Received: 20 March 2019 | Revised: 13 November 2019 | Accepted: 8 December 2019

DOI: 10.1111/cpsp.12321

LITERATURE REVIEW

CLINICAL PSYCHOLOGY SCIENCE AND PRACTICE WILEY

When helping hurts? Toward a nuanced interpretation of adverse effects in gang-focused interventions

Miriam P. Rubenson | **Katharine Galbraith** | **Olivia Shin** | **Christopher R. Beam** | **Stanley J. Huey Jr.**

Department of Psychology, University of Southern California, Los Angeles, CA, USA

Correspondence
Miriam P. Rubenson, Department of Psychology, University of Southern California, 3620 S McClintock Ave, Ste 501, Los Angeles 90089.
Email: rubenson@usc.edu

Abstract

We examined gang interventions that produced adverse effects, or worse outcomes for treated participants in controlled evaluations. Prior studies describe harmful psychological processes and flawed research methods that can cause adverse effects in gang and crime prevention programs. We tested whether several program elements and methodological issues predicted adverse effects across 41 controlled evaluations. Only the presence of law enforcement officers as intervention providers significantly predicted adverse outcomes, and it is unclear whether law enforcement officers increased participants' antisocial behaviors or simply detected more offending among treated participants. Including law enforcement officers in gang interventions may increase the risk of adverse outcomes; however, further research is needed to determine why.

KEYWORDS

adverse effects, controlled evaluation, gang, intervention

1 | INTRODUCTION

Gang members are involved in more delinquency and serious crimes than non-gang members (Esbensen & Huizinga, 1993; Gatti, Tremblay, Vitaro, & McDuff, 2005; Krohn & Thornberry, 2008; Pyrooz, Turanovic, Decker, & Wu, 2016) and are responsible for a disproportionate number of homicides (Egley & Howell, 2014). Many intervention programs have attempted to reduce or prevent gang involvement and related antisocial behavior, and some of these have been evaluated in controlled outcome studies (Huey, Lewine, & Rubenson, 2016; Wong, Gravel, Bouchard, Morselli, & Descormiers, 2012) providing the best evidence for effectiveness. Occasionally, these evaluations produce adverse effects, that is, outcomes that favor control participants. In the gang and crime prevention literatures, these are often interpreted as harmful or iatrogenic, suggesting the interventions worsened participants' offending or gang involvement.

However, determining whether a program is harmful can be challenging. Adverse evaluation effects, with adverse referring to the contrary, opposing or unfavorable direction (Oxford English Dictionary, 2019a), *sometimes* indicate that programs harmed participants or made offending worse; in other cases, interventions are beneficial or neutral, and adverse effects arise from flawed or inappropriate evaluation methods. We present a meta-analytic review of controlled evaluations of gang-focused interventions that aimed to reduce gang involvement or antisocial behavior. We explore whether elements of the intervention programs or research methods increase the likelihood that studies find worse outcomes for treated participants, in other words, adverse effects. Studying predictors of adverse effects may improve the safety of future interventions and can help distinguish harmful effects from methodological errors, so that beneficial or neutral programs are not misjudged as harmful.

2 | BACKGROUND

Scholarship on adverse effects in crime prevention programs has largely focused on non-gang offenders (e.g., Barnett & Howard, 2018; Welsh & Rocque, 2014) or on gang interventions that measure outcomes for whole communities (Braga, 2016), rather than for gang-involved individuals. For example, systematic reviews find that surveillance-based programs and group treatment for juvenile offenders can lead to adverse effects with non-gang offenders (Barnett & Howard, 2018; Welsh & Rocque, 2014). Gang scholars have proposed that programs using street outreach or employing former gang members as intervention providers may be harmful (Braga, 2016; Klein, 2011; Wilson & Chermak, 2011), but these ideas have not been tested in a systematic review. Broadly, we suggest that program evaluations finding adverse effects can suffer from two types of problems: harmful psychological processes that increase participants' antisocial behavior and gang involvement, and methodological challenges with program evaluation that can produce apparently worse outcomes for treated participants, even if the programs are not clearly harmful (Ekblom & Pease, 1995; Petersilia & Turner, 1993; Welsh & Rocque, 2014). Here, we explore each of these issues in more depth.

2.1 | Psychological causes of adverse effects

Deviancy training, increases in gang cohesion, and reactance may be harmful social psychological processes that harm intervention participants, and may increase offending and gang involvement. These can rightly be called iatrogenic, that is, caused unintentionally by the intervention (Oxford English Dictionary, 2019b).

2.1.1 | Deviancy training

Often blamed for adverse effects in programs targeting youth offending (Welsh & Rocque, 2014), deviancy training can occur when an intervention provides opportunities for a group of antisocial and at-risk peers to reinforce each other's antisocial behaviors (Dishion, Spracklen, Andrews, & Patterson, 1996), for example, by laughing along with delinquent acts. In their review of crime prevention programs, Welsh and Rocque (2014) found that group treatment was associated with increased likelihood of producing adverse effects among non-gang offenders, and proposed deviancy training as a possible mechanism. In addition to group treatment modalities, interventions that include recreational activities may also provide opportunities for deviancy training (Gottfredson, 2010; McCord, 2003). For example, in an after-school program for youth at risk for behavior problems, Gottfredson (2010) found that peers reinforced each other's negative behaviors significantly more often during recreation than during other activities. Recreation may involve less structure and supervision than other intervention activities and may promote social interaction between participants who can encourage delinquency.

Other scholars have been critical of the idea that deviancy training causes much harm, arguing that it likely only accounts for a fraction of negative peer influence that youth are exposed to (Weiss et al., 2005), and that deviancy training effects may be overstated (Handwerk, Field, & Friman, 2000). It is not clear whether group-based interventions for gang-involved youth risk promoting deviancy training beyond what youth already experience by spending time in gangs. We examine whether group-based treatment or recreation predict increased odds of adverse effects in gang-focused interventions.

Public Health Significance

Interventions to reduce delinquency and gang involvement should at least do no harm. However, practitioners and researchers must be cautious when interpreting what appear to be harmful effects. Methodological problems with program evaluation can produce outcomes that favor control participants, even if programs are not harmful.

2.1.2 | Gang cohesion

Members' identification with and sense of belonging to gangs, or gang cohesion, may also promote delinquency and offending (Braga, 2016). Certain intervention strategies including streetworker programs (e.g., Braga, 2016; Miller, 1962; Schlossman & Sedlak, 1983; Spergel et al., 2003) may inadvertently increase gang cohesion. Streetworker programs target gang members in public spaces and use multiple strategies to prevent and control gang crime, including police suppression, individual and group counseling, and connecting youth to social services (Spergel, Wa, & Sosa, 2006). Streetworker programs often employ former gang members as providers. Given their familiarity with gang life, former members are presumed to have the necessary "street cred" to build rapport with current gang members and diffuse disputes before they escalate (Klein, 2011; Spergel et al., 2006;

Wilson & Chermak, 2011). However, scholars note that former gang members can also inadvertently glorify gang life when sharing past experiences and strengthen gang identity among participants (Klein, 2011). Former members can also hinder collaborations with police, be reluctant to share information about youth with other providers, and are sometimes themselves arrested for illegal behavior during their employment (Braga, 2016; Klein, 2011; Spergel, Wa, & Sosa, 2005b). Braga (2016) reviewed several streetworker programs that seemed to increase violence among targeted gangs, relative to untreated comparison gangs, and cautioned that even well-implemented streetworker programs may inadvertently strengthen gang identity and cohesion. In this article, we examine whether streetworker programs or former gang members as intervention providers predict adverse effects across studies.

2.1.3 | Reactance

Interventions that include threatened consequences, deterrence elements, or authority figures may sometimes produce the very behaviors they aim to deter. Scared Straight, a classic example of this "backfire" effect, may have inadvertently increased delinquency and arrests among youth participants because it conveyed a threat they did not believe would be carried out (Petrosino, Turpin-Petrosino, & Finckenauer, 2000; Petrosino, Turpin-Petrosino, Hollis-Peel, & Lavenberg, 2013). Scared Straight exposed youth to prisons and graphic descriptions of prison life (Finckenauer, 1982; Petrosino et al., 2013) and was based on the theory that seeing the harsh reality of prison would deter offending (Finckenauer, 1982). Elements of gang interventions that could be seen as threatening or deterrent, such as, perhaps, the presence of law enforcement officers or intensive supervision methods (Petersilia, 1990), may be viewed as symbols of authority to deliberately oppose. In this article, we test whether the presence of law enforcement as intervention providers or intensive supervision is associated with adverse effects in gang interventions. However, these factors also have methodological implications for evaluation outcomes, as discussed below.

2.2 | Methodological causes of adverse effects

Certain research methods may also increase the likelihood of adverse outcomes in program evaluations, regardless of an intervention's true effect on participants. Problems with study design and measurement may make it difficult to discern whether an intervention helped or harmed (Ekblom & Pease, 1995; Rosenbaum, 1986; Welsh & Rocque, 2014). Here, we describe how differential rates of detecting crime across treatment and control groups, assignment to treatment conditions, and measurement validity problems can muddle interpretation of outcomes.

2.2.1 | Differential detection

Assessing outcomes differently for treatment and control groups can alter the likelihood of detecting behaviors of interest. Often, an intervention's structure makes differential detection of outcomes hard to avoid. For example, in intensive supervision programs, treated participants under intensive supervision may be more likely than controls to be caught committing crimes and violations, even if the intervention does not increase the rate of offending (Hyatt & Barnes, 2017; Petersilia, 1990). Controlled evaluations of intensive supervision probation (ISP), a community-based alternative to incarceration or regular probation (Hyatt & Barnes, 2017; Petersilia, 1990), are susceptible to this issue. ISP probationers have more frequent contacts with probation officers, drug testing, and home visits compared to standard probation (Hyatt & Barnes, 2017). The supervision is meant to deter crime and can also increase the likelihood that law enforcement will detect crimes and technical violations for treated participants more than for controls, regardless of any reactance effect (Altschuler, Armstrong, & MacKenzie, 1999; Turner & Petersilia, 1992). For example, in Hyatt and Barnes's (2017) ISP evaluation, underlying rates of offending were the same for ISP and control groups, but officially detected parole violations were significantly higher for ISP clients. Interpretation of group differences in rates of arrests and probation violations becomes challenging when groups are monitored differently.

A similar effect may occur when law enforcement officers serve as intervention providers. With police delivering an intervention, participants might have more contact with police than would untreated controls, giving police more opportunities to observe offenses among intervention participants. Even if actual offending rates are similar for both groups, increased police contacts with intervention participants could result in more arrests for the treatment group. The evidence in this area is mixed; some studies find that law enforcement officers deter crime, reducing both official crime rates and delinquent behavior (e.g., Curry, Decker, & Pyrooz, 2013; Mello, 2019). There is also evidence that police presence increases official offense rates, irrespective of any change in delinquent behavior. In a nationally representative study, Na and Gottfredson (2013) found that police officers stationed in U.S. schools increased rates of recorded misbehaviors and crimes, compared to schools without police officers. They suggest that the observed increases may be due to changes in measurement practices rather than increased offending; police may more readily interpret ambiguous situations as illegal, and more easily detect crime than other school staff (Na & Gottfredson, 2013). In this article, we test whether the presence of law enforcement on intervention provider teams is associated with increased likelihood of detecting adverse effects.

2.2.2 | Unmeasured confounds

Study design can also affect the validity of reported outcomes. Some designs may increase the likelihood of producing significant effects in either direction (i.e., favoring treated or control participants) that do not reflect real program effects.

Randomly assigning individuals to treatment and control groups is considered the gold standard in program evaluation research (Shadish, Cook, & Campbell, 2002). However, randomizing gang members can be challenging because members of one gang may be assigned to different conditions and "contaminate" assignment via regular contact with one another (Spergel, Wa, & Sosa, 2002). Quasi-experimental evaluations or clustered designs in which entire gangs or classrooms are assigned to conditions are reasonable alternatives when randomly assigning individuals is impractical (Shadish et al., 2002). However, in such designs, discerning which observed changes are attributable to an intervention becomes more challenging, mainly because unmeasured confounds cannot be eliminated (Ekblom & Pease, 1995). Baseline differences in recruitment methods, location, time, or level of individual risk can produce differences in outcomes for treatment and control groups, regardless of intervention effects. In programs that draw comparison groups from untreated neighborhoods (e.g., Spergel et al., 2006), events unrelated to the intervention can affect outcomes. For example, local police may increase or decrease patrols in one area and not the other (Ekblom & Pease, 1995). Here, we test whether evaluations that did not randomly assign individuals to treatment conditions, with their added potential for confounds that can favor either treatment or control groups, were more likely to produce adverse effects than designs that randomized individuals.

2.2.3 | Other measurement validity problems

Certain ways studies measure constructs can also introduce bias or validity problems. For example, using follow-up periods of different lengths may allow more opportunities to recidivate for one group (Cohen, Williams, Bekelman, & Crosse, 1995; Williams, Cohen, & Curry, 1999). Likewise, a handful of interventions use the number of arrests or parole violations as an outcome, without accounting for participants' time at risk (i.e., time when they are not incarcerated and can be rearrested, e.g., Spergel et al., 2006). Within-person arrests may not be independent events, since a first arrest may reduce the likelihood of subsequent arrests, if it leads to incarceration. Once incarcerated, individuals have less "opportunity time" to recidivate (Hyatt & Barnes, 2017), so accounting for time at risk is crucial for drawing meaningful conclusions about differences between groups. More severe offending might result in longer incarceration and less opportunity to accrue multiple arrests than less-severe offending. We wondered whether these problems, common to lower-quality studies, increase the likelihood of false positives, that is, adverse or beneficial outcomes that reflect a failure to accurately measure program effects.

2.3 | Current study

In the current study, we examine whether any program characteristics or methodological issues predict adverse outcomes across controlled gang intervention trials. Fundamentally, do adverse effects reflect harmful psychological processes, methodological problems with evaluations, or both? A recent systematic review and meta-analysis (Huey et al., 2016) provides an opportunity to study these questions quantitatively. In their study, Huey and colleagues examine whether gang-focused interventions effectively reduce gang involvement or antisocial behavior. In this article, we code the same program evaluations for content, providers, research methods, and outcomes and then assess whether any variables related to the harmful psychological processes or methodological issues described in the prior sections predicted adverse effects.

2.3.1 | Hypotheses

If deviancy training or increases in gang cohesion are driving adverse outcomes across studies, we would expect that streetworker programs, or including former gang members as providers, group treatment modalities, or recreation activities, would predict adverse outcomes. If adverse outcomes reflect methodological problems with research design and measurement, we would expect the presence of clear measurement validity problems or nonrandom assignment of individuals to treatment conditions to be significant predictors. Finally, if the presence of law enforcement officers or intensive supervision of treated participants is significant predictors of adverse outcomes, these may indicate either a methodological issue, a harmful psychological effect like reactance, or both.

3 | METHOD

3.1 | Literature search

An online literature search was conducted using the following databases: ERIC, Library and Information Science Abstracts (LISA), ProQuest Dissertations and Theses, ProQuest Research Library, PsycARTICLES, PsycCRITIQUES, PsyclNFO, Social Services Abstracts, and Sociological Abstracts. Search terms reflected controlled evaluations (e.g., *controlled, random*), interventions (e.g., *treatment, prevention, intervention*), and gang involvement (e.g., *gangs*). We supplemented our online literature search using references listed in gang-focused meta-analyses and reviews (e.g., Esbensen, 2000; Wong et al., 2012) and included unpublished, in press, and published studies recommended by gang researchers. Two hundred and nineteen titles and abstracts were screened for eligibility (Huey et al., 2016).

3.2 | Inclusion criteria

Eligible studies included (a) either predominantly gang-involved participants, participants who were described as at risk for joining gangs, or any gang-involved participants (even if only a minority of the study population) if the study included separate outcome data for the gang-involved participants; (b) a control or comparison group; and (c) posttreatment or follow-up outcomes for individual participants' gang involvement or antisocial behavior. Studies were excluded if they reported outcomes at the neighborhood or community level only, rather than for individuals, or if they were not published or written in English.

3.3 | Study-level coding

We coded each study at two levels: (a) study-level characteristics we hypothesized were related to adverse effects, and (b) study outcomes as favoring treatment, controls, or neither. We sought to test whether study-level characteristics predicted adverse effects.

All variables were coded by consensus between the first, second, and third authors. The first author coded all studies for all variables. The second author coded intensive supervision and measurement validity problems, and all other variables were coded by the third author.

We coded studies for the presence of variables related to harmful psychological processes (i.e., gang cohesion, deviancy training, and reactance), and to methodological problems with evaluations, including street outreach, former gang members as intervention providers, group treatment modality, inclusion of recreation activities, more intensive supervision of the treatment group than the comparison group, law enforcement officers as intervention providers, random assignment of individuals to treatment conditions, and any clear measurement validity problems. To code for these variables, we carefully read method sections and program descriptions and searched for terms related to these variables (e.g., *streetworker, street outreach worker, youth worker*, etc.).

The measurement validity problems code included studies that used outcome measures that differed by either the frequency, instrument, or method across treatment and control groups, as well as any study that used arrests or other nonindependent outcome measures without accounting for time at risk. For example, studies that used a longer follow-up period for one group (e.g., Cohen et al., 1995), studies that asked interview questions differently across groups (e.g., Williams et al., 1999), and studies that used the number of arrests for participants as an outcome measure without accounting for time at risk (Gold & Mattick, 1974; Spergel et al., 2003, 2005b; Spergel, Wa, & Sosa, 2005a) were coded as having clear measurement validity problems. Study-level predictor variables are displayed in Table 1.

3.4 | Outcome-level coding

All outcomes were coded by consensus between the first and second authors; for each study, we listed all reported antisocial behavior and gang involvement outcomes. We defined antisocial behavior as justice system involvement (e.g., arrests, incarceration, time to first probation violation, court appearances, parole violations, etc.) and self- or observer-reported delinquent, illegal, or overtly aggressive behaviors (e.g., scores on delinquency scales, illegal substance use). We defined gang-involvement outcomes as any individual behavior related to gangs (e.g., self-reported membership, wearing gang colors, involvement in gang fights). We included only outcomes that were compared across treated and untreated groups; results of moderation and mediation analyses or changes within the treatment group across time were excluded. We also excluded outcomes for neighborhoods and or other geographical areas, such as police reports of gang size changes, and outcomes that were clearly not antisocial behavior or gang involvement (e.g., school performance, clinical diagnoses).

TABLE 1 Variables tested as predictors of adverse outcomes

Group intervention modality
Recreation component
Streetworker program
Intensive supervision
Law enforcement officer providers
Any measurement validity problems
Nonrandom assignment

We coded outcomes based on statistical significance and direction of effect. Unfortunately, most studies did not provide complete reports of statistical analyses; often, the type of statistical tests, *p*-values, means, standard deviations, or sample sizes were missing (e.g., Gold & Mattick, 1974; Schlossman & Sedlak, 1983). Several studies reported only proportions for treatment and control groups (e.g., proportion arrested) without statistical tests (e.g., Miller, 1962; Reckless & Dinitz, 1972; Willman & Snortum, 1982). Some studies reported that results were statistically significant, but included no details of statistical tests. Furthermore, reporting tended to be less clear for older studies and particularly for those with adverse effects (1979), which limited our ability to explore our research questions in a traditional meta-analysis using effect sizes as the data needed to calculate effect size were missing for a substantial proportion of outcomes. As such, we chose to analyze effects by reports of statistical significance and direction, rather than effect size, which allowed for the largest number of studies to be included in our analyses, maximizing both our statistical power and the validity of data available to explore our research questions.

All outcomes were coded as either significantly favoring treated participants, significantly favoring controls, or not significant. We defined adverse effects as any antisocial behavior or gang-involvement outcome that significantly favored the control group; however, study authors reported them. This differed across studies; for example, some studies reported arrest outcomes in the form of changes in arrest rates from pretreatment to follow-up, in which case a relative increase for the treatment group was an adverse outcome (e.g., Spergel et al., 2005a). Others reported proportions of the total sample arrested at follow-up, in which case a larger proportion for the treatment group was an adverse outcome (e.g., Willman & Snortum, 1982). To determine significance, we used an alpha level of 0.05, or relied on authors' reports that comparisons were "significant" or "statistically significant" if *p*-values were not reported. For any outcome reported as "not significant" or with a p-value $> .05$, we coded the outcome as not significant. For studies with no mention of statistical tests for relevant outcomes but with sufficient data to determine effect size (i.e., sample sizes, means, and standards deviations or proportions), we calculated Hedge's g and the corresponding *p*-values using Comprehensive Meta-Analysis (CMA; Borenstein, Hedges, Higgins, & Rothstein, 2005). The total number of outcomes per study and their codes were determined by consensus between the first and second authors.

4 | ANALYSES

We performed two sets of analyses to explore whether study-level characteristics were related to adverse effects in controlled gang intervention trials. The first tested these variables at the study level only; each study was coded as either reporting one or more adverse effects or not. The second set of analyses accounted for the number of outcomes reported in each study.

First, we ran two-tailed Fisher's exact tests to determine whether reporting one or more statistically significant adverse outcomes depended on the presence of any of the variables in Table 1. Fisher's exact test computes the probability of obtaining an observed contingency table (or a more extreme one) by chance alone for small samples, given expected frequencies for each cell (Agresti, 1992; Fisher, 1922). Data were grouped in two-by-two contingency tables according to adverse outcomes (i.e., present or not present) and the given predictor variable (e.g., intervention included recreation or not), for a total of eight Fisher's exact tests. We used IBM SPSS Statistics version 25 to analyze the data.

Next, to account for multiple outcomes per study, we evaluated study-level variables as predictors of adverse effects using generalized linear mixed effects regressions (GLMER) for repeated measures using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) for R version 3.6.0 (R Core Team, 2018). GLMER is a mixed effects model for categorical outcomes that makes similar assumptions as general linear mixed effects models for continuous models (Singer & Willett, 2003). In this set of analyses, we dichotomized each outcome as adverse or not. GLMER models assume the outcome is binomially distributed, and the logit link is used to relate all predictors to the conditional means of the adverse responses. GLMER analysis accounts for correlations between binary outcome variables nested within studies (Fitzmaurice, Laird, & Ware, 2011). We first ran separate GLMER models each with a single predictor to test the bivariate correlation between each predictor and occurrence of adverse outcomes. The models are specified as follows:

$$IE_{ij} = e^{(\gamma 00 + u0j + \beta 1Xj)}$$

In this model, IE represents the i^{th} adverse outcome in the j^{th} study. The right-hand side of the equation exponentiates the parameters in the model, as the logit link function estimates parameters in logit units. Exponentiated values are interpreted as odds ratios (OR) where values that do not significantly differ from 1 indicate equal risk of the study variables whereas OR values significantly greater than 1 indicate that the study variable increases risk of adverse outcomes. γ_{00} represents the exponentiated mean logit intercept when all predictors are zero; u_{0j} indicates the individual study deviation around the mean logit intercept. β_{1} represents the exponentiated effect of each study predictor variable, X_{j}. We note that generalized linear models lack residual scores (Singer & Willett, 2003), so no residual appears in the above expression.

We then fit multivariate models to identify whether significant predictors in the bivariate models remain significant adjusting for other predictor variables. All models were fit using maximum likelihood, and the likelihood ratio test is used to compare nested models.

5 | RESULTS

5.1 | Descriptive statistics

Thirty-eight documents including 41 evaluations met inclusion criteria. Across studies, target populations and intervention strategies varied widely; some were prevention programs targeting gang-joining in middle school children; others were treatment programs for convicted gang offenders. Intervention strategies included suppression services delivered by law enforcement, support groups, cultural programming for indigenous youth, and family therapy, among others. For brief descriptions of each study, including sample sizes, demographics, and core outcomes, see Huey et al. (2016).[1]

Table 2 shows the distribution of adverse, apparently beneficial, null, and missing outcomes across studies. A total of 560 antisocial behavior and gang involvement outcomes were reported across all studies, with between 1 and 44 outcomes per study. One hundred and six represented significant beneficial effects of treatment, 15 were significant adverse effects, 401 were nonsignificant, and 38 were uninterpretable or missing.[2] Eight out of the 41 evaluations reported one or more statistically significant adverse effects, all of which were related to antisocial behavior, and included arrests, institutional misconduct, incarceration time, and self-reported illegal activities; none was related to gang involvement (Table 3). Two studies reported no interpretable outcomes and were excluded from subsequent analyses.

Table 4 shows a correlation matrix for the eight study-level predictor variables and adverse effects. Correlations with adverse effects were close to zero; however, collinearity was high among other predictors, particularly those that were conceptually related. For example, presence of law enforcement officers was correlated with intensive supervision at $r = .61$, likely because intensive supervision was typically provided by police and probation officers.

TABLE 2 Distribution of outcomes across studies

Outcome	*N*	%
Beneficial	106	18.9
Null	401	71.6
Adverse	15	2.6
Missing	38	6.8
Total	560	100.0

5.2 | Fisher's exact tests

We ran Fisher's exact tests as exploratory analyses to examine the distribution of our predictor variables across the 39 studies with one or more interpretable outcomes for antisocial behavior or gang involvement. We tested whether there were any associations at the study level between our predictor variables and reporting one or more adverse effects.

Fisher's exact probabilities for observed contingency tables are displayed in Table 5, in ascending order. Adverse effects were significantly related to the presence of law enforcement officers on the provider team ($p = .049$). Seven out of eight studies with adverse outcomes included police or probation officers as providers. We observed no significant relationships between adverse effects and random assignment ($p = .682$), nor to intensive supervision of one group ($p = .090$), nor to variables related to deviancy training or gang cohesion (p-values ranged from .355 to .695). Interventions reporting one or more adverse outcomes were no more likely to use streetworkers or former gang members as providers, be delivered in groups, or include recreation than studies with only apparently beneficial or null outcomes.

5.3 | Generalized linear mixed effects regressions

Table 6 shows the results of GLMER analyses with a single predictor variable in each. Presence of law enforcement officers on the provider team was significantly associated with increased likelihood of reporting one or more adverse effects ($\beta = 2.54$, $OR = 12.64$, $p = .029$). Thus, we fit a multivariate GLMER model to estimate whether presence of law

TABLE 3 Gang-focused interventions reporting one or more adverse outcomes

Study	Adverse Outcome(s)
Agopian (1990)	% incarcerated
Peters et al. (1996)	Time to recidivate
Spergel et al. (2002; Mesa)	Property arrest changes
Spergel et al. (2005b; Riverside)	% with drug arrests
Spergel et al. (2005b; Bloomington-Normal)	Total arrests; violence arrests; property arrests; other arrests; drug arrests
Willman and Snortum (1982)	Total crimes against persons; total robberies
Wiebush et al. (2005; Las Vegas)	% with major institutional misconduct
Wodarski et al. (1979)	Engaging in aggressive behavior to achieve goals

TABLE 4 Correlations Among Study-Level Predictor Variables and Adverse Effects

	Adverse effects	Law enforcement	Intensive supervision	Group modality	Streetworker program	Former gang members	Random assignment	Measurement validity problems	Recreation component
Adverse effects	1	0.13	0.09	−0.05	−0.02	−0.01	−0.02	0.07	0.06
Law enforcement	0.13	1	0.61	0.12	0.37	0.37	−0.19	0.53	0.48
Intensive supervision	0.09	0.61	1	0.17	0.60	0.50	−0.13	0.93	0.59
Group modality	−0.05	0.12	0.17	1	0.08	0.13	−0.03	0.08	0.30
Streetworker program	−0.02	0.37	0.6	0.08	1	0.54	−0.03	0.63	0.49
Former gang members	−0.01	0.37	0.5	0.13	0.54	1	−0.34	0.46	0.33
Random assignment	−0.02	−0.19	−0.13	−0.03	−0.03	−0.34	1	−0.18	−0.19
Measurement validity problems	0.07	0.53	0.93	0.08	0.63	0.46	−0.18	1	0.58
Recreation component	0.06	0.48	0.59	0.30	0.49	0.33	−0.19	0.58	1

TABLE 5 Results of 2 × 2 contingency tables with adverse outcomes

Variable	Two-Tailed Fisher's Exact Probability (p)
Law enforcement officer providers	**.049**
Intensive supervision	.090
Any measurement validity problems	.220
Streetworker program	.355
Former gang member providers	.587
Group treatment modality	.617
Nonrandom assignment	.682
Recreation component	.695

Note: Bolded values indicate statistically significant associations (p < .05).

enforcement remained a significant predictor of adverse outcomes, adjusting for other covariates.

We next ran multivariate GLMER models using all predictors. As presence of law enforcement was highly correlated with intensive supervision ($r = .61$), recreation ($r = .48$), and any measurement validity problems ($r = .53$), these three variables were removed from the model, as the predictive utility of these variables is redundant with the predictive utility of law enforcement and leads to statistical suppression of the effect of law enforcement. Results are presented in Table 7. The multivariate model did not perform significantly better than the bivariate model that included only presence of law enforcement (likelihood ratio test = 1.40, $df = 4$, $p = .844$), as expected. We found that the presence of law enforcement officers remained significant, with an odds increase of adverse effects equal to 21.28, indicating that presence of law enforcement officers on the provider team increased the risk.

6 | DISCUSSION

Despite increasing attention to potentially harmful effects from crime prevention programs (Barnett & Howard, 2018; Braga, 2016; McCord, 2003; Welsh & Rocque, 2014), effects on individuals in gang interventions had not been studied systematically. We find that adverse effects in evaluations of gang interventions are related to the presence of law enforcement officers as providers. However, we do not interpret this to mean that police and probation officers necessarily increase offending. Our finding suggests a few possible explanations; increased crime *detection* for treatment groups may contribute to adverse effects, and law enforcement officers may inspire reactance in gang-involved or at-risk participants who want to oppose authority by committing more delinquent acts, or both. Given the limitations of this meta-analysis, we could not determine whether methodological issues with evaluation or psychological processes caused adverse

TABLE 6 Results of single-predictor GLMER models

Predictor Variable	$\sigma^2_{\text{Intercept}}$	β	OR	*SE*	*z*-value	*p*
Intensive supervision	2.156	1.368	3.927	0.938	1.458	.145
Law enforcement providers	**1.139**	**2.537**	**12.642**	**1.165**	**2.178**	**.029**
Random assignment	2.377	−0.189	0.828	0.966	−0.196	.845
Group modality	2.268	−1.110	0.330	1.080	−1.028	.304
Recreation		0.535	1.707	0.909	0.589	.556
Streetworker program	2.509	0.199	1.220	1.107	0.179	.858
Former gang member providers	2.491	0.321	1.378	1.243	0.258	.796
Any measurement validity problems	2.208	1.116	3.053	0.917	1.217	.224

Note: Bolded values indicate statistically significant parameters ($p < .05$).

outcomes, or whether programs that reported adverse effects were in fact harmful.

Our results may support suppression theory, in that increased crime detection may increase documented offense rates, regardless of actual changes in delinquent behavior (Na & Gottfredson, 2013; Petersilia & Turner, 1993). Police and probation officers may be more likely to detect offenses than other types of intervention providers. In contrast, crimes committed by control participants would only be detected through normal police operations, rather than through additional contact with law enforcement during an intervention. Our results are consistent with evaluations of ISP (Hyatt & Barnes, 2017; Petersilia & Turner, 1991, 1993) and of school-based police officer programs (Na & Gottfredson, 2013), where rates of minor and technical offenses in particular are higher when police and probation officers are present. If law enforcement officers detected more crime among active treatment groups, we should not assume that interventions reporting adverse effects worsened participants' behavior.

Alternatively, the presence of law enforcement could have provoked a harmful oppositional reaction in youth, if they wanted to demonstrate that they were not intimidated by law enforcement, for example. This backfiring effect has been suggested as an explanation for the failure of Scared Straight (Finckenauer, 1982; Petrosino et al., 2013), which more often increased, rather than decreased, rates of future incarceration for participants relative to controls (Petrosino et al., 2000). That program may have harmed youth by inadvertently challenging them to prove they were not scared, increasing their motivation for delinquency (Finckenauer, 1982).

The presence of police and probation officers in gang interventions may have inspired similar reactance, but it is unknown whether the association with adverse effects reflects methodological, psychological, or both problems. Evaluations in our study that reported adverse effects tended to include intensive supervision by law enforcement (e.g., Agopian, 1990; Wiebush, Wagner, Mcnulty, Wang, & Le, 2005) and did not emerge in trials of G.R.E.A.T., a classroom-based gang prevention program led by uniformed police officers (Esbensen, Peterson, Taylor, & Osgood, 2012; Esbensen & Wayne Osgood, 1999). Future research could examine how the presence of law enforcement affects program participants psychologically. Data on how many offenses are detected by law enforcement officers during interventions versus during normal police operations, or measures of participants' attitudes toward police, could help distinguish methodological and psychological effects on antisocial behavior.

The lack of evidence for deviancy training in our data is somewhat surprising given prior research on apparently harmful and ineffective crime prevention programs. Here, we find that group treatment and recreation activities were unrelated to adverse effects in gang-focused programs. Our findings raise the question of whether deviancy training has potent effects among youth who are already gang-involved. Perhaps gang youth in group treatments are less susceptible to additional deviancy training than non-gang youth, since reinforcement from deviant peers during an intervention may be negligible compared to the reinforcement gang members already receive simply by spending time in gangs (Weiss et al., 2005). Several of the interventions in our review were delivered to groups of delinquent gang youth (e.g., Josi & Sechrest, 1999; Spergel et al., 2003), or were primary prevention programs delivered in schools (e.g., Esbensen et al., 2012), and perhaps neither type provided much additional opportunity for deviancy training beyond what youth already experienced in gangs or in school. Although group treatment for delinquent or at-risk youth predicts adverse effects in some studies (e.g., Welsh & Rocque, 2014), risk for deviancy training may depend on youths' daily context (Weiss et

TABLE 7 Results of multivariate GLMER models

	Model 1		Model 2	
Fixed Effects	**β**	***SE***	**β**	***SE***
Intercept	**−5.98**	**1.18**	**−5.14**	**1.36**
Law enforcement	**2.54**	**1.17**	**3.06**	**1.26**
Random assignment	—	—	0.63	0.99
Group treatment	—	—	−1.44	1.04
Streetworker program	—	—	−1.02	1.08
Former gang member	—	—	0.52	1.33
Random effects				
$\sigma^2_{\text{Intercept}}$	1.14	—	0.80	—
Model fit				
Log likelihood (*df*)	−60.50 (519)		−59.10 (515)	

Note: Bolded values indicate statistically significant parameters (p < .05).

al., 2005). Future studies could compare deviancy training effects across gang and non-gang youth.

Deviancy training may also depend on the type or format of group treatment. Unfortunately, too few studies in our sample specified details about group treatments that would have allowed for a more rigorous analysis of deviancy training opportunities. Group treatment varied widely across studies in our sample, and different formats may facilitate deviancy training to differing degrees (e.g., group skills training vs. group camping trips). A more fine-grained coding of group treatments in future studies could help determine whether any formats are more likely to be associated with adverse effects.

We also found no evidence that streetworker programs were more likely than other intervention strategies to be harmful for individuals, despite scholarly concerns that such programs risk increasing gang cohesion and area-level crime rates (Braga, 2016; Klein, 2011). Our analyses also did not support concerns about the use of former gang members as intervention providers (Klein, 2011) relative to any other provider type. However, independent effects of streetworker programs may have been masked in our analysis; these programs tended to include recreation, intensive supervision, and group treatment, employ law enforcement officers and former gang members as providers, and be evaluated using quasi-experimental designs (Miller, 1962; Schlossman & Sedlak, 1983; Spergel et al., 2006). A few older streetworker programs also vaguely described what may have been adverse effects but lacked sufficient data for analysis (e.g., Gold & Mattick, 1974). Harmful effects of streetworker programs may not have been detectable in our analysis. Our findings should not be taken as endorsements of these practices, but rather as a call for closer examination of streetworker program effects.

Type 1 error, or false positives, may explain some of our findings. Several studies in our review ran multiple statistical tests without clearly accounting for the type 1 error rate, which would have increased the rate of false-positive results in either direction (i.e., adverse or beneficial; Benjamini & Hochberg, 1995). This was particularly the case in unpublished reports and older journal articles (e.g., Arbreton & McClanahan, 2002; Spergel et al., 2006; Wodarski, Filipczak, McCombs, Koustenis, & Rusilko, 1979). For example, Wodarski and colleagues (1979) ran separate *t* tests on every item of a self-report questionnaire, and several of the Spergel Model program evaluations reported dozens of statistical comparisons between treatment and control groups (e.g., Spergel et al., 2005a). The 15 adverse effects out of 560 total effects across all studies, or 2.68%, are similar to the 2.5% expected by random chance. However, we would not expect truly random effects to be associated with law enforcement presence, which suggests our finding reflects differential detection of crime, reactance, or both, rather than type-1 error alone.

6.1 | Limitations

Our analyses were limited to a focus on statistically significant effects, rather than effect sizes. Several studies in our sample did not report sufficient data to calculate effect sizes, and those also tended to be the studies that reported adverse effects (e.g., Wodarski et al., 1979), so exploring our research questions in a traditional meta-analysis would have limited validity. This focus on significant effects may have masked the true incidence of adverse effects in small-sample studies, which lacked power to produce significance in either direction (e.g., Dole, 2005; Huey, McDaniel, Smith, Pearson, & Griffin, 2014). As a result, our study lacks some of the detail that a traditional meta-analysis could provide. For example, we could not test whether study-level characteristics like group treatment or law enforcement presence reduced effect sizes overall, even among studies that found no adverse effects.

Our analyses were limited by low statistical power, because our literature search found only 41 controlled gang-focused interventions with outcomes for individuals. Other study or program characteristics may indeed be related to adverse outcomes but did not emerge significant due to low power. Publication bias may have contributed to the problem. Given the tendency for researchers to favor publishing significant positive findings over null and adverse findings

(Levine, Asada, & Carpenter, 2009; McCord, 2003), our review may underestimate the extent to which adverse effects are present in this literature.

Our analyses were also limited by multicollinearity, which made it difficult to tease apart independent effects of each predictor. Several variables were closely related in the delivery of the interventions; police and parole officers were providers of most interventions with intensive supervision, and former gang members were providers mostly in streetworker programs. Streetworker programs in particular added to the multicollinearity problem by including multiple modalities and provider types. These program elements may have had different impacts on participant outcomes, but were not separately detectable in our analyses.

We were also unable to systematically examine implementation failure, another potential cause of adverse effects (Ekblom & Pease, 1995; Welsh & Rocque, 2014). Implementation failure occurs when an intervention is not delivered as intended, even if its theory is sound (Ekblom & Pease, 1995). Staff turnover, budget problems, and conflicts with partner agencies, for example, can lead to iatrogenic effects if treatment is chaotic and disorganized, or the delivery undermines the original intent of the intervention (Welsh & Rocque, 2014). Evaluations in our review discussed implementation to varying degrees; some studies measured treatment fidelity, tracked worker contacts with participants (e.g., Agopian, 1990; Miller, 1962; Spergel et al., 2002), and even calculated how implementation issues contributed to outcomes (e.g., Spergel et al., 2003; Wiebush et al., 2005). However, most studies did not mention implementation quality at all (e.g., Cohen et al., 1995; Thompson & Jason, 1988), so we chose not to test it as a predictor, rather than exclude the majority of studies or make assumptions about studies that failed to discuss it. A qualitative review of the core theories, implementation strengths and weaknesses, and evaluation methods in studies that reported adverse outcomes could help inform what caused them and whether interventions were harmful (Rubenson, Galbraith, & Huey, in revision).

Finally, while law enforcement presence significantly predicted adverse outcomes across studies, our findings are purely correlational and should be interpreted with caution. Deviancy training and increases in gang cohesion, for example, could have caused some of the adverse outcomes in these studies and were not detected in our analyses.

6.2 | Implications

We found that interventions using law enforcement officers as providers were more likely to report adverse outcomes than other interventions. When programs aim to reduce gang membership and antisocial behavior, police and probation officers may be counterproductive, especially when outcomes are measured by indicators like police contacts and arrests. However, our results are correlational and it remains unclear why this relationship emerged. Additionally, whether this finding seems concerning depends on an intervention's goals; if the goal is to detect more crime and bring offenders to justice (Petersilia, 1990), increases in arrest rates among treated participants may not be considered adverse. Future evaluations should include corroborating measures for the same constructs, such as self- and observer reports along with official reports of delinquency, particularly when there are likely to be different rates of detecting crime across treatment and control groups (Petersilia & Turner, 1993).

Overall, adverse outcomes in program evaluations should not be called "harmful" without careful consideration. Understanding the difference between harm and adverse effects due to methodological problems is critical when trying to address serious issues like delinquency and gang violence. Harmful practices should be modified or eliminated (Welsh & Rocque, 2014), and evaluation methods prone to validity problems should be avoided or used with caution. Potentially helpful interventions should not be discarded because researchers simply failed to measure them effectively.

Some studies in our sample inaccurately refer to their own findings as harmful when undesirable behaviors increased for treated participants, even when the causes were unclear. For example, Miller (1962) concluded that the Total Community Gang Control Project increased delinquency. Although it appeared to increase for treated gangs between baseline and follow-up, delinquency was only measured for the treatment group. For the only measure compared to controls, the program appeared to have no effect; court appearance rates were nearly identical across groups. Without control groups, single-group effects provide no information about the effectiveness of an intervention (McCord, 2003). Offending in particular has predictable increases from adolescence into early adulthood and then declines (DeLisi, 2015; Hirschi & Gottfredson, 1983). When calling a program "harmful," it should be clear that it made participants worse—more criminal or more gang-involved—than they would be without the intervention and that changes are not due to some other cause. If a program is not clearly harmful, and adverse effects appear to be caused by differential detection of offending or evaluation design issues, the program might be worth reexamining with different methods.

This research highlights the continued need for rigorous evaluations of gang and violence reduction programs. The prevalence of studies with adverse outcomes in our sample (20% of studies) suggests that McCord's (2003) call for monitoring the safety of interventions and the need to do no harm is still relevant, even though adverse effects require nuanced interpretation.

ORCID

Miriam P. Rubenson https://orcid.org/0000-0002-9013-4971

Katharine Galbraith https://orcid.org/0000-0002-7054-827X

ENDNOTES

[1] Three studies were found when we updated the literature search for this paper: Weinrath et al. (2016), Cook, Kang, Braga, Ludwig, & O'Brien (2015), and Minnis et al. (2014).

[2] These include outcomes that were reportedly measured, but no results were reported, and studies that reported only raw data or group means without sufficient information to compute significance.

REFERENCES

Agopian, M. W. (1990). The impact of intensive supervision probation on gang-drug offenders. *Criminal Justice Policy Review*, *4*(3), 214–222. https://doi.org/10.1177/088740349000400302

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, *7*(1), 131–153. https://doi.org/10.1214/ss/1177011454

Altschuler, D. M., Armstrong, T. L., & MacKenzie, D. L. (1999). *Reintegration, supervised release, and intensive aftercare*. Washington, DC: U.S. Department of Justice.

Arbreton, A. J. A., & McClanahan, W. S. (2002). *Targeted outreach: Boys & Girls Clubs of America's approach to gang prevention and intervention. Public/Private Ventures*. Retrieved from https://eric.ed.gov/?xml:id=ED466518

Barnett, G. D., & Howard, F. F. (2018). What doesn't work to reduce reoffending? A review of reviews of ineffective interventions for adults convicted of crimes. *European Psychologist*, *23*(2), 111–129. https://doi.org/10.1027/1016-9040/a000323

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis, version 3*. Englewood, NJ: Biostat.

Braga, A. A. (2016). The continued importance of measuring potentially harmful impacts of crime prevention programs: The academy of experimental criminology 2014 Joan McCord lecture. *Journal of Experimental Criminology*, *12*, 1–20. https://doi.org/10.1007/s11292-016-9252-4

Cohen, M. I., Williams, K., Bekelman, A. M., & Crosse, S. (1995). Evaluation of the national youth gang drug prevention program. In M. W. Klein & C. L. Maxson (Eds.), *Modern gang reader* (pp. 266–275). Los Angeles, CA: Roxbury.

Curry, D. G., Decker, S. H., & Pyrooz, D. C. (2013). *Confronting gangs: Crime and Community*. New York, NY: Oxford University Press.

DeLisi, M. (2015). Age–crime curve and criminal career patterns. In J. Morizot & L. Kazemian (Eds.), *The development of criminal and antisocial behavior* (pp. 51–63). London, UK: Springer.

Dishion, T. J., Spracklen, K. M., Andrews, D. W., & Patterson, G. R. (1996). Deviancy training in male adolescent friendships. *Behavior Therapy*, *27*(3), 373–390. https://doi.org/10.1016/S0005-7894(96)80023-2

Dole, K. (2005). *The effect of the Balance Program on aggression in the classroom*. San Francisco, CA: University of San Francisco.

Egley, A., & Howell, J. C. (2014). *Highlights of the 2010 national youth gang survey*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Ekblom, P., & Pease, K. (1995). Evaluating crime prevention. *Crime and Justice*, *19*, 585–662. https://doi.org/10.1086/449238

Esbensen, F.-A. (2000). *Preventing adolescent gang involvement*. Washington, DC: US Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

Esbensen, F.-A., & Huizinga, D. (1993). Gangs, drugs, and delinquency in a survey of urban youth. *Criminology*, *31*(4), 565–589. https://doi.org/10.1111/j.1745-9125.1993.tb01142.x

Esbensen, F.-A., Peterson, D., Taylor, T. J., & Osgood, D. W. (2012). Results from a multi-site evaluation of the G.R.E.A.T. program. *Justice Quarterly*, *29*(1), 125–151. https://doi.org/10.1080/07418825.2011.585995

Esbensen, F.-A., & Wayne Osgood, D. (1999). Gang Resistance Education and Training (GREAT): Results from the national evaluation. *Journal of Research in Crime and Delinquency*, *36*(2), 194–225. https://doi.org/10.1177/0022427899036002004

Finckenauer, J. O. (1982). *Scared straight! and the panacea phenomenon*. Edgewood Cliffs, NJ: Prentice-Hall.

Fisher, R. A. (1922). On the interpretation of $\chi 2$ from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, *85*(1), 87–94. https://doi.org/10.2307/2340521

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Gatti, U., Tremblay, R. E., Vitaro, F., & McDuff, P. (2005). Youth gangs, delinquency and drug use: A test of the selection, facilitation, and enhancement hypotheses. *Journal of Child Psychology and Psychiatry*, *46*(11), 1178–1190. https://doi.org/10.1111/j.1469-7610.2005.00423.x

Gold, M., & Mattick, H. W. (1974). *Experiment in the streets: The Chicago youth development project*. Ann Arbor, MI: Institute for Social Research.

Gottfredson, D. C. (2010). Deviancy training: Understanding how preventive interventions harm: The Academy of Experimental Criminology 2009 Joan McCord award lecture. *Journal of Experimental Criminology*, *6*(3), 229–243. https://doi.org/10.1007/s11292-010-9101-9

Handwerk, M. L., Field, C. E., & Friman, P. C. (2000). The iatrogenic effects of group intervention for antisocial youth: Premature extrapolations? *Journal of Behavioral Education*, *10*(4), 223–238. https://doi.org/10.1023/A:1012299716053

Hirschi, T., & Gottfredson, M. (1983). Age and the explanation of crime. *American Journal of Sociology*, *89*(3), 552–584. https://doi.org/10.1086/227905

Huey, S. J. J., Lewine, G., & Rubenson, M. (2016). A brief review and meta-analysis of gang intervention trials in North America. In C. L. Maxson & F.-A. Esbensen (Eds.), *Gang transitions and transformations in an international context* (pp. 217–233). New York, NY: Springer.

Huey, S. J. J., McDaniel, D. D., Smith, C. A., Pearson, C., & Griffin, J. P. J. (2014). Gang-involved African American youth: An overview. In K. C. Vaughans & W. Spielberg (Eds.), *The psychology of Black boys and adolescents* (Vols. *1–2*, pp. 313–333). Santa Barbara, CA: Praeger.

Hyatt, J. M., & Barnes, G. C. (2017). An experimental evaluation of the impact of intensive supervision on the recidivism of high-risk probationers. *Crime and Delinquency*, *63*(1), 3–38. https://doi.org/10.1177/0011128714555757

Josi, D. A., & Sechrest, D. K. (1999). A pragmatic approach to parole aftercare: Evaluation of a community reintegration program for high-risk youthful offenders. *Justice Quarterly*, *16*(1), 51–80. https://doi.org/10.1080/07418829900094051

Klein, M. W. (2011). Comprehensive gang and violence reduction programs: Reinventing the square wheel. *Criminology & Public Policy*, *10*(4), 1037–1044. https://doi.org/10.1111/j.1745-9133.2011.00758.x

Krohn, M. D., & Thornberry, T. P. (2008). Longitudinal perspectives on adolescent street gangs. In M. D. Krohn & T. P. Thornberry (Eds.), *The long view of crime: A synthesis of longitudinal research* (pp. 128–160). New York, NY: Springer New York.

Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against nonsignificant findings. *Communication Monographs*, *76*(3), 286–302. https://doi.org/10.1080/03637750903074685

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. https://doi.org/10.1177/0002716202250781

Mello, S. (2019). More COPS, less crime. *Journal of Public Economics*, *172*, 174–220. https://doi.org/10.1016/j.jpubeco.2018.12.003

Miller, W. B. (1962). The impact of a "Total-Community" delinquency control project. *Social Problems*, *10*(2), 168–191. https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=148501

Na, C., & Gottfredson, D. C. (2013). Police officers in schools: Effects on school crime and the processing of offending behaviors. *Justice Quarterly*, *30*(4), 619–650. https://doi.org/10.1080/07418825.2011.615754

Oxford English Dictionary. (2019a). Adverse, adj. and n. Retrieved from https://www-oed-com.libproxy2.usc.edu/view/Entry/2960?rskey=iP426x&result=1#eid

Oxford English Dictionary. (2019b). Iatrogenic, adj. Retrieved from https://www-oed-com.libproxy2.usc.edu/view/Entry/90705?redirectedFrom=iatrogenic#eid

Peters, M., Bullman, S., Gimbel, C., Thomas, D., Laxton, G., Opanga, M., … Croan, G. (1996). *Evaluation of the impact of boot camps for juvenile offenders*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Petersilia, J. (1990). When probation becomes more dreaded than prison. *Federal Probation*, *23*, 23–27. https://psycnet.apa.org/record/1990-22862-001

Petersilia, J., & Turner, S. (1991). An evaluation of intensive probation in California. *The Journal of Criminal Law and Criminology*, *82*(3), 610–658. https://doi.org/10.2307/1143747

Petersilia, J., & Turner, S. (1993). Intensive probation and parole. *Crime and Justice*, *17*, 281–335. https://doi.org/10.1086/449215

Petrosino, A., Turpin-Petrosino, C., & Finckenauer, J. O. (2000). Well-meaning programs can have harmful effects! Lessons from experiments of programs such as Scared Straight. *Crime & Delinquency*, *46*(3), 354–379. https://doi.org/10.1177/0011128700046003006

Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M. E., & Lavenberg, J. G. (2013). "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency. *Cochrane Database of Systematic Reviews*, *5*, 1–55. https://doi.org/10.1002/14651858.CD002796.pub2

Pyrooz, D. C., Turanovic, J. J., Decker, S. H., & Wu, J. (2016). Taking stock of the relationship between gang membership and offending: A meta-analysis. *Criminal Justice and Behavior*, *43*(3), 365–397. https://doi.org/10.1177/0093854815605528

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Reckless, W. C., & Dinitz, S. (1972). *The prevention of juvenile delinquency: An experiment* (Vol. *36*). Columbus, OH: Ohio State University Press.

Rosenbaum, D. P. (1986). *Community crime prevention: Does it work?* London, UK: Sage.

Schlossman, S., & Sedlak, M. (1983). The Chicago area project revisited. *Crime & Delinquency*, *29*(3), 398–462. https://doi.org/10.1177/001112878302900305

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, UK: Oxford University Press.

Spergel, I. A., Wa, K. M., Grossman, S., Sosa, R. V., Anderson, L., Arata, L., … Lyndes, K. (2003). *The Little Village gang violence reduction project in Chicago*. Chicago, IL: University of Chicago, School of Social Service Administration.

Spergel, I. A., Wa, K. M., & Sosa, R. V. (2002). *Evaluation of the Mesa Gang Intervention Program (MGIP)*. Chicago, IL: University of Chicago.

Spergel, I. A., Wa, K. M., & Sosa, R. V. (2005a). *Evaluation of the Bloomington-Normal comprehensive gang program*. Chicago, IL: University of Chicago.

Spergel, I. A., Wa, K. M., & Sosa, R. V. (2005b). *Evaluation of the Riverside comprehensive community-wide approach to gang prevention, intervention & suppression*. Chicago, IL: University of Chicago.

Spergel, I. A., Wa, K. M., & Sosa, R. V. (2006). The comprehensive, community-wide gang program model: Success and failure. In J. F. Short & L. A. Hughes (Eds.), *Studying youth gangs* (pp. 203–224). Lanham, MD: AltaMira Press.

Thompson, D. W., & Jason, L. A. (1988). Street gangs and preventive interventions. *Criminal Justice and Behavior*, *15*(3), 323–333. https://doi.org/10.1177/0093854888015003005

Turner, S., & Petersilia, J. (1992). Focusing on high-risk parolees: An experiment to reduce commitments to the Texas Department of Corrections. *Journal of Research in Crime and Delinquency*, *29*(1), 34–61. https://doi.org/10.1177/0022427892029001003

Weiss, B., Caron, A., Ball, S., Tapp, J., Johnson, M., & Weisz, J. R. (2005). Iatrogenic effects of group treatment for antisocial youth. *Journal of Consulting and Clinical Psychology*, *73*(6), 1036–1044. https://doi.org/10.1037/0022-006X.73.6.1036

Welsh, B. C., & Rocque, M. (2014). When crime prevention harms: A review of systematic reviews. *Journal of Experimental Criminology*, *10*(3), 245–266. https://doi.org/10.1007/s11292-014-9199-2

Wiebush, R. G., Wagner, D., Mcnulty, B., Wang, Y., & Le, T. N. (2005). *Implementation and outcome evaluation of the intensive aftercare program: Final report*. Washington, DC: National Council on Crime and Delinquency.

Williams, K., Cohen, M. I., & Curry, D. (1999). *Evaluation of youth gang drug intervention/prevention programs for female adolescents*. Washington, DC: U.S. Department of Justice.

Willman, M. T., & Snortum, J. R. (1982). A police program for employment of youth gang members. *International Journal of Offender Therapy and Comparative Criminology*, *26*(3), 207–214. https://doi.org/10.1177/0306624X8202600302

Wilson, J. M., & Chermak, S. (2011). Community-driven violence reduction programs. *Criminology & Public Policy*, *10*(4), 993–1027. https://doi.org/10.1111/j.1745-9133.2011.00763.x

Wodarski, J. S., Filipczak, J., McCombs, D., Koustenis, G., & Rusilko, S. (1979). Follow-up on behavioral intervention with troublesome adolescents. *Journal of Behavior Therapy and Experimental Psychiatry*, *10*(3), 181–188. https://doi.org/10.1016/0005-7916(79)90057-0

Wong, J., Gravel, J., Bouchard, M., Morselli, C., & Descormiers, K. (2012). *Effectiveness of street gang control strategies: A systematic review and meta-analysis of evaluation studies*. Ottawa, ON: Research and National Coordination Organized Crime Division.

How to cite this article: Rubenson MP, Galbraith K, Shin O, Beam CR, Huey SJJr.. When helping hurts? Toward a nuanced interpretation of adverse effects in gang-focused interventions. *Clin Psychol Sci Pract*. 2020;00:e12321. https://doi.org/10.1111/cpsp.12321