



Post-experimental follow-ups—Fade-out versus persistence effects: The Rialto police body-worn camera experiment four years on



Alex Sutherland^{a,d}, Barak Ariel^{b,e,*}, William Farrar^{c,f}, Randy De Anda^c

^a Communities, Safety & Justice RAND Europe, Westbrook Centre, Milton Road, Cambridge CB4 1YG, United Kingdom

^b Institute of Criminology, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, United Kingdom

^c Rialto Police Department, CA, United States

^d Jerry Lee Centre of Experimental Criminology, Institute of Criminology, University of Cambridge, United Kingdom

^e Institute of Criminology, Faculty of Law, Hebrew University, Mount Scopus, Jerusalem 91905, Israel

^f Administration of Justice Instructor, Mt. San Jacinto College, San Jacinto, CA, United States

ARTICLE INFO

Keywords:

Body-worn cameras
 Fade-out effect
 Persistent effect
 Randomized controlled trials
 Police
 Natural experiment

ABSTRACT

Purpose: Under certain conditions, experimental treatment effects result in behavioral modifications that persist beyond the study period, at times, even after the interventions are discontinued. On the other hand, there are interventions that generate brief, short-term effects that “fade out” once the manipulation is withdrawn or when the in-study follow-up period is completed. These scenarios are context specific.

Methods: This study reports the results from a three-year post-experimental follow-up from the world's first randomized controlled trial of police body-worn cameras.

Results: The results show that initial falls in rates of complaints against police and police use of force during arrest were sustained during the four years following the cameras being introduced.

Conclusions: The findings suggest that police officers do not become habituated to the effect of the body-worn cameras, and that persistence rather than fade-out effects may characterize this emerging technology.

1. Introduction

The Rialto Police Department was the first police department in the world to participate in a randomized controlled trial of police body-worn cameras. That study, known as the “Rialto experiment,” was first published in 2014 (Ariel, Farrar, & Sutherland, 2015), and quickly gained attention following a renewed focus on critical incidents involving officers' shootings in the United States, which sadly continues to this day. Concerns with police accountability, police legitimacy, and use of force in police–public contacts have led to two intertwined phenomena: public upheaval on the one hand (Ransby, 2015), and de-policing (i.e., police withdrawal from proactive engagement with the public; see Oliver, 2015; Pyrooz, Decker, Wolfe, & Shjarback, 2016), on the other. From both sides of this spectrum, body-worn cameras were proposed as a potent solution. Civil liberties organizations such as the American Civil Liberties Union (ACLU) have promoted the use of body-worn cameras to increase the accountability of armed police officers (Stanley, 2013). The police profession pushed for mass rollout as a strategy to reduce some of the tensions with minority groups that recently surfaced, as well as to provide much-needed evidence on police–public encounters (see Lum, Koper, Merola, Scherer, & Reioux, 2015).

The Rialto Experiment (Ariel et al., 2015) provided evidence on the benefits of body-worn cameras in three major ways: first, the study suggested that using body-worn cameras causes a reduction of about 50% in the use of police force compared with control conditions. It also suggested a dramatic reduction in complaints lodged against Rialto police officers, of > 90%, compared with the year prior to the experiment. Finally, the study suggested that the benefits of the equipment justify the costs, with about a 4:1 ratio (see also Ariel, 2016).

There are at least two critical questions about the findings from the Rialto Experiment and they are both linked to the issue of study validity (Shadish, Cook, & Campbell, 2002). First, are the findings from the Rialto Experiment replicable in other settings? Rialto might have been “special” in some way; therefore, the conclusions may have been susceptible to a site selection bias (Allcott, 2015), as Rialto is just one police department from the “universe” of police departments. If this is the case, the findings would not be generalizable. However, this question has been at least partly answered through the Cambridge University Replication Study (CURS) (Ariel et al., 2016b, 2016c, 2016a; see also Drover & Ariel, 2015; Henstock & Ariel, 2017). CURS used an identical methodology to Rialto in a dozen other jurisdictions in English-speaking police departments. CURS discovered virtually identical

* Corresponding author.

E-mail addresses: as2140@cam.ac.uk (A. Sutherland), ba285@cam.ac.uk, barak.ariel@mail.huji.ac.il (B. Ariel), wfarrar@msjc.edu (W. Farrar), ltlojack@aol.com (R. De Anda).

trends in terms of complaints against the police: an *average* overall reduction of 93% on a year-to-year comparison ($Z = -3.234$; $p \leq 0.001$; between-sites variation $Q = 4.905$; $p = 0.428$). In terms of the use of force, a similar pattern emerged, with significant reductions on a between-groups basis (SMD = -0.346 ; SE = 0.137 ; 95% CI -0.614 to -0.077), however only in sites that were characterized by high treatment fidelity (Ariel et al., 2016c; see also Slothower, Sherman, & Neyroud, 2015).

The second critical question deals with what we may call “Fade-out” effects: is the impact of body-worn cameras time dependent, and will the rate of complaints and/or use of force regress back to a pre-implementation mean, as if the body-worn cameras were never introduced? The present report is meant to deal with this practical as well as theoretical question, which underpins a key causal mechanism behind body-worn cameras. In other words, do officers (and suspects) eventually become desensitized to being videoed by a camera during interactions (Ariel, Sutherland, Henstock, Young, & Sosinski, 2017), limiting the effects of body-worn cameras in the longer term? As the findings suggest, responses to these queries both provide deeper insight into self-awareness theory and have direct implications for how to design interventions more optimally.

1.1. Study follow-up periods: taking a longer view

It is widely accepted that randomized controlled trials should have “appropriate” follow-up periods. There are known concerns with the lack of completeness of follow-up *during the in-trial period* of experiments—that is, during or immediately after an intervention has been administered to experimental units. These are mainly issues associated with biased causal estimates of the treatment effect and threats to the statistical power of the test (Juni, Altman, & Egger, 2001; Moher, Schulz, & Altman, 2001). There is no “recommended” in-trial follow-up period, and some flexibility is needed depending on intervention type and discipline. The majority of experiments follow up on the study participants during the grant lifecycle—usually not > 6, 12, or 24 months after the last case was randomly allocated into the study conditions (see Farrington & Welsh, 2005; for more on cascaded random allocation sequences, see Ariel & Farrington, 2010; Wittes, 2002).

In spite of the often substantial costs associated with experiments, it is rare that follow-ups are conducted beyond the life of the original study. This is a concern because “a treatment response restricted to this brief ‘in-trial’ period can potentially underestimate the long-term benefits of treatment and also may fail to detect delayed hazards” (Llewellyn-Bennett, Bowman, & Bulbulia, 2016: 1). This means that our knowledge about effectiveness is typically limited to the short-term, covering one or two years’ post-allocation of units into treatment conditions at most. These issues have been noted particularly in decision-making and education studies (see Allcott & Rogers, 2014; Protzko, 2015, respectively), but there are no apparent reasons they would not also characterize experiments in criminology.

Studies that did measure medium- and long-run effects of interventions have provided critical insight into various interventions.¹ These studies were able to unravel “legacy effects” (Ford, Murray, McCowan, & Packard, 2016), as well as delayed hazards, which are likely to materialize only years after participants were exposed to the treatment (see Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Leventhal, Fauth, & Brooks-Gunn, 2005; Schweinhart et al., 2005; Sherman & Harris, 2013, 2015). For example, the Milwaukee Domestic Violence Experiment (MILDVE) found, with a 23-year follow-up of domestic violence arrests, that death was more

prevalent in treatment compared with control groups (Sherman & Harris, 2013, 2015; see also Harris, Polans, Mazeika, & Sherman, 2016). Compared to control cases, victims whose partners were arrested and jailed (than if warned and allowed to remain at home) had a 60% greater risk of all-cause mortality ($p = 0.037$, 95% CI = risk ratio of 1:1.024 to 1:2.628). At 23 years after enrolment, suspects assigned to arrest were almost three times more likely to have died of homicide (at 2.25% of suspects) than suspects assigned to a warning (at 0.81%), a small to moderate effect size ($d = 0.39$; $p = 0.096$; relative risk ratio = 2.79:1; 90% CI = 1.0007 to 7.7696). These findings would not be known with the relatively short follow-up period of the original experiment (Sherman, 1990; Sherman, Smith, Schmidt, & Rogan, 1992).

To our knowledge, Schweinhart et al. (2005) carried out the longest follow-up in a criminological intervention study, on the “Perry Preschool Program”. Children were followed up over 40 years after attending a cognitively oriented preschool program aimed to increase thinking and reasoning abilities and school achievement and the children in the program, compared with control children, showed 35% fewer arrests. Treatment children eventually worked harder, were less likely to commit a crime, and participated in many fewer social pathologies than did control group members. Rightly so, their follow-up study was termed “lifetime effects” (see also Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010).

Two additional examples are noteworthy. First, Olds et al. (1998) conducted a 15-year follow-up of the effect of nurse home visitation on children’s criminal and antisocial behavior. Their study has shown that the children of visited mothers were arrested at a significantly (54%) lower rate than the children of nonvisited mothers. Second, Henggeler, Clingempeel, Brondino, and Pickrel (2002) had a four-year follow-up of multisystemic therapy with substance-abusing and substance-dependent juvenile offenders. Analyses demonstrated significant long-term treatment effects for aggressive criminal activity (0.15 versus 0.57 convictions per year) but not for property crimes. We note these additional yet rare studies as they indicate the limited extent of medium and long-term follow-up periods in our field.

1.2. Persistence, durability, and “fade-out” effects

The motivation for long-term follow-ups of interventions is to understand if treatment effects “persist” or “fade out” over time.² Allcott and Rogers (2014:3) differentiate between sustained treatment/control differences when treatments are continued for long periods of time (‘durability’), and if treatment effects are observed even after interventions have discontinued (‘persistence’). To illustrate, it has been shown that early childhood interventions are beneficial during, or immediately after, the intervention has been administered; however, as children move on to poorer quality schools after early childhood intervention, the treatment effect vanishes. Protzko (2015) conducted a meta-analysis of 39 randomized controlled trials aimed at increasing children’s IQ scores and investigated whether the effects were durable and persistent. The meta-analysis shows that after an intervention that successfully raised intelligence scores, the effects reduce to nil [effect size immediately after the intervention completed $d = 0.523$ (95% CI = 0.451 to 0.666); over time $b = -0.132$ /year (95% CI = -0.243 to -0.021)]. Protzko (2015) suggests that these reductions occur because those in the experimental group lose their IQ advantage over time. It may also be the case that control cases “catch up” with treatment cases in the long run. Although the end result is a nil difference, whether a jump in IQ then reduces in the treatment group, or control group eventually catch up have quite different

¹ Some of these measured effects were reviewed by Allcott and Rogers, 2014: p. 6 [internal references omitted]: “exercise, smoking, weight loss, water conservation, academic performance, voting, charitable donations, labor effort.”

² Sherman (1990) discusses some of these “after-the-fact” phenomena; within the context of policing, he refers to these as “residual deterrence” and “deterrence decay.”

implications for an intervention's utility and likely cost-effectiveness.³

In this respect, we may be mainly concerned with “fade-out effects” in criminology experiments as interventions are not typically sustained beyond the life of an intervention unless scaled up (which is discussed further below). Fade-out effects refer to circumstances in which the outcomes of an experimental manipulation diminish after a particular intervention ends (i.e., the effects do not “persist”). Fade-out can be ‘complete’ or ‘partial’. The former is where intervention effects reduce to nil, meaning treatment and control participants are indistinguishable from one another. Partial fade-out is where there is a reduction in the magnitude of the effect over time, but there is still a discernable difference.

There is also a distinction between diminished impact from a manipulation that is completed (i.e., is not administered anymore), versus a reduction in the potency of an intervention that is still in place but the experiment is completed. This might be termed ‘adaptation’ to an intervention – which could be a limitation of simplistic interventions such as ‘nudges’ (e.g. one would expect the effect of repeatedly receiving letters saying ‘most of your neighbors have already paid their taxes’ would eventually be zero). The former refers to instances where both the experiment as well as the application of the treatment have ceased, while the latter indicates that the experimentalist is “out of the picture” but participants continue to be exposed to the intervention. This distinction creates new challenges when there are no direct tangible outcome measurements to observe the *direct* treatment persistence effects. For example, there could be an interaction effect between the randomly allocated treatment and any number of post-experimental variables; the fundamental change is a result of the randomly allocated treatment, but the interaction effects are challenging to interpret.

Finally, long-term fade-out effects are not necessarily linear. Bailey, Duncan, Odgers, and Yu (2016) refer to this phenomenon as “fade out and re-emergence.” One documented example was found by Dodge et al. (2014) in the context of social skills training in young childhood. The “Fast Track program” provided a range of behavioral and academic services to a random subset of first-grade boys exhibiting behavioral problems. Impacts in elementary school were uniformly positive, producing improvements in the boys’ prosocial behaviors and reductions in their aggressive behaviors (Conduct Problems Prevention Research Group, 1999). By middle or high school, most of these effects had disappeared for all but the highest-risk boys (Conduct Problems Prevention Research Group, 2011), although impacts on some of these outcomes reappeared when the participants were assessed in their mid-20s (Dodge et al., 2014).

There are theoretical and practical implications of different long-term effects. If a particular treatment effect is to be sustained, how much “dosage” is required over time (see Sherman, 1990; Wain & Ariel, 2014)? Is there a risk of treatment effect reduction, even though the treatment provider continues the intervention—and if so, when? Likewise, if the treatment effect persists, then cost-effectiveness assessments require a longer time-horizon. But, again, for how long? While the literature cannot answer these questions directly, such questions are pertinent to any criminologist interested in measuring causal impacts. Further, cost-benefit calculations require long-term follow-up measures, despite the political pressures to reach policy-oriented conclusions following a study.

2. Fade-out versus persistence effects in the context of police body-worn cameras

In the case of body-worn cameras, the few published studies using rigorous designs (i.e., Ariel et al., 2015, 2016b, 2016c, 2016a) included

a few months of data before and after the experimental allocation, or up to one-year post-randomization (Ariel et al., 2015). The evidence from these studies suggests some equivocal effects of body-worn cameras on police use of force, complaints, and assaults against officers; this relatively small pool of research evidence does not match the extent of rollout of body-worn cameras, reaching hundreds of thousands of police officers in thousands of police departments worldwide (Cubitt, Lesic, Myers, & Corry, 2016). The discussion has already moved beyond “should this police force have cameras?” to “why doesn't this police force have cameras?” (Sutherland & Ariel, 2016a).

To inform the debate about the long-term efficacy of body-worn cameras in policing, this paper reports results for three years' post-experiment in Rialto. There are several possible expectations. On the one hand, it could be that rates of use of force and complaints would *increase* once the knowledge that the body-worn cameras were no longer part of an experiment. There are several explanations for this hypothesis. First, as soon as the researcher is “out of the picture,” police could revert to pre-camera behavior patterns. By implication, a measurable fade-out effect would be observed. This is somewhat similar to the Experimenter Effect, or the Research Participation effects (see Rosenthal & Jacobson, 1992). Another way of viewing this is as a “novelty effect”; the police officers performed differently at first because of the novelty of body-worn cameras in this profession, which may change their expectations and/or perform differently—but that is followed by regression to the officers’ “old ways.” We should be clear that the observed pattern would be the same in both scenarios.

On the other hand, the effect of body-worn cameras may persist over time, which in practice means that the use of force and complaints will remain the same, post-random assignment. Other research (Ariel et al., 2016b) suggests that repeated exposure to the cameras for officers may lead to “spill-over effects” and an overall suppression of complaints, which the authors refer to as “contagious accountability.” This implies that there is no experimenter effect and that the intervention continues to promulgate after the experiment is completed but the treatment is nevertheless still applied (given a force-wide rollout, as discussed below).

3. Data and methods

We report rates of use of force and complaints during an arrest for the year preceding the Rialto Experiment, the experimental year, and three years' post-experiment; five years in total (Table 1). In the original Rialto Experiment (Ariel et al., 2015), outcomes were reported as rates as per 1000 *contacts* with citizens, but the data below update this to per 1000 *arrests* so as to make the results comparable to other published literature (e.g., Ariel et al., 2016a; Henstock & Ariel, 2017).

It is important to understand that once the experimental year had finished, Rialto PD issued cameras to all frontline staff. This means that the intervention did not finish, but was rolled out across the police department. This means we are (i) assessing the impact of the rollout, but also, unusually, we are able to (ii) compare force-wide roll out with

Table 1
Complaints, use of force, arrests, and frontline officers in Rialto preexperiment, during experiment, and postexperiment figures.

	Preexperimental			During	Postexperimental		
	2009	2010	2011	2012	2013	2014	2015
Use of force	70	65	67	25 (17 v 8) ^b	32	33	35
Complaints	36	51	24	3	4	4	4
Arrests	– ^a	– ^a	3495	3823	3912	4143	4023
Frontline officers	54	54	54	54	88	88	88

^a Rialto PD used a different tracking system; comparability with following year may produce skewed results.

^b Control versus Experimental.

³ For a recent example of a catch-up effect in child development but where treatment differences persisted, see Hagen, Melby-Lervåg, and Lervåg (2017).

the results from a randomized controlled trial. The latter comparison allows us to assess whether there was an additional impact from the rollout.

To statistically assess changes over time we used an interrupted time-series design, implemented via the user-written *ITSA* command in Stata 14 (Linden, 2016). The analysis model consists of a continuous variable capturing the pre-implementation trend; a binary variable indicating when the intervention started that quantifies the difference between the pre- and post-implementation intercepts; and an interaction between the pre-implementation trend and intervention that captures the change in the overall trend following the intervention. From this, it is straightforward to calculate the post-implementation trend (by adding the pre-intervention trend and the difference in the trend following intervention). This is done during model estimation to obtain an associated standard error. Importantly, *ITS* models also accurately account for the autocorrelated error structure that is common to time-series data (Linden, 2015), adjusting standard errors accordingly.

4. Results

Table 1 below shows that in the baseline year and the experimental year the overall numbers of recorded arrests were very similar. There is a slight increase in the number of arrests in the experimental year (8%, or 328 more arrests), so we have to acknowledge that another side effect of the body-worn cameras could be a small proportional increase in arrests as the number of available police officers did not change over these years. However, it seems unlikely that this small increase in arrests—well within what we might expect in terms of normal variation—could drive the fall in complaints or incidents requiring the use of force noted when comparing the treatment and control groups. (It would require a degree of foresight and anticipation bordering on premonition to know that a given contact should be avoided as it would lead to a complaint and/or use of force. In a reactive service such as policing, such selection seems unlikely.)

We also see that while at baseline the number of recorded uses of force incidents is relatively stable, the number of complaints is more “chaotic,” jumping from 36 to 51 and then dropping to 24, with a stable number of available frontline officers ($n = 54$). When looking at the post-experimental year, both measures of interest—use of force and complaints—are stable over time. The number of arrests during the post-experimental period seems higher than the pre-experimental period (although we do not have data for 2009 and 2010 as a different recording system was then used by Rialto PD), but it can logically be explained by the increase in the number of frontline officers (54 versus 88)—more police means more arrests.

We now turn to the results for rates of complaints and use of force. Fig. 1 shows the rate of complaints per 1000 arrests and Fig. 2 shows

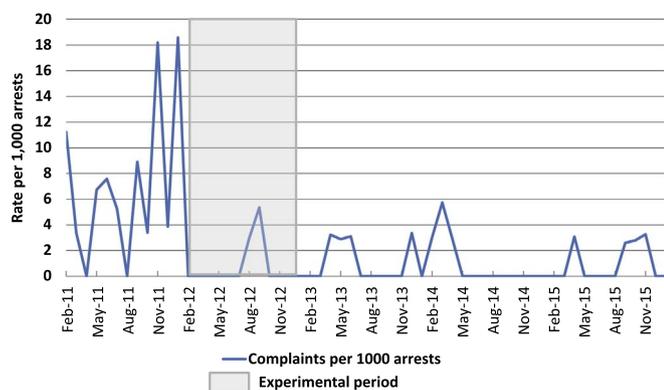


Fig. 1. Rate of complaints per 1000 arrests in Rialto before, during, and after experimental period.

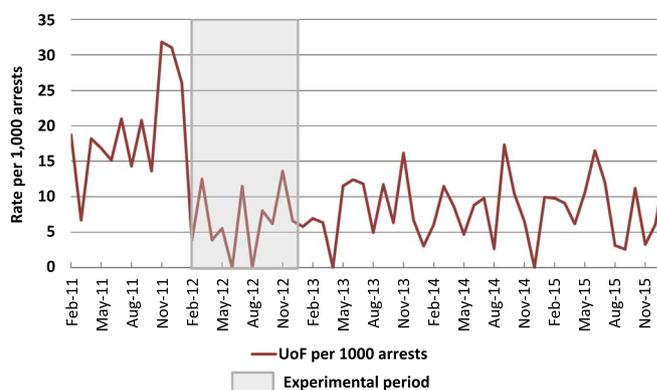


Fig. 2. Rate of use-of-force incidents per 1000 arrests in Rialto before, during, and after experimental period.

the rate of use-of-force incidents per 1000 arrests. In both cases, we can see that in the 12 months prior to the body-worn cameras experiment starting, both complaints and use of force were higher. During the experimental period, both fell (complaints almost to zero), as we have reported previously (Ariel et al., 2015). However, what is clear and striking is that in the three years after the experiment finished, both complaints and use of force remained comparatively low. All frontline officers were issued with cameras following the end of the experiment, and the results suggest that the effects associated with the cameras' introduction were sustained over this period.

As previously noted, between-group comparisons for complaints were not possible when reporting the original experiment because there were so few complaints (Ariel et al., 2015). This led us to assess pre/post differences using time-series models. We do the same below, with the justification being that although such comparisons are generally regarded as weaker evidence for causal effects (Sherman et al., 1998), one can arguably treat the exact timing of a policy change as a natural experiment. Furthermore, Fig. 1 suggests there may be some seasonality with the use of force, perhaps to be expected given the seasonality noted in crime, which is captured through the inclusion of time as a covariate.

Table 2 reports the results for complaints (Panel A) and use of force (Panel B). We see that, as previously reported, there was a sharp drop in complaints in the month cameras were introduced ($b = -11.56$; $SE = 4.46$; $p = 0.012$; 95% CI -20.493 to -2.618). This drop may have been exaggerated by the small spike in complaints in the months before the cameras were introduced (which hints at other factors potentially also changing), but it is clear that there was a relationship between the fall in complaints and the camera trial starting.

For use of force (Panel B), the pattern of results is very similar. A sharp drop associated with the month of the cameras' introduction ($b = -23.19$; $SE = 4.35$; $p \leq 0.000$; 95% CI -31.909 to -14.470), followed by a new “normal”; which is to say a much lower rate in the four years after the cameras began being used. It is interesting to note that although use of force was increasing over time before the cameras were introduced ($b = 1.36$; $SE = 0.48$; $p = 0.007$; 95% CI 0.39 to 2.33), there were no differences in the trends following either the introduction of the cameras (in-trial) or wider roll out (post-trial). This suggests that the effects of the cameras were already embedded by the time rollout took place, but also that the main effect of the cameras was a one-off fall in use of force in the month they were introduced, which was then sustained during rollout. It is also noted that there were no significant differences in complaints or use of force incidents between the original trial (which involved 54 officers wearing cameras for some shifts and not others) and the complete roll-out to the original 54 + 34 additional officers who were not in the original experiment).

Table 2
Interrupted time-series models for complaints and use of force.

	Coefficient	SE	<i>t</i>	<i>p</i>	95% CI LB ^b	95% CI UB ^c
Panel A: complaints per 1000 arrests						
Pre-BWC ^a trend (m1–12)	0.682	0.562	1.210	0.230	– 0.444	1.808
BWC trial starts (m13)	– 11.555	4.458	– 2.590	0.012	– 20.493	– 2.618
Δ post–pre trend (m13–24)	– 0.558	0.575	– 0.970	0.337	– 1.711	0.596
BWC trial ends (m24)	– 0.353	1.334	– 0.260	0.793	– 3.027	2.322
Δ post–pre trend (m24–60)	– 0.134	0.127	– 1.060	0.295	– 0.388	0.120
Intercept	2.826	3.509	0.810	0.424	– 4.210	9.861
m13–24 trend	0.124	0.124	0.999	0.322	– 0.125	0.374
m24–60 trend	– 0.010	0.024	– 0.407	0.685	– 0.058	0.038
Panel B: use of force per 1000 arrests						
Pre-BWC trend (m1–12)	1.359	0.483	2.810	0.007	0.391	2.327
BWC trial starts (m13)	– 23.190	4.349	– 5.330	0.000	– 31.909	– 14.470
Δ post–pre trend (m13–24)	– 1.086	0.616	– 1.760	0.083	– 2.321	0.148
BWC trial ends (m24)	– 0.362	2.907	– 0.120	0.901	– 6.190	5.466
Δ post–pre trend (m24–60)	– 0.248	0.389	– 0.640	0.526	– 1.027	0.531
Intercept	10.679	3.427	3.120	0.003	3.809	17.550
m13–24 trend	0.273	0.382	0.713	0.479	– 0.494	1.039
m24–60 trend	0.025	0.070	0.356	0.723	0.115	0.164

^a Body-worn cameras

^b Lower bound of confidence interval.

^c Upper bound of confidence interval.

5. Discussion

Our results show that following the introduction of body-worn cameras there was a sharp fall in the rate of complaints and use of force in Rialto, as previously reported in Ariel et al. (2015). Importantly, evidence from the three years after the end of the experiment shows that these initial falls persisted. We acknowledge that the results presented lack a comparison group, but we do not believe there is a credible alternative explanation for the fall in complaints. We know when the policy was implemented. The exact date of implementation could be regarded as random, but we know the results of the RCT itself led to a fall in complaints, and we know that the policy carried on following the trial. Under the assumption of “contagious accountability” (Ariel et al., 2016b), the medical analogy is that Rialto police officers are still “taking the pill” after the trial finished. An alternative explanation would require an intervention that covered all frontline officers, at the same time as the RCT began and persisted beyond the study period, and an intervention that directly influenced police–citizen interactions in such a way as to reduce complaints. To the best of our knowledge, there are no candidates that fulfill these requirements in Rialto—although they might in other police departments.

This new evidence from Rialto tells us that the effects of cameras have been maintained long after the experiment concluded. Our interpretation of this is that the cameras and associated changes in police practice (e.g., issuing a verbal warning; see Sutherland & Ariel, 2016b), once embedded as part of the experiment, simply became “habit” for officers. We do not have data to tell us whether officers actually changed how they approached interactions, for example by being more respectful, or adhering to practices believed to improve compliance and trust (i.e., procedural justice; see Bottoms & Tankebe, 2012; Lind & Tyler, 1988; however, cf. Nagin & Telep, 2017). At the same

time, that does not mean that all officers responded in the same way (Noppe, 2016). Nor do we know whether greater knowledge among citizens (and thus potential suspects) about the use of cameras by police influenced the general population's approach to interacting with the police. We are skeptical about the extent to which residents are aware of new police policy, although the recent focus on policing in the United States, in particular, may mean there is more awareness of some policies, but it may take residents years to acknowledge fully and consequently react to a body-worn camera policy. Similarly, one can imagine that among the subset of residents more inclined to break laws routinely, knowing that police wore cameras all the time would be quite important information, but that does not mean all would be sensitive to it (see Wikström & Treiber, 2007).

A more practical question arises here about police forces implementing body-worn cameras. Our results suggest that there was no additional impact following the force-wide roll out. This could be because officers were already used to behaving differently, or that the cameras were already being used by the officers most prone to using force/being complained about, but we cannot rule out other explanations such as senior-staff turnover or other time-varying factors. Similarly, we cannot know whether it was the roll-out itself that led to persistence of effects (e.g. because the roll-out was accompanied by a renewed focus on police–citizen encounters and/or additional training on why the cameras were being used. For more on the wider effects of single interventions, see Weisburd et al., 2017).

Although Rialto is typical of many US police forces both in terms of the size of the population it serves and the size of the police force itself, there is a question about whether the results are generalizable. Some (e.g., Deaton & Cartwright, 2016) argue for more care in making generalizations outside of the original context of a given study, and Cartwright and Hardie (2012) make clear that without the same supporting conditions being present, it is far from clear whether results will “travel.” We believe this may be the case for body-worn cameras' persistence effects; however, more evidence is needed from similar experimental sites over the long term before drawing strong conclusions.

Within policing research, our findings suggest that body-worn cameras are likely not due to short-term novelty effects. Whether Rialto officers remained sensitive to being observed by cameras, or changed the way they behave when dealing with members of the public, or members of the public behaved differently with Rialto officers, the result is the same: a reduction in police–public interactions resulting in use of force and/or complaints that was sustained. We cannot characterize the pathways of change that explain why the treatment effects persisted, however, strictly from a policy perspective, body-worn cameras may have long-lasting implications for police forces and the citizens they serve.

6. Methodological limitations

As with any pre-experimental design with a single group pretest–posttest methodology, there exists threats to the validity of the above assertions (Shadish et al., 2002). We are not able to rule out other alternative explanations for the potential (longitudinal) persistent causal relationship between the intervention and the outcomes. For example, possible changes in the departmental policies on use-of-force, training schemes that were instituted during the follow-up period, as well as the force-wide introduction of BWCs, are all possible explanations to the observed trend. Another potential phenomenon is de-policing (Ariel et al., 2016c), however this might explain the trend in some but not all follow-up years of the present study. Similarly, one could make the case that the absence of a significant *further* reduction in complaints or use-of-force, given the expansion in the number of police officers in Rialto, leads to additional explanations beyond the treatment effect. From a ‘purist’ perspective, these rival hypotheses create challenges in interpreting the observed trajectories.

However, we argue that there *cannot* be a satisfying control group,

to which the observed trends can be compared, when the unit of analysis is the entire police department ($n = 1$). Introducing a comparison group like another police department would be artificial, given the number of potential discrepancies and baseline imbalances between Rialto, its officers, organizational culture, crime trends, socio-demographic makeup of its residents and transient population, and the characteristics of the comparison site. Similarly, any comparison group within Rialto – for example, contrasting the behavior of the officers who took part in the original experiment (Ariel et al., 2015) with the behavior of the additional officers who joined the department post-experiment, is ‘synthetic’; given the virtually-guaranteed cross-over and contamination effect on the additional 34 new recruits, whom are all exposed to the treatment, organizational culture and departmental policies, the comparison will yield invalid results. Similarly, matching procedures (e.g., propensity-score matching) with such a small sample would be inadvisable (Austin, 2009:3091).

Finally, the observed reductions in both use of force as well as complaints were so dramatic, that even when substituting the causal explanation with a correlational explanation of body worn cameras for these persistent effects, our findings still carry a clear policy implication. The results show that the break in trend observed at the time of the experiment starting persisted to the end of the extended follow-up. If there are alternative explanations, they are linked to the introduction of body-worn cameras. These additional variables can potentially mask a different persistent effect size, but we fail to see how they would change its statistical significance, let alone its direction. The alternative would lead us to consider, for instance, that post-experimentally body worn cameras can increase the overall number of complaints against the police – a pattern that has not been reported anywhere, yet (Ariel, 2017; Ariel et al., 2016b). Expressing an accurate magnitude of causal estimates is desirable, however given the maximin rule of selecting the most fitting research design possible given the circumstances, our findings remain informative: we interpret these finding as consistent with behavioral learning or “contagious accountability,” (Ariel et al., 2016b), which are conceivable since all officers began wearing cameras at the end of the experiment.

7. Additional future research implications

There are three important questions which researchers should be mindful of. First, our study provides evidence against the idea that fade-out effects of body-worn cameras might arise as a function of experimenter effects. The experiment finished but the results persisted following rollout. Our study does not consider what would happen if the treatment itself was discontinued (i.e., removal of the cameras), although that has strong appeal. Such a design, especially if implemented experimentally with multiple treatment arms, could provide evidence about whether body-worn cameras are a sufficient and/or necessary condition for police behavior change that can be removed and effects still persist. Future research may provide the answer to this theoretically driven question. This has wide implications, particularly for learning theorists and social control scholars interested in understanding whether adaptation of behavior can be achieved when the manipulation that causes self-awareness to being observed is removed. A follow-up question would be: for how long should a relatively expensive intervention like the cameras be carried out for before a change in police officers' perception is fully embedded?

Second, it would be interesting see the type of evidence that BWCs provide toward successful prosecutions and internal disciplinary proceedings, and whether these outcomes vary over time in the same sort of pattern as we show for use of force and complaints. Should BWCs become useful in court, a long-term perspective should be implemented as well, as the effect of BWCs can fade out or vary over time. There is already laboratory evidence on “body-worn camera perspective bias” (Boivin, Gendron, Faubert, & Poulin, 2017:125): individuals' opinions on the appropriateness of policing are different when presented

evidence comes from a BWC than when it is seen from other types of testimonies (e.g., surveillance cameras). These considerations seem important to us, as the efficiency of BWCs evidence is discussed in the context of both criminal justice system outcomes as well as disciplinary hearings, on a longitudinal basis.

Furthermore, while it is clear that complaints against police fell in Rialto, we do not know, indeed we cannot know, whether the cameras deterred only spurious complaints, or whether they may also have deterred genuine complainants. Deterrence of genuine complaints with good grounds could arise through a lack of trust in how the police may use video evidence, but again it is not possible to know this without evidence. (The other conclusion, that only spurious complaints were deterred, seems unlikely—but we do not know the true ratio of genuine-to-spurious complaints.) If genuine complaints are being deterred, this could present risks to civil liberties and could foster a culture of impunity among officers—precisely the issues that led to initial calls to police to use body-worn cameras. This concern is not only Rialto-specific, given the global results from CURS, which detected similar reductions in complaints on a before–after basis. Future research should observe more closely these distinctions.

Finally, a more robust cost–benefit analysis of body-worn cameras is urgently required. Our crude analyses (Ariel et al., 2015) are an insufficient basis for the continuing (large) expenditure on this technology. If such a study takes place, it must take into account the total cost of the equipment (including support costs) along with the “total benefit” of the policy should be calculated. In light of the present study, policymakers and economists alike should consider the full economic costing and “profits” of body-worn cameras on a long-term basis. If the effect of body-worn cameras persists over time, measuring the cost of the equipment and the immediate contribution to policing on a short-term basis would be unnecessarily conservative.

References

- Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics* (qjv015).
- Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *The American Economic Review*, 104(10), 3003–3037.
- Ariel, B. (2016). The puzzle of police body cams. *IEEE Spectrum*, 53(7), 32–37.
- Ariel, B. (2017). The effect of police body-worn videos on use of force, complaints and arrests in large police departments. *Journal of Criminal Law and Criminology*, 106(4), 729–768.
- Ariel, B., Farrar, W. A., & Sutherland, A. (2015). The effect of police body-worn cameras on use of force and citizens' complaints against the police: A randomized controlled trial. *Journal of Quantitative Criminology*, 31(3), 509–535.
- Ariel, B., & Farrington, D. P. (2010). Randomized block designs. *Handbook of quantitative criminology* (pp. 437–454). Springer.
- Ariel, B., Sutherland, A., Henstock, D., Young, J., Drover, P., Sykes, J., ... Henderson, R. (2016b). “Contagious accountability” a global multisite randomized controlled trial on the effect of police body-worn cameras on citizens' complaints against the police. *Criminal Justice and Behavior* (0093854816668218).
- Ariel, B., Sutherland, A., Henstock, D., Young, J., Drover, P., Sykes, J., ... Henderson, R. (2016c). Report: Increases in police use of force in the presence of body-worn cameras are driven by officer discretion: A protocol-based subgroup analysis of ten randomized experiments. *Journal of Experimental Criminology*, 12(3), 453–463.
- Ariel, B., Sutherland, A., Henstock, D., Young, J., Drover, P., Sykes, J., ... Henderson, R. (2016a). Wearing body cameras increases assaults against officers and does not reduce police use of force: Results from a global multi-site experiment. *European Journal of Criminology*, 13(6), 744–755.
- Ariel, B., Sutherland, A., Henstock, D., Young, J., & Sosinski, S. (2017). The deterrence spectrum: Explaining why police body-worn cameras ‘work’ or ‘backfire’ in aggressive police–public encounters. *Policing*. <http://dx.doi.org/10.1093/policing/paw051>.
- Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28(25), 3083–3107.
- Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2016). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, 1–33.
- Boivin, R., Gendron, A., Faubert, C., & Poulin, B. (2017). The body-worn camera perspective bias. *Journal of Experimental Criminology*, 13(1), 125–142.
- Bottoms, A., & Tankebe, J. (2012). Beyond procedural justice: A dialogic approach to legitimacy in criminal justice. *The Journal of Criminal Law and Criminology*, 119–170.
- Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002).

- Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science*, 6(1), 42–57.
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.
- Conduct Problems Prevention Research Group (1999). Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology*, 67(5), 648–657.
- Conduct Problems Prevention Research Group (2011). The effects of the Fast Track preventive intervention on the development of conduct disorder across childhood. *Child Development*, 82(1), 331.
- Cubitt, T. I., Lesic, R., Myers, G. L., & Corry, R. (2016). Body-worn video: A systematic review of literature. *Australian & New Zealand Journal of Criminology*. <http://dx.doi.org/10.1177/0004865816638909>. The published version is here <http://journals.sagepub.com/doi/abs/10.1177/0004865816638909>.
- Deaton, A., & Cartwright, N. (2016). *Understanding and misunderstanding randomized controlled trials*.
- Dodge, K. A., Bierman, K. L., Coie, J. D., Greenberg, M. T., Lochman, J. E., McMahon, R. J., & Pinderhughes, E. E. (2014). Impact of early intervention on psychopathology, crime, and well-being at age 25. *American Journal of Psychiatry*, 172(1), 59–70.
- Drover, P., & Ariel, B. (2015). Leading an experiment in police body-worn video cameras. *International Criminal Justice Review*, 25(1), 80–97.
- 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(1), 9–38.
- Ford, I., Murray, H., McCowan, C., & Packard, C. J. (2016). Long term safety and efficacy of lowering LDL cholesterol with statin therapy: 20-year follow-up of west of Scotland Coronary Prevention study. *Circulation*, 115, 019014.
- Hagen, A. M., Melby-Lervåg, M., & Lervåg, A. (2017). Improving language comprehension in preschool children with language difficulties: A cluster randomized trial. *The Journal of Child Psychology and Psychiatry*. <http://dx.doi.org/10.1111/jcpp.12762> (online first).
- Harris, H. M., Polans, D. S., Mazeika, D., & Sherman, L. W. (2016). Retrieving administrative data to assess long-term outcomes: A case study of the 23-year follow-up of the Milwaukee domestic violence experiment. *Journal of Experimental Criminology*, 12(4), 599–608.
- Heckman, J. J., Moon, S. H., Pinto, R., Savellyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1), 114–128.
- Henggeler, S. W., Clingempeel, W. G., Brondino, M. J., & Pickrel, S. G. (2002). Four-year follow-up of multisystemic therapy with substance-abusing and substance-dependent juvenile offenders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 41(7), 868–874.
- Henstock, D., & Ariel, B. (2017). Testing the effects of police body-worn cameras on use of force during arrests: A randomised controlled trial in a large British police force. *European Journal of Criminology* (1477370816686120).
- Juni, P., Altman, D. G., & Egger, M. (2001). Assessing the quality of controlled clinical trials. *British Medical Journal*, 323(7303), 42.
- Leventhal, T., Fauth, R. C., & Brooks-Gunn, J. (2005). Neighborhood poverty and public policy: A 5-year follow-up of children's educational outcomes in the New York City moving to opportunity demonstration. *Developmental Psychology*, 41(6), 933.
- Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer Science & Business Media.
- Linden, A. (2015). Conducting interrupted time series analysis for single and multiple group comparisons. *Stata Journal*, 15, 480–500.
- Linden, A. (2016). *ITSA: Stata module to perform interrupted time series analysis for single and multiple groups*. (Statistical Software Components).
- Llewellyn-Bennett, R., Bowman, L., & Bulbulia, R. (2016). Post-trial follow-up methodology in large randomized controlled trials: A systematic review protocol. *Systematic Reviews*, 5(1), 214.
- Lum, C., Koper, C. S., Merola, L. M., Scherer, A., & Reiou, A. (2015). *Existing and ongoing body worn camera research: Knowledge gaps and opportunities*. (Report for the Laura and John Arnold Foundation).
- Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Medical Research Methodology*, 1(1), 2.
- Nagin, D. S., & Telep, C. W. (2017). Procedural justice and legal compliance. *Annual Review of Law and Social Science*, 13(1).
- Noppe, J. (2016). Are all police officers equally triggered? A test of the interaction between moral support for the use of force and exposure to provocation. *Policing and Society*, 1–14.
- Olds, D., Henderson, C. R., Jr., Cole, R., Eckenrode, J., Kitzman, H., Luckey, D., ... Powers, J. (1998). Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *JAMA*, 280(14), 1238–1244.
- Oliver, W. M. (2015). Depolicing rhetoric or reality? *Criminal Justice Policy Review* (0887403415586790).
- Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, 53, 202–210.
- Pyrooz, D. C., Decker, S. H., Wolfe, S. E., & Shjarback, J. A. (2016). Was there a Ferguson Effect on crime rates in large US cities? *Journal of Criminal Justice*, 46, 1–8.
- Ransby, B. (2015). The class politics of Black Lives Matter. *Dissent*, 62(4), 31–34.
- Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom* (Expanded ed.). New York: Irvington.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40*.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage Learning.
- Sherman, L. W. (1990). Police crackdowns: Initial and residual deterrence. *Crime and Justice*, 12, 1–48.
- Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P., & Bushway, S. D. (1998). *Preventing crime: What works, what doesn't, what's promising*. Research in Brief/National Institute of Justice.
- Sherman, L. W., & Harris, H. M. (2013). Increased homicide victimization of suspects arrested for domestic assault: A 23-year follow-up of the Milwaukee Domestic Violence Experiment (MildVE). *Journal of Experimental Criminology*, 9(4), 491–514.
- Sherman, L. W., & Harris, H. M. (2015). Increased death rates of domestic violence victims from arresting vs. warning suspects in the Milwaukee Domestic Violence Experiment (MildVE). *Journal of Experimental Criminology*, 11(1), 1–20.
- 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*, 680–690.
- Slothower, M., Sherman, L. W., & Neyroud, P. (2015). Tracking quality of police actions in a victim contact program: A case study of training, tracking, and feedback (TTF) in evidence-based policing. *International Criminal Justice Review*, 25(1), 98–116.
- Stanley, J. (2013). *Police body-mounted cameras: With right policies in place, a win for all*. New York: ACLU.
- Sutherland, A., & Ariel, B. (2016a). *Body worn cameras: Technology meets complexity*. International Council of Police Representative Associations (ICPRA).
- Sutherland, A., & Ariel, B. (2016b). *How body cameras curbed police use of force in Rialto: Monitoring of police–community interactions can improve behavior and ease tension*. Zócalo Public Square.
- Wain, N., & Ariel, B. (2014). Tracking of police patrol. *Policing* (pau017).
- Weisburd, D., Braga, A. A., Groff, E. R., & Wooditch, A. (2017). Can hot spots policing reduce crime in urban areas? An agent-based simulation. *Criminology*, 55(1), 137–173.
- Wikström, P.-O. H., & Treiber, K. (2007). The role of self-control in crime causation beyond Gottfredson and Hirschi's general theory of crime. *European Journal of Criminology*, 4(2), 237–264.
- Wittes, J. (2002). Sample size calculations for randomized controlled trials. *Epidemiologic Reviews*, 24(1), 39–53.