is reviewed and monitored by institutional review boards (IRBs). IRBs are composed of individuals with subject matter expertise who review project proposals and dictate changes to designs and procedures to insure projects adhere to these principles. Weisburd (2000) notes, however, that criminal justice researchers often avoid experimental designs because RCTs are vulnerable to ethical problems not commonly found with less rigorous methods. These ethical issues may result in greater scrutiny by funding agencies and IRBs, and that scrutiny may inhibit the use of RCT designs. For example, an RCT is grounded in a randomization process that will deprive half of the research subjects from receiving the intervention (which is often presumed to have some benefit). Moreover, in the case of an untested intervention, exposure to the treatment could actually generate harm for research subjects, thereby violating the beneficence principle of research (see, e.g., McCord, 2003). Importantly, however, this concern is not unique to RCTs, and in fact the RCT experiment provides a forum for participants to report on any harm

incurred that can then be formally documented. Further, advocates of experimental methods have pushed back against these ethical concerns. Boruch (1975) argued that the use of nonrandomized designs represents an ethical violation because they produce weaker evidence and force researchers to make equivocal judgments about treatment efficacy. Weisburd (2003) likewise observed that “ethical barriers can be overcome and that randomized experiments are appropriate in a very diverse group of [criminal justice] circumstances” (p. 337).

Practically speaking, RCTs come with a very high degree of difficulty and can be expensive, especially when compared to secondary data analysis (Hyman, 1972; Laub, Sampson, & Kiger, 1990). The researcher may need to convince a policy maker or practitioner to allow the research to take place. In the case of a police department, the chief must agree to a range of difficult requirements, including a randomization process that may disrupt normal operations (Eck, 2002) and will deprive half of the officers the intervention. Police officers may choose not to participate in the study for a host of reasons, or they may push back against their placement in a particular study group (Fienberg & Tanur, 1986; Goodman & Blum, 1996). Weisburd (2000, p. 187) noted that “experiments will be most difficult to implement when the researcher attempts to limit the discretion of criminal justice agents who generally operate with a great degree of autonomy and authority.” An RCT is also a long-term commitment that may not produce publishable findings for months or even years (as is any prospective study).

Criminal justice experiments also do not occur in a laboratory setting and as a consequence there are “real-world” methodological challenges to overcome. There may be departures to randomization, occurring when research subjects switch groups, and selection bias if true randomization is not used. In the Kansas City Preventative Patrol Experiment, for example, “the judgment of task force members and other police officials was used to override the computer-derived matching of beats in order to take into account idiosyncrasies of various beats” (Larson, 1975, p. 268). In White, Gaub, & Todak (2018)’s study of the impact of BWCs on officer use of force, the researchers were asked by agency executives to assign some officers to the experimental group so as to avoid discontent (i.e., some officers had worn cameras prior to the study and did not want to have them taken away).

Maintaining treatment fidelity, including applying the appropriate dosage of the intervention, is also a challenge, especially in longer RCTs (Weisburd, Petrosino, & Mason, 1993). The intended dosage of patrol in reactive beats was not realized in Kansas City, as patrol officers increased

their use of lights and sirens to get to the location of patrol calls, and the activities of dispatchers and specialty units were not controlled for in these areas (Larson, 1975). Researchers carrying out the Minneapolis Domestic Violence Study asked police officers to randomly employ different interventions upon arriving at calls involving domestic violence. They found it difficult to convince police officers to override their “street sense” and employ the randomly selected interventions on scene. In studies of offender populations, it has been noted that “clients who stand to benefit most from treatment (i.e., high-risk, high-needs) are the least likely to complete it” (Olver, Stockdale, & Wormith, 2011, p. 6). We may reasonably expect the same to be true for criminal justice actors. To use the BWC example, police officers who are most likely to receive citizen complaints or use force may also be less likely to activate their body cameras and thus realize the benefits of the technology. The concerns regarding treatment dosage and fidelity are certainly not unique to the RCT and can be problematic in less rigorous designs.

Contamination (or diffusion) across groups can also occur in real-world settings, when control subjects are either exposed to the intervention or are affected by the treatment of the experimental group. Contamination violates the Stable Unit Treatment Value Assumption (SUTVA) that the interventions received by individuals in one group do not affect the individuals in other groups (Imbens & Rubin, 2015; Maskaly, Donner, Jennings, Ariel, & Sutherland, 2017). Contamination is currently a major concern of scholars studying the effects of police BWCs since researchers cannot prevent treatment and control officers from responding to the same calls (Wallace, White, Gaub, & Todak, 2018). Violations of SUTVA compromise the researcher’s ability to make causal inferences between the independent and dependent variables. In addition to internal validity, there are trade-offs between the time investment of conducting an RCT and external validity (or generalizability), though replication studies can address this concern (Farrington, 2003). Drawing on a case study example, this article seeks to provide guidance for researchers regarding how to address these challenges.

## **The TASER Cognitive Functioning Study**

Conducted electrical weapons such as the TASER have become a fundamental part of the police arsenal. More than 17,000 law enforcement agencies spanning 107 countries have adopted the TASER, and the device has been deployed by law enforcement more than 3 million times in the field (Eisler, Szep, Reid, & Smith, 2017). Widespread use of the device has led to

questions about the physiological effects of TASER exposure (White et al., 2013). Among the 1,206 known deaths proximate to TASER exposure in the United States, 150 autopsies have identified the TASER as a contributing or causal factor, though the manufacturer of the device has vigorously disputed these claims (Eisler et al., 2017). Thus, a large body of research has studied the effects of TASER exposure on the body, especially the heart, respiration, metabolic effects, and stress. Generally, the research shows that the physiological risks are low and targeted to a specific group of vulnerable persons (Holder, Robinson, & Laub, 2011; Todak, Cesar, & Louton, 2015).

Virtually, no studies have explored the effects of TASER exposure on the brain, and this is problematic for two reasons. First, a body of research has documented deficits in neuropsychological functioning after exposure to electricity, particularly in memory, attention, and concentration (see Kane & White, 2016, for a review). Given the TASER generates a high-voltage (up to 50,000 V), low-amperage (2.1 mA) current of electricity, the literature on electrical injury provides a backdrop for considering the potential neuropsychological effects among suspects who receive a TASER exposure. Second, lawyers have also questioned whether such declines violate citizens' Fifth Amendment rights by reducing their capacity to knowingly, intelligently, and voluntarily waive their *Miranda* rights (Kane & White, 2016).

In order to address this knowledge gap, our research team carried out an RCT to test the effects of TASER exposure on cognitive functioning (funded by the U.S. Department of Justice, National Institute of Justice [NIJ]; Project # 2011-IJ-CX-0102). The project began with a small pilot study involving 21 police recruits who received a 5-s TASER exposure as part of their academy training (White, Ready, Kane, & Dario, 2014). Recruits were given a battery of cognitive tests 3–4 hr before exposure, within 5 min after exposure, and 24 hr after exposure. The study found the recruits experienced significant declines in several cognitive measures immediately after exposure, but all recruits returned to baseline within 24 hr. The experiences and results from the pilot study informed the design of the full RCT, in which 142 volunteers were split randomly into four treatment groups: control, TASER exposure only, physical exertion only, and physical exertion + TASER exposure (32–38 per group; White et al., 2015). During the study, participants completed a battery of cognitive tests at six points in time—several days prior to treatment, 1 hr prior to treatment, immediately after treatment, 1 hr after treatment, 1 day after treatment, and 1 week after treatment. Participants were required to attend four visits: Visit 1 (at the authors' university for screening, informed consent, and initial

cognitive testing), Visit 2 (at a hospital in Scottsdale, Arizona for additional screening, pretreatment testing, exposure to treatment, and posttreatment testing), Visit 3 (at the hospital for 1-day follow-up testing), and Visit 4 (at the hospital for 1-week follow-up testing). Among participants who received a TASER exposure, the research team documented statistically significant declines in scores for auditory learning and memory, as well as in subjective states of the participants (i.e., concentration, anxiety, and feeling overwhelmed), with the effects lasting for approximately 1 hr.

Study results were published in psychology (White et al., 2015) and criminology journals (Kane and White, 2016), with authors recommending that law enforcement agencies consider requiring officers to wait at least 1-hr post-TASER exposure before reading a suspect their *Miranda* rights and asking questions. The next section reviews the primary ethical, methodological, and logistical hurdles encountered during the study, as well as how those hurdles were overcome. Many of the challenges that emerged were a result of the distinctive nature of this study, but those challenges were grounded in the larger principles of experimental research (reviewed above), and the research team's responses reflected our effort to adhere to those principles.

## Challenges and Responses

### *Ethical*

The TASER cognitive functioning study presented a number of ethical challenges for the team because TASER exposure is a painful experience, there are known health risks associated with TASER exposure (Holder et al., 2011; Vilke et al., 2011), and the effects of the TASER on cognitive functioning are not known. The team employed a range of measures to minimize risk such as (1) creation of an interdisciplinary advisory board to guide the study, (2) review of the study protocol by three different IRBs, (3) inclusion of a pilot study involving police recruits already slated to receive a TASER study, (4) rigorous and multistage medical screening of potential participants, (5) thorough informed consent procedures, (6) implementation of the study in a hospital with a doctor and medical staff on hand, and (7) employment of TASER-trained law enforcement officers to administer the treatment condition. These measures are described below.

*Advisory board.* The researchers organized a collaborative advisory board of experts to guide the planning, design, and implementation of the study. The



board included two physicians (both medical consultants for the TASER manufacturer), two neuropsychologists (including the leading expert in the country on electrical injury), an attorney, and a police practices expert. In addition, the research team included three principal investigators (PIs; all criminologists), a team of physicians and medical professionals, two faculty from the college of nursing (with expertise conducting research with risk), three police officers, one biostatistician, and three full-time doctoral students. The advisory board provided critical input that guided all elements of the study. For example, both physicians serve as medical consultants for the TASER manufacturer, and as part of their own research, they have tased hundreds of research subjects. Their experience in this key area provided valuable input on the necessary safety protocols to reduce risk of injury. The neuropsychologists offered guidance on the dimensions of cognitive functioning to exam; they selected the validated cognitive tests that were employed during the study; and they trained the research team to administer and score those tests.<sup>3</sup> The neuropsychologists also recommended the use of “subjective state” questions that allowed participants to self-report their own levels of anxiety, clarity of thinking, and so on. The police practices expert recommended using alligator clips to administer the TASER exposure rather than the traditional mode which causes puncture wounds.<sup>4</sup> He also recommended that participants receive their exposure while lying face-down on a mat (to prevent injuries from falling over).

*Oversight by multiple IRBs.* The TASER cognitive functioning study was reviewed by three IRBs: the Arizona State University (ASU) IRB, the IRB for NIJ, and the Western IRB (WIRB; <https://www.wirb.com/Pages/default.aspx>). The ASU IRB served as the “IRB of record” since the PIs were faculty at this institution, and the NIJ human subjects protection officer (and other NIJ staff) also reviewed and approved the study. Since ASU does not have a medical school, the university IRB and the researchers decided to seek an additional level of review by WIRB, an independent for-profit organization (staffed by physicians, psychiatrists, etc.) that specializes in the review of biomedical research. WIRB conducted the full review (and annual reviews) of the study protocol to insure protection of human subjects, with consultation from NIJ and the ASU IRB. The multiple layers of human subjects review provided significant protections for research participants.

*Pilot study.* The research team conducted a pilot study over a 2-week period, from April 24, 2012 to May 2, 2012, with police recruits who were required to receive a TASER exposure as part of their training ( $n = 21$ ). The pilot

study served as a “test run” in a number of different ways. First, the pilot study allowed the research team and advisory board to assess whether the appropriate cognitive tests had been selected, and whether they were administered at the appropriate time periods. For example, the pilot study included cognitive testing immediately following TASER exposure, as well as the next day. Results from the pilot study indicated statistically significant declines in multiple dimensions of cognitive functioning following TASER exposure, but all subjects returned to baseline by the 24-hr mark. For the RCT, the researchers added a testing point at the 1-hr mark to better isolate the length of any effect. They also added testing at the 1-week mark to insure no longer term effects were observed. The pilot study also allowed the team to ensure that they were properly trained in the administration and scoring of the cognitive tests, and it gave the team an opportunity to “get a feel” for the nature of the TASER exposure and the logistics in terms of administering tests immediately following an exposure. Last, the research team and the IRBs agreed the study would have specific stopping rules, based on the results from the pilot study. That is, if the pilot study documented significant and persistent deficits in cognitive functioning among police recruits, the study would be halted.

*Medical screening of potential participants.* The RCT included medical and psychological screening of potential participants at multiple stages of the study. The screening procedures were developed based on the available research on risk of injury after TASER exposure (White et al., 2013). All participants were rigorously vetted and anyone who possessed a risk factor (see below) was excluded in the study. After an initial in-person recruitment session, all volunteers were first screened by phone to assess their initial eligibility. A “yes” response to any of the following questions resulted in the individual’s exclusion from the study (except for questions 1 and 2 where a “no” response is exclusionary).

- Are you able to read and understand the English language?
- Are you between the ages of 18 and 65 years old?
- Do you weigh less than 100 pounds?
- Do you have any significant health problems?
- (If female) Are you now or could you be pregnant?
- Do you now or have you ever used drugs that were not prescribed by your doctor?
- Have you ever been diagnosed with a psychiatric problem?
- Are you currently homeless?

- Do you now or have you ever been diagnosed with any cognitive problems, like problems remembering, problems with reading, or problems with attention?
- Have you ever been diagnosed with high blood pressure?
- Have you ever had an abnormal electrocardiogram?
- Have you ever had a heart attack, stroke, or transient ischemic attack (a ministroke that goes away in an hour or so)?
- Has your doctor told you that you should not exercise?
- Do you have chronic back pain?
- Have you ever been “tased” before (or been exposed to another kind of conducted electrical device)?
- Have you ever had an electrical injury before?

If the individual passed the phone screening, he or she was scheduled for an in-person visit with a PI, which involved an in-depth physical and mental health screening based on self-report. If an individual passed the second screening (and completed the consent process outlined below), then he or she was scheduled for Visit 2 at the hospital. During Visit 2 (the day treatment conditions were delivered, 3–5 days after Visit 1), the participant received a complete medical exam from a physician that included a review of health history, a breathalyzer test, a urine test for illicit drug use and pregnancy, measurement of blood pressure and pulse rate, and a 5-lead electrocardiogram to examine heart functioning. Upon completion of the medical exam, the PI and physician consulted to determine whether the individual was eligible for participation, and if so, the individual was formally admitted to the study.

*Rigorous informed consent procedures.* The research team developed rigorous informed consent procedures to ensure prospective participants were fully aware of the potential risks and benefits of study participation, as well as to convey the voluntary nature of their involvement (i.e., they could withdraw at any time). The PI led the informed consent process, which lasted approximately 45 min, with each prospective participant. The consent process began with each individual reading the 15-page consent document, followed by the PI then reviewing each section of the document (risks, benefits, voluntariness, requirements of study participation, etc.).<sup>5</sup> One of the key elements of this discussion involved compensation for study participation. All participants, regardless of group assignment, received US\$200 cash as compensation for their time and effort.<sup>6</sup>

Informed consent was ongoing at every stage of the study. For example, the PI emphasized the voluntariness of study participation at multiple key points during Visit 2 (when the treatment was delivered). This included when the participant first arrived at the hospital, when the participant completed the medical exam, when the participant was notified of group assignment, and when the participant was laying on the mat immediately before exposure.<sup>7</sup>

*Hospital setting for the RCT.* The original grant proposal to NIJ stated that the study would be carried out in the student health services building on the main campus of ASU. However, discussions with the advisory board and IRBs led to the researchers to rent out the first floor of a fully equipped hospital for study completion.<sup>8</sup> Shifting the study to a hospital helped to minimize risk of participants in the event of an adverse reaction to TASER exposure, and it facilitated completion of the comprehensive medical exams that were a critical part of participant screening. The hospital location paid dividends during the study, as several participants required medical observation following TASER exposure (e.g., drops in blood pressure that led to near-fainting), and one participant experienced an adverse event.<sup>9</sup>

*TASER-certified police officers administer the treatment.* The research team used grant funds to purchase a TASER device. However, researchers hired three certified TASER instructors from a local police department to administer the TASER exposures. The officers' experience and training insured that the TASER exposures were administered properly, lowering risk to participants and increasing the external validity of the study (approximating, to the extent possible, exposures in the field).

### *Methodological*

Just like any other experiment in criminal justice, the researchers had to address a number of threats to internal and external validity, as well as the SUTVA (Imbens & Rubin, 2015; Maskaly et al., 2017), such as participant attrition, random assignment departures, and contamination (control subjects exposed to the treatment). First, the researchers opted to perform the experiment in a laboratory setting. This decision allowed the research team to avoid several of these threats. Namely, there were no departures from random assignment. Potential participants were randomly assigned to a study group early in the screening process, after recruitment but well before formal admission to the study. In this way, the researchers were able to

monitor sample size among all four study groups throughout the study. All 142 participants who completed the study were retained in their originally assigned group.<sup>10</sup> The researchers also avoided treatment contamination by blinding participants from their group assignment until the moments before they received their treatment. Also, each day at the hospital involved all four treatment conditions, so there was no way for a participant to determine group assignment prior to formal notification from the PI.

The research team did struggle with participant attrition, however. The researchers started with a large recruitment pool, given the time commitment for study participation (four visits totaling about 8 hr), the off-site location for Visits 2–4, the restrictive screening criteria (medical and psychological), and the nature of the target population (young undergraduate college students). Research team members successfully recruited approximately 900 students through in-person visits to the various campuses of ASU. Within a month of that recruitment visit, research team members contacted individuals by phone to gauge their continued interest, to conduct the initial screening, and to schedule Visit 1. A total of 214 individuals attended Visit 1 at ASU, passed the self-report medical/mental health screening and informed consent, and completed the first administration of cognitive tests. All 214 were then scheduled for Visit 2 at the hospital 3–5 days later. In the interim, research team members sent reminder e-mails and text messages regarding the upcoming visit in an effort to reduce no-shows. We also made calls during Visit 2 once participants were late for their appointment. Of the 214, 58 failed to appear at the hospital (27.1%). Of the 156 individuals who appeared at the hospital for Visit 2, 10 were screened out for drug/alcohol use (9 were documented through either self-report or urinalysis; 1 participant showed up intoxicated, as evidenced by breathalyzer results).<sup>11</sup>

All of these withdrawals happened prior to individuals' formal admission to the study, which occurred upon completion of the Visit 2 medical exam and consultation between the PI and physician. As a consequence, 146 participants were formally admitted to the study. Attrition following formal study admission was quite low ( $n = 4$  or 2.7%). Two participants withdrew following notification of group assignment, one experienced an adverse medical event, and one left the state for a family emergency following Visit 3 and never returned. The low attrition rate after Visit 2 is explained, in part, by the staggered compensation schedule. Participants received US\$50 at the end of Visit 2 (day of their treatment), US\$50 at the end of Visit 3 (1 day after treatment), and US\$100 at the end of Visit 4

(1 week after treatment). The compensation schedule was back-loaded to entice subjects to continue participation after receiving their treatment.

The decision to perform the study in a controlled environment and to use healthy college student participants as a proxy for suspects limited generalizability of the findings to police–citizen interactions as they occur in the real world. These participants were not injured, intoxicated, or in crisis, nor were they members of the population who are most likely to engage in violent encounters with the police—active offenders. Their college attendance also indicates that they are less likely to suffer from the consequences of an impoverished lifestyle common among those who routinely interact with police. As such, our ability to infer that the results would be similar in a real-world setting is limited. At the same time, the laboratory setting allowed us to minimize the methodological challenges, reviewed above, that often plague experiments in criminal justice. Moreover, the current study documented deficits in cognitive functioning among a young, healthy, well-educated sample. Arguably, the effects of TASER exposure on cognition would be exacerbated by the characteristics of a real-world population (older, less physically and psychologically healthy, intoxicated, etc.).

We recommend that researchers employ experiments in both laboratory and real-world settings and then compare the results from these different methods. This approach is currently underway in research examining police officer deadly force decision-making. Researchers have examined the impact of suspect and situational characteristics on police officer decisions to shoot in video game and computer simulators, virtual reality simulators, as well as in real-life deadly force contexts from incident reports. Although differing methodologies do tend to influence results, some consistencies can be observed across research designs—for example, the influence of suspect demeanor has been found to predict police use of force in laboratory settings (James, James, & Vila, 2018), but research from field studies has produced mixed results (see, e.g., Engle, Sobol, & Worden, 2000; Terrill & Mastrofski, 2002).

### *Logistical*

The research team faced logistical challenges in managing and carrying out the 6-monthlong experiment. The first challenge involved costs, especially unanticipated costs. The TASER cognitive functioning study was funded by NIJ (US\$408,377), which covered many of the personnel costs of the study (faculty time, graduate students, etc.). However, not all of the costs were foreseen. The study was originally set to take place at the student health

services center. Once it became clear the study needed to take place at a fully equipped medical facility (as a result of discussion with the home university's IRB), the researchers searched for a location that would allow full and exclusive access for consecutive weekends over 4 months. The researchers also did not budget for the cost of the hospital, so the PIs forfeited most of their compensation to cover it.<sup>12</sup> The researchers also did not budget for the cost of an external, for-profit IRB (WIRB), which totaled approximately US\$6,000.

The second logistical challenge involved recruitment and screening. During fall 2012, the team developed a proactive recruitment plan that involved a co-PI and student volunteers making weekly trips the four campuses at the university. Team members approached persons on campus and read from an approved script. If an individual expressed interest, the team member recorded their name, phone number, and e-mail address. Over a period of 3 months, this process produced a list of 900 persons. All 900 were then called for the initial screening and scheduling of Visit 1. Because of the impending winter break, the research team delayed start of the RCT until early January 2013. Unfortunately, the delayed start meant that a month or more lapsed between recruitment and follow-up phone calls. The phone screening process was arduous and time-consuming. Many students failed to respond to phone calls, were no longer interested in participating, or had given incorrect cell phone numbers. Many students also failed the screening questions, in particular regarding prior use of illicit substances.

Third, managing the Visit 2 days at the hospital, when treatments were delivered, was a complex undertaking. We scheduled 30 participants each day, and each participant spent 4 hr at the hospital. Days lasted from 8:00 a.m. to 6:00 p.m., with participant arrivals every 15 min. There were multiple stages of processing: a short review of informed consent by the PI (10 min), urinalysis and breathalyzer (variable time, depending on the participant),<sup>13</sup> medical exam (45 min), PI/physician consultation on study admittance (5 min), pretreatment cognitive testing (20 min), 1-hr wait period, study group notification and treatment (10 min), posttreatment cognitive testing (20 min), 1-hr wait period, 1-hr posttreatment cognitive testing (20 min), and payment and scheduling for Visit 3 (5 min). In order to keep the flow of participants moving, there were three medical exam rooms (one for breathalyzer/urinalysis, two for medical exams), four cognitive testing rooms, and a separate room to deliver the treatment. Each Visit 2 was staffed with approximately 20 research team members: the PI, 2 co-PIs and 3 PhD students (who conducted the cognitive testing), 6 medical personnel (hospital administrator, physician, 4 nurses), 3 police officers, 2

nursing professors, and 6–8 graduate student volunteers. The most difficult logistical part involved maintaining the appropriate cognitive testing times when participants were late to arrive, experienced delays during the medical exam<sup>14</sup> or experienced delays after receiving their treatment.<sup>15</sup> Fatigue among the research staff, especially the cognitive testers, was also a concern and PIs (who were also trained) would step in when needed.

The last challenge involved data entry and analysis. All of the cognitive testing was administered with paper and pencil. There were 142 study participants and 6 data collection points per participant, for a total of 852 data collection points. Eight cognitive tests were administered at each data collection point, for a total of nearly 7,000 individual cognitive tests. The PI created a data shell in SPSS Version 18 and two PhD students independently entered 68,600 cells of data in separate data sets. The PI then used a statistical function to conduct cell-by-cell reliability checks for the two independent data sets. There was an error rate of less than half of 1%, and discrepancies were resolved by consulting the original testing documents. Once the data were clean, the PI provided the data set to the team biostatistician for analysis.

## **Lessons Learned and Conclusions**

RCT is the most rigorous methodological design available to researchers. It is also the most difficult to carry out in the field, and as a consequence, RCTs are relatively rare in criminal justice (Farrington & Welsh, 2005; Weisburd, 2000), though their prevalence in the field is growing. Nevertheless, Weisburd (2003) and others have argued that researchers have an obligation to find answers about the impact of programs and practices that are employed in policing, court systems, and correctional settings. Does a particular police strategy reduce crime? Does a specialized court program reduce recidivism? Does a correctional treatment program help individuals overcome substance abuse? These are critically important questions that affect the lives, liberty, and quality of life of citizens. Results from an RCT represent the most definitive evidence on these questions about impact and consequences. If criminologists are to remain relevant, they will need to increasingly embrace rigorous research methods that provide evidence-based results that can inform criminal justice policy and practice.

Given there is little in the way of guidance for researchers on how to overcome the myriad challenges that arise during an RCT, the current article described how one team of researchers addressed ethical, methodological, and logistical challenges during an experiment that tested the



impact of TASER exposure on cognitive functioning. The ethical, methodological, and logistical challenges that arose during the TASER cognitive functioning study were both complex and a bit unique given the nature of the study. Nevertheless, the obstacles that emerged were tied to many of the core principles of experimental research, such as beneficence, voluntariness, and more common methodological issues in experimental research, including attrition and concerns about external validity. As a consequence, the solutions employed in the current study have relevance for experimental researchers.

Although the current study was successful, there are a number of things the research team could have done differently or better. Our first set of lessons concern our recruitment strategies. First, we recruited only college students primarily for the sake of convenience. The inclusion of older adults (who would still have to pass the screening processes) would have increased the external validity of the results, and it may also have lessened attrition. Older adult participants may have been more reliable in terms of attendance and less likely to be excluded because of drug use. Second, the monthlong delay between participant recruiting and scheduling of first visits (a consequence of the university's scheduled winter break) produced significant attrition. The research team could have either recruited earlier in the fall academic semester or waited to begin recruitment until after the spring semester had started. Either approach would have lessened the attrition. Generally, having a long break between recruitment and data collection was not ideal. Finally, the first visit occurred on one of the campuses of the research team's university, but the remaining visits occurred at the hospital which was a 30- to 40-min drive from the university campuses. This distance certainly contributed to attrition. In future situations like ours, researchers should provide participants with money for Lyft or some other rideshare program, or public transportation.

We learned additional lessons through the trials and errors of this study that could save future researchers both time and money. For one, the study advisory board recommended the use of "paper and pencil" cognitive tests, but a number of the tests used in the study can be administered electronically which would have streamlined the data collection and data entry process. The PI also delivered the informed consent "speech" on more than 200 occasions because we believed the personal approach was best given the potential risks of the treatment. In hindsight, the team could have recorded the consent speech and simply hit "play," rather than delivering the speech in-person, and had the PI present to answer questions.

In addition to these logistical hurdles, there are a number of additional lessons we learned that represent valuable guidance for other researchers carrying out RCTs. We discuss each of these lessons below and, to highlight their relevance for experiments in policing more broadly, we also discuss how these lessons would inform RCTs testing two timely and controversial interventions in policing: BWCs and de-escalation training.<sup>16</sup> Although we limit our focus here to two examples, we argue the lessons apply equally well to the rigorous study of other police technologies (e.g., license plate readers) and practices (e.g., problem-oriented policing, foot patrol). First, the three PIs had experience conducting RCTs, and they were well-versed in the literature on police use of the TASER, the effects of TASER exposure, and the basics of *Miranda* rights and waiver. However, they had no experience conducting research with risk, and their knowledge of neuropsychology, neuropsychological tests (including how to administer and score them), electrical injury, and the nuances of *Miranda* waiver were limited. The three PIs knew their limitations and they built a team of experts to fill the knowledge gaps. The inclusion of physicians, medical professionals, neuropsychologists, police officers, nursing professors, and a lawyer inevitably complicated some aspects of the process, but their inclusion insured the safety of research participants.

The importance of building an interdisciplinary team applies equally well to RCTs testing the impact of police BWCs and de-escalation training. BWCs are a relatively new technology (White, 2014), and the impact of cameras on the functioning of a police department, on police and citizen behavior, and on criminal justice case processing has been largely unknown (though the research base has grown rapidly in the last few years; see Maskaly et al., 2017). Although there has been much discussion recently about the importance of de-escalation training for police (Police Executive Research Forum, 2016), there has never been an evaluation of such training. As a result, the design of an RCT to study police BWCs or de-escalation training should be informed by a wide range of stakeholders internal and external to the police department, from patrol officers and training staff to prosecutors and victim advocates (depending on “the intervention”). With regard to BWCs, the U.S. Department of Justice has developed best practice guidelines for the planning and implementation of a BWC program, and two of the core recommendations involve developing an interdisciplinary working group to guide the project, and partnering with researchers to evaluate the program (Bureau of Justice Assistance, 2015; White, Gaub, & Todak, 2018). The Tempe (AZ) Police Department was recently funded to develop and evaluate de-escalation training (with an RCT; <http://www.strategiesfor>

policinginnovation.com/spi-sites/tempe-arizona-2017), and they too have developed an interdisciplinary working group that includes internal staff (patrol officers, training staff, leadership), curriculum developers, and researchers.

Second, the pilot study with police recruits provided valuable insights that guided the larger RCT. It was, by all accounts, a practice run giving the research team experience that facilitated successful implementation of the study RCT. The research team modified (and improved) the RCT in a number of ways based on the pilot study results. Additional cognitive tests were included. Additional post-TASER exposure testing points (1 hr, 1 week) were included. But most importantly, the pilot study provided the first-ever results regarding cognitive deficits following TASER exposure. The research team, advisory board, NIJ, and the three IRBs all agreed the study RCT would be stopped if the pilot study uncovered significant and persistent cognitive deficits. The research team shared the pilot study results with all of the relevant parties, and based on review and discussion of those results, the research team received approval to carry out the full RCT.

The value of a pilot study extends to the deployment and evaluation of police BWCs and de-escalation training. BWCs come with a very high degree of difficulty. They require an enormous investment in resources, and they impact nearly every aspect of a police department's operations (White, 2014). The deployment of police BWCs brings into play numerous weighty issues from citizen privacy and the recording of vulnerable populations to data security and technological infrastructure requirements. A pilot study allows a department to wade slowly into the deployment of this technology, which minimizes the burden, allows the department to troubleshoot problems, and limits the unintended effects of the cameras. The U.S. Department of Justice guidelines recommend a phased deployment of BWCs via a small pilot study (Bureau of Justice Assistance, 2015). The same logic applies with de-escalation training. Given there are no existing studies of police de-escalation training, a pilot study represents a viable mechanism for rolling out such a training. Rigorous research enhances the value of a pilot study by increasing the understanding of the impact and consequences of either BWCs or de-escalation training. In fact, a number of police departments have partnered with researchers to evaluate their pilot studies of BWCs with either RCTs or rigorous quasi-experiments (e.g., Mesa, AZ; Mesa Police Department, 2013; Phoenix, AZ; Katz, Kurtenbach, Choate, & White, 2015; Anaheim, CA; McClure et al., 2017).

Third, the research team had to be both flexible and patient in terms of overcoming obstacles. As stated previously, the location for the study RCT

was changed at significant unforeseen cost. Individuals who volunteered to participate were undergraduate college students. They were frequently late or, in many cases, failed to show up at all. The PI performed the informed consent process more than 200 times during a 4-month period. Ten individuals appeared at the hospital for Visit 2 but had to be excluded from the study and sent home because of a failed drug test (or self-reported drug use they failed to disclose during the phone screening and at the first visit). One individual showed up to the hospital intoxicated. Another failed to disclose a preexisting injury that led to an adverse event following TASER exposure. As a result of the adverse event, the police officers changed the positioning of individuals' arms during the TASER exposure (flat along the side rather than raised above the head). Finally, for a period of 4 months, the research team spent every weekend at the hospital, with each Saturday lasting nearly 12 hr and each Sunday lasting 5 hr (for 1-day and 1-week testing).

The same flexibility and patience extends to RCTs of BWCs or de-escalation training. For example, attrition can be a significant problem in BWC and de-escalation studies. Police officers may choose to not participate in the research. They may move to different assignments that jeopardize the study design (contamination), or they may leave the department. Some officers, or the department leadership, may challenge the random assignment. For example, several of the authors carried out an RCT to test the impact of BWCs. The authors randomly assigned all patrol officers into treatment (BWC) and control (no BWC) groups, but several of the officers assigned to the control group had participated in the BWC pilot study and resisted giving up their BWCs (resulting in departures to random assignment). In one of the author's current studies of de-escalation training, there has been discussion about forgoing random assignment for some officers who the department believes are in need of the training (e.g., low performers). Also, in the case of both BWCs and de-escalation training, the participants may not use the "intervention" as intended (e.g., failure to activate the BWC, or failure to use the de-escalation skills), thereby complicating researchers' attempts to evaluate the intervention. Similar challenges were reported by researchers working on the Minneapolis Domestic Violence Experiment, who struggled to convince police officers to employ the randomly selected interventions and not those the officers felt were best suited for the situations at hand (Sherman & Berk, 1984).

Last, while flexibility was a key, there were several instances where the research team "held the line," most notably with methodological challenges. The three PIs monitored in real time the gender and racial/ethnic makeup of each of the four study groups. As the end of the study neared,

groups began to differ on gender. Moreover, the “any prior drug use” screening question proved to be problematic for the study’s target population and led to numerous exclusions. The PIs had animated discussions about altering random assignment to address the gender gaps in groups (e.g., placing female participants in specific groups to equalize gender, rather than random assignment), and about loosening the restrictions on prior drug use (potentially increasing risk, as drug use is a risk factor for experiencing physiological complications after TASER exposure). The team also discussed the possibility of having “TASER” and “no-TASER” weekends to reduce the costs (only paying the police officers to be present on certain weekends, not all) and burden on the police officers (who had significant downtime because each study day included participants who were randomly assigned to non-TASER groups). However, the research team believed these mid-project changes would have weakened both the rigor of the design and confidence in the study findings.

The same lesson applies to studies of BWCs and de-escalation training. Researchers may need to push back on police department’s suggestions for violations to random assignment. Researchers may have to overcome officer resistance to the application of research protocols. They may also need to encourage the leadership to not make assumptions about an intervention’s impact (especially if the intervention has not been tested before; e.g., de-escalation training) and to be wary of implementation failure [e.g., low BWC activation rates]). There is little question that “holding the line” on these issues can make RCTs more difficult. In the end, however, the PIs’ decisions in the current study were guided by two important principles: preserving the integrity of the experimental design and prioritizing the safety of research participants.

### **Acknowledgments**

The authors thank Robert Kane, Justin Ready, Carl Yamashiro, Sharon Goldsworthy, and Darya McClain for their contributions to this project, the physicians and health-care professionals at Hope Research Institute and Freedom Pain Hospital in Scottsdale, AZ, and the project’s advisory group: Dr. George P. Prigitano, Dr. Neil H. Pliskin, Dr. Jeffrey D. Ho, Dr. Donald M. Dawes, Jeremy D. Mussman, Esq., and Chief Frank Balkcom. They also thank police officers Shawn Dirks and Brian Ong from the Glendale (AZ) Police Department, who administered the TASER exposures. Finally, the authors also thank the more than 20 graduate students from Arizona State University who assisted with data collection, and the 142 individuals who participated in the study.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Notes

1. Arguably, some of the challenges associated with randomized controlled trials (RCTs) described in this article could also occur in a rigorous quasi-experiment (e.g., attrition, challenges to beneficence), especially those that are prospective like the RCT. However, we believe the ethical, methodological, and logistical challenges of an RCT are more acute and warrant special focus.
2. As an exception, Volume 8, Issue 3 (2012) of the *Journal of Experimental Criminology* was devoted to RCTs. The issue included two valuable essays on the challenges of carrying out RCTs (Boruch, 2012; Strang, 2012) and a handful of articles describing specific RCTs (see also Pettus-Davis et al., 2016).
3. Early in the study, one of the neuropsychologists witnessed two of the PIs get tased, and he questioned them immediately after exposure. This experience allowed the neuropsychologist to observe the potential cognitive effects post-exposure, and it allowed the PIs to experience the treatment their study would deliver to participants.
4. Obviously, a treatment that creates visible wounds would draw additional scrutiny by the study institutional review boards (IRBs). Attaching the device to participants' clothing via alligator clips provides the same TASER exposure but avoided causing any sort of visible injury or damage to the participants' clothing.
5. He also shared personal anecdotes of his own experience getting tased.
6. Researchers originally proposed US\$250 as compensation, but The Western Institutional Review Board (WIRB) felt that amount was overly coercive, given participants were undergraduate students. The US\$200 was paid in installment across Visits 2–4 at the hospital, upon completion of each visit.
7. Participants were not aware of their group assignment until immediately before receiving their assigned treatment (e.g., blinded assignment). Upon group notification, the principal investigator (PI) asked each individual if he or she wanted to continue. For those receiving a TASER exposure, the PI again asked about continued participation as the individual lay on the mat with TASER clips attached (with the PI crouching next to the individual). Notably, three

- participants withdrew upon group notification, and one withdrew while lying on the mat immediately prior to exposure.
8. The study took place at the hospital on Saturdays and Sundays for 4 consecutive months, January to April, 2013.
  9. The participant failed to disclose a preexisting shoulder injury. When the participant was tased, the exposure caused a shoulder separation that was immediately diagnosed by the physician on scene. The participant was transported to a nearby emergency room for treatment. The adverse event was report to all three IRBs.
  10. The research team and WIRB agreed on a sample size not to exceed 105. This would provide enough statistical power to detect treatment effects, while also limiting risk to a relatively small number of study participants.
  11. Recall that any prior illicit drug use (even just once) served as an exclusionary criterion.
  12. The PI forfeited three course releases. One of the co-PIs left ASU just prior to the start of the RCT, and he forfeited nearly all of his compensation (except for travel funds to return on multiple weekends to help with data collection). The cost of the hospital exceeded US\$40,000.
  13. One participant needed about 90 min before he could urinate.
  14. In addition to delays in being able to urinate, 5–7 participants had blood pressure that exceeded the study parameters. One nurse on the team was especially skilled at relaxing anxious participants to bring their blood pressure down, but this would take on average 20–25 min.
  15. The adverse event and medical observation of a few participants (e.g., near-fainting) were mentioned previously. Several participants also had emotional reactions (e.g., crying) to the TASER exposure and required several minutes to collect themselves.
  16. Several of the authors have carried out RCTs with police BWCs and de-escalation training.

## References

- 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*, 1–27.
- Boruch, R. (1975). On common contentions about randomized field experiments. In V. Glass (Ed.), *Evaluation studies review annual* (Vol. 1, pp. 158–194). Beverly Hills, CA: Sage.
- Boruch, R. (2012). Deploying randomized field experiments in the service of evidence-based crime policy. *Journal of Experimental Criminology*, *8*, 331–341.

- Boruch, R., Snyder, B., & DeMoya, D. (2000). The importance of randomized field trials. *NCCD News*, 46, 156–180.
- Braga, A. A. (2005). Hot spots policing and crime prevention: A systematic review of randomized controlled trials. *Journal of Experimental Criminology*, 1, 317–342.
- Braga, A. A., & Bond, B. J. (2008). Policing crime and disorder hot spots: A randomized controlled trial. *Criminology*, 46, 577–607.
- Braga, A. A., Coldren, J. R., Sousa, W., Rodriguez, D., & Alper, O. (2017). *The benefits of body-worn cameras: New findings from a randomized controlled trial at the Las Vegas Metropolitan Police* (No. 2013-IJ-CX-0016). Final Report to the National Institute of Justice.
- Bureau of Justice Assistance. (2015). *Law enforcement implementation checklist. Body-worn camera toolkit*. Washington, DC: Author. Retrieved from [bja.gov/bwc/pdfs/BWCImplementationChecklist.pdf](http://bja.gov/bwc/pdfs/BWCImplementationChecklist.pdf)
- Eck, J. E. (2002). Learning from experience in problem-oriented policing and situational prevention: The positive functions of weak evaluations and the negative functions of strong ones. *Crime Prevention Studies*, 14, 93–117.
- Eisler, P., Szep, J., Reid, T., & Smith, G. (2017). *Shock tactics part 1: The toll*. Ontario, CA: Reuters Investigates.
- Engle, R. S., Sobol, J. J., & Worden, R. E. (2000). Further exploration of the demeanor hypothesis: The interaction of suspects' characteristics and demeanor on police behavior. *Justice Quarterly*, 17, 235–258.
- Farrington, D. P. (1983). Randomized experiments on crime and justice. *Crime and Justice*, 4, 257–308.
- Farrington, D. P. (2003). A short history of randomized experiments in criminology: A meager feast. *Evaluation Review*, 27, 218–227.
- Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology*, 1, 9–38.
- Feder, L., Jolin, A., & Feyerherm, W. (2000). Lessons from two randomized experiments in criminal justice settings. *NCCD News*, 46, 380–400.
- Fienberg, S. E., & Tanur, J. M. (1986). The design and analysis of longitudinal surveys: Controversies and issues of cost and continuity. In R. Pearson & R. Boruch (Eds.), *Survey research designs: Towards a better understanding of their costs and benefits*. New York, NY: Springer.
- Goodman, J. S., & Blum, T. C. (1996). Assessing the non-random sampling effects of subject attrition in longitudinal research. *Journal of Management*, 22, 627–652.
- Holder, E. H., Robinson, L. O., & Laub, J. H. (2011). *Police use of force, tasers and other less-lethal weapons (NIJ Research in Brief)*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.



- Hyman, H. H. (1972). *Secondary analysis of sample surveys: Principles, procedures, and potentialities*. New York, NY: Wiley.
- Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. New York, NY: Cambridge University Press.
- James, L., James, S., & Vila, B. (2018). Testing the impact of citizen characteristics and demeanor on police officer behavior in potentially violent encounters. *Policing: An International Journal*, 41, 24–40.
- Jennings, W. G., Lynch, M. D., & Fridell, L. (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.
- Kane, R. J., & White, M. D. (2016). TASER<sup>®</sup> exposure and cognitive impairment. *Criminology & Public Policy*, 15, 79–107.
- Katz, C. M., Kurtenbach, M., Choate, D. E., & White, M. D. (2015). *Phoenix, Arizona, smart policing initiative: Evaluating the impact of police officer body-worn cameras* (Smart Policing Initiative Spotlight Report). Washington, DC: CNA.
- Kelling, G. L., Pate, T., Dieckman, D., & Brown, C. E. (1974). *The Kansas City preventive patrol experiment: A summary report*. Washington, DC: Police Foundation.
- Larson, R. C. (1975). What happened to patrol operations in Kansas City? A review of the Kansas City preventative patrol experiment. *Journal of Criminal Justice*, 3, 267–297.
- Laub, J. H., Sampson, R. J., & Kiger, K. (1990). Assessing the potential of secondary data analysis: A new look at the Gluecks' unraveling juvenile delinquency data. In K. L. Kempf (Ed.), *Measurement issues in criminology* (pp. 241–257). New York, NY: Springer.
- Maskaly, J., Donner, C., Jennings, W. G., Ariel, B., & Sutherland, A. (2017). The effects of body-worn cameras (BWCs) on police and citizen outcomes: A state-of-the-art review. *Policing: An International Journal of Police Strategies & Management*, 40, 672–688.
- 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*. Washington, DC: Urban Institute.
- McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *The ANNALS of the American Academy of Political and Social Science*, 587, 16–30.
- Mesa Police Department. (2013). *On-officer body camera system: Program evaluation and recommendations*. Mesa, AZ: Mesa Police Department.

- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Bethesda, MD: Office of Human Subjects Research, National Institute of Health.
- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2011). A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. *Journal of Consulting and Clinical Psychology, 79*, 6–21.
- Pettus-Davis, C., Howard, M. O., Dunnigan, A., Scheyett, A. M., & Roberts-Lewis, A. (2016). Using randomized controlled trials to evaluate interventions for releasing prisoners. *Research on Social Work Practice, 26*, 35–43.
- Police Executive Research Forum. (2016). *Use of force: Taking policing to a higher standard: 30 guiding principles* (Critical Issues in Policing). Washington, DC: Author.
- Ratcliffe, J. H., Taniguchi, T., Groff, E. R., & Wood, J. D. (2011). The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology, 49*, 795–831.
- Sherman, L. W. (2009). Evidence and liberty: The promise of experimental criminology. *Criminology & Criminal Justice, 9*, 5–28.
- Sherman, L. W., & Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. *American Sociological Review, 49*, 261–272.
- Sherman, L. W., Smith, D. A., Schmidt, J. D., & Rogan, D. P. (1992). Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review, 57*, 680–690.
- Sherman, L. W., Strang, H., Barnes, G. C., Braithwaite, J., Inkpen, N., & Teh, M. M. (1998). *Experiments in restorative policing: A progress report on the Canberra reintegrative shaming experiments*. Canberra, AU: Law Program Research School of Social Sciences Institute of Advanced Studies, Australian National University.
- Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime “hot spots”: A randomized, controlled trial. *Justice Quarterly, 12*, 625–648.
- Sorg, E. T., Haberman, C. P., Ratcliffe, J. H., & Groff, E. R. (2013). Foot patrol in violent crime hot spots: The longitudinal impact of deterrence and posttreatment effects of displacement. *Criminology, 51*, 65–101.
- Strang, H. (2012). Coalitions for a common purpose: Managing relationships in experiments. *Journal of Experimental Criminology, 8*, 211–225.
- Telep, C. W., Garner, J. H., & Visher, C. A. (2015). The production of criminological experiments revisited: The nature and extent of federal support for experimental designs, 2001–2013. *Journal of Experimental Criminology, 11*, 541–563.

- Terrill, W., & Mastrofski, S. D. (2002). Situational and officer-based determinants of police coercion. *Justice Quarterly*, 30, 143–155.
- Todak, N. E., Cesar, G. T., & Louton, B. (2015). Forensic reporting of TASER exposure: An examination of situational and exposure characteristics. *Journal of Forensic and Legal Medicine*, 35, 4–8.
- Vilke, G. M., Chan, T. C., Sloane, C. M., Neuman, T., Castillo, E. M., & Kolkhorst, F. (2011). *The effect of TASER on cardiac, respiratory, and metabolic physiology in human subjects*. Washington, DC: U.S. Department of Justice.
- Wallace, D., White, M. D., Gaub, J. E., & Todak, N. (2018). Body-worn cameras as a potential source of de-policing: Testing for camera-induced passivity. *Criminology*, 56, 481–509.
- Weisburd, D. (2000). Randomized experiments in criminal justice policy: Prospects and problems. *Crime & Delinquency*, 46, 181–193.
- Weisburd, D. (2003). Ethical practice and evaluation of interventions in crime and justice: The moral imperative for randomized trials. *Evaluation Review*, 27, 336–354.
- Weisburd, D., Petrosino, A., & Mason, G. (1993). Design sensitivity in criminal justice experiments. In M. H. Tonry (Ed.), *Crime and justice* (Vol. 17, pp. 337–379). Chicago, IL: University of Chicago Press.
- Weisburd, D., Telep, C. W., Hinkle, J. C., & Eck, J. E. (2010). Is problem-oriented policing effective in reducing crime and disorder? Findings from a Campbell systematic review. *Criminology and Public Policy*, 9, 139–172.
- White, M. D. (2014). *Police officer body-worn cameras: Assessing the evidence*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.
- White, M. D., Gaub, J. E., & Todak, N. (2018). 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., Ready, J. T., Kane, R. J., & Dario, L. M. (2014). Examining the effects of the TASER on cognitive functioning: Findings from a pilot study with police recruits. *Journal of Experimental Criminology*, 10, 1–24.
- White, M. D., Ready, J. T., Kane, R. J., Yamashiro, C. T., Goldsworthy, S., & Bonds, McClain, D. (2015). Examining cognitive functioning following TASER exposure: A randomized controlled trial. *Applied Cognitive Psychology*, 29, 600–607.
- White, M. D., Ready, J.T., Riggs, C., Dawes, D. M., Hinz, A., & Ho, J. D. (2013). An incident-level profile of TASER device deployments in arrest-related deaths. *Police Quarterly*, 16, 85–112.
- Yokum, D., Ravishankar, A., & Coppock, A. (2017). *Evaluating the effects of police body-worn cameras: A randomized controlled trial*. Washington, DC: The Lab @ DC, Office of the City Administrator, Executive Office of the Mayor.

## Author Biographies

**Natalie Todak** is an assistant professor of Criminal Justice at the University of Alabama at Birmingham. Her research covers a range of topics in American policing, with a current focus on police body-worn cameras, diversity, and de-escalation strategies. Her work has been published in leading criminology and criminal justice journals, such as *Criminology*, *Criminology and Public Policy*, *Women & Criminal Justice*, and *Police Quarterly*.

**Michael D. White** is a professor in the School of Criminology and Criminal Justice at Arizona State University (ASU), and he is the Associate Director of ASU's Center for Violence Prevention and Community Safety. His primary research interests involve the police, including use of force, technology, and misconduct. His recent work has been published in *Criminology*, *Criminology and Public Policy*, and *Criminal Justice and Behavior*.

**Lisa M. Dario** is an assistant professor in the School of Criminology and Criminal Justice at Florida Atlantic University. Her current research includes police legitimacy perceptions, use of force, and problem-oriented policing. Her recent work has been published in *Justice Quarterly*, the *Journal of Experimental Criminology*, and *Criminal Justice and Behavior*.

**Andrea R. Borrego** is an assistant professor of Criminal Justice and Criminology at Metropolitan State University of Denver. Her research areas include use of force, fatal police-citizen encounters and LGBTQ victimization.