

When crime prevention harms: a review of systematic reviews

Brandon C. Welsh · Michael Rocque

Published online: 21 January 2014

© Springer Science+Business Media Dordrecht 2014

Abstract

Objectives In a series of important scholarly works, Joan McCord made the case for the criminological community to take seriously harmful effects arising from individual-based crime prevention programs. Building on these works, two key questions are of central interest to this paper: What has been the state of research on harmful effects of these crime prevention programs since McCord's works? And what are the theoretical, methodological, and programmatic characteristics of individual-based crime prevention programs with reported harmful effects?

Methods This paper reports on the first empirical review of harmful effects of crime prevention programs, drawing upon 15 Campbell Collaboration systematic reviews. Altogether, 574 experimental and quasi-experimental studies (published and unpublished) with 645 independent effect sizes were reviewed.

Results A total of 22 harmful effects from 22 unique studies of individual-based crime prevention programs were identified. Almost all of the studies have been reported since 1990, all but 2 were carried out in the United States, and two-thirds can be considered unpublished. The studies covered a wide range of interventions, from anti-bullying programs at schools, to second responder interventions involving police, to the Scared Straight program for juvenile delinquents, with more than half taking place in criminal justice settings. Boot camps and drug courts accounted for the largest share of studies with harmful effects.

B. C. Welsh
Northeastern University, Boston, MA, USA

B. C. Welsh
Netherlands Institute for the Study of Crime and Law Enforcement, Amsterdam, The Netherlands

M. Rocque
Maine Department of Corrections, Augusta, ME, USA

M. Rocque
University of Maine, Orono, ME, USA

B. C. Welsh (✉)
School of Criminology and Criminal Justice, Northeastern University,
Churchill Hall, 360 Huntington Avenue, Boston, MA 02115, USA
e-mail: b.welsh@neu.edu

Conclusions Theory failure, implementation failure, and deviancy training were identified as the leading explanations for harmful effects of crime prevention programs, and they served as key anchors for a more focused look at implications for theory and policy. Also, the need for programs to be rigorously evaluated and monitored is evident, which will advance McCord's call for attention to safety and efficacy.

Keywords Crime prevention · Harmful effect · Systematic review · High quality evaluation · Theory failure · Implementation failure

Introduction

In a series of important scholarly works, Joan McCord (2002, 2003; Dishion et al. 1999) made the case for the criminological community to take seriously harmful effects arising from individual-based crime prevention programs. One passage from her 2003 article is particularly telling: “Unless social programs are evaluated for potential harm as well as benefit, safety as well as efficacy, the choice of which social programs to use will remain a dangerous guess” (p. 17). Several key arguments are at the heart of this view.

The first one is that the prevailing question in the search for effective crime prevention measures is far too narrow: “Does the program work or not?” (McCord 2003, p. 17). There is often no acknowledgment that these programs may cause harm. This concern has also been raised by others (e.g., Gatti et al. 2009; Sherman 2007). As Werch and Owen (2002, p. 581) note, “Common thinking regarding prevention and health promotion programs is that they are helpful, or perhaps benign at worst, but rarely if ever harmful.”

Second, bias exists in the reporting of negative effects arising from evaluation studies. McCord (2003) takes special aim at the well-known issue of publication bias, which holds that scientific journals are more likely to publish studies that report positive treatment effects (Wilson 2009). Closely related is the “file-in-the-drawer” problem (Rosenthal 1979), whereby null or negative effects are suppressed from public let alone academic scrutiny. McCord's main point here is that even when harmful effects are found, they often go unreported in the literature.

Another key argument advanced by McCord is that program evaluation matters and high quality designs are needed to be able to identify any harmful (or beneficial) treatment effects. (Properties of high quality designs are briefly discussed in the next section.) This has everything to do with the safety of the participants of crime prevention programs—the children, families, or offenders. For one, experimental trials can be shut down if the treatment condition is faring worse than the control condition. Closely monitored experiments, including wait-list control designs, allow for this possibility. Also, scientifically valid conclusions about harmful effects can serve as a warning for any planned replications or wider dissemination.

In making this argument, McCord did not shy away from its implications for “making social science more experimental” (Sherman 2003): “Recognizing that programs can have harmful effects may be critical to acceptance of experimental designs for evaluating social interventions” (McCord 2003, p. 29). Not surprisingly, McCord's point is altogether absent from recent critical commentary on randomized experiments (Hough 2010; Sampson 2010). Perhaps more surprisingly is that those on the other side

of the debate have not paid much attention to the safety element proffered by experiments (Farrington and Welsh 2006; Weisburd 2010).

These arguments for investigating potential harmful effects of crime prevention programs are just as salient today. As MacKenzie (2013) reminds us, the idiom “first do no harm”—from the Hippocratic Oath—must continue to be the guiding principle for all of those involved in the development, implementation, and evaluation of crime prevention programs. The recent experiences of some states moving toward evidence-based practice in juvenile justice (Greenwood and Welsh 2012) and correctional treatment more generally (Cullen 2013; MacKenzie 2012, 2013), not to mention increased innovation and experimentation in policing (Braga et al. 2012; Braga and Weisburd 2012), would certainly be welcomed by Joan McCord.¹ But as promising as these events may be, they call for just as much vigilance to avoid causing harm.

Two key questions are of central interest to this paper. The first, while not meant as a test of the scholarly influence of McCord’s works on the subject, seems like a natural starting point: What has been the state of research on harmful effects of individual-based crime prevention programs over the last decade? To address this, we carry out a review of the theoretical and empirical literature. The second question is: What are the theoretical, methodological, and programmatic characteristics of individual-based crime prevention programs with reported harmful effects? Here, we turn to a methodologically-informed review of studies drawn from 15 Campbell Collaboration systematic reviews. This question is guided by the absence of any systematic research on harmful effects of crime prevention programs as well as our interest to unpack the explanations for why some programs cause harm. Taken together, these two questions also provide a more meaningful framework for drawing lessons for crime prevention policy and theory. We begin with a review of the theoretical and empirical literature over the last decade.

Literature review

When and how do we know if a crime prevention program is harmful? What is required at minimum is a high quality research design in which the observed effect—whether it be harmful, positive, or null—can be isolated in a way that rules out alternative plausible explanations (Cook and Campbell 1979; Shadish et al. 2002). Here, some kind of control condition is necessary to estimate what would have happened to the experimental units (e.g., people or areas) if the intervention had not been applied to them—termed the “counterfactual inference.”

In principle, a randomized experiment has the highest possible internal validity because it can rule out many threats to internal validity. Importantly, this requires implementation with full integrity, and some threats may still be problematic (e.g., differential attrition, selection by history). Randomization is the only method of assignment that controls for unknown and unmeasured confounders as well as those that are known and measured (Weisburd and Hinkle 2012). The next best method is to

¹ In addition to her academic writings on these topics, McCord was a founding member of the Crime and Justice Steering Committee of the Campbell Collaboration and a president of the Academy of Experimental Criminology.

use a nonrandomized experiment or high quality quasi-experimental design, for example, regression discontinuity or interrupted time series (Cook and Campbell 1979).

It is noteworthy to begin this review with some of the other key points raised by McCord (2003). At its core, her work is a narrative review of 5 crime prevention programs that reported harmful effects, including the Cambridge-Somerville Youth Study for which she had conducted an extensive follow-up (McCord 1978). She surmises that the reason for harmful effects is because of deviant peer influence or deviancy training. Her “construct theory” is premised on the notion that youths take their cues from each other in terms of behavioral motivation. When it is expected that deviance will result in their peers holding them in higher esteem, juveniles are more likely to engage in it. This situation, according to McCord, is especially relevant to crime prevention programs that mix high-rate or chronic delinquents with more minor delinquent youths (Dishion et al. 1999; Dodge et al. 2005). Thus, to McCord, the cause of the harmful effect found in crime prevention programs is a result of a social learning process.

Several articles have focused on the harmful effects of individual-based crime prevention programs since McCord’s series of works in the early 2000s. The balance of these more recent works have been theoretical (Dodge et al. 2005; Dishion and Dodge 2005) or policy-oriented (Dodge et al. 2006; Gottfredson 2010). They have drawn attention to the theoretical underpinnings of harmful effects, the need for policymakers and practitioners to reconsider the grouping of deviant peers, and the need for continued research in this area.

There have also been a number of recent reviews of the literature. Cécile and Born (2009), Rhule (2005), and Weiss et al. (2005) are particularly relevant. In their review, Cécile and Born (2009) conclude that when programs cause harm it is due to the peer contagion effect (or social learning), similar to the construct theory. With this in mind, the authors suggest several possible “moderating” factors that may reduce the negative influence of antisocial peers in crime prevention programs, including more structured activities, familial involvement, and contact with pro-social peers.

Rhule (2005) reviews 5 programs (including the Cambridge-Somerville Youth Study) that report harmful effects. Relying on the conclusions of the original study authors, Rhule argues that group aggregation leading to deviancy training is the likely cause of the harmful effects of the programs. However, Rhule recognizes that not all scholars agree on the specific mechanisms driving harmful effects. The degree of harm associated with prevention programs varies by the type of program as well as individual factors. Rhule, like Cécile and Born (2009), offers several pieces of advice to reduce harmful effects, including greater awareness of the issue, more structured and supervised programs, and further research on the possible causes of harmful effects.

While negative peer influence remains the dominant explanation thus far proffered for harmful effects, some scholars have been quite critical of this view. For example, Weiss et al. (2005) argue that any deviancy training effect is likely to be only a fraction of the amount of negative peer influence that such youth are exposed to in everyday life. They go on to discuss several studies purporting to find a deviancy training effect and identify various limitations (e.g., small to marginally significant effects). Weiss et al. also suggest that a finding that unstructured peer socializing leads to increased deviance cannot always be called a harmful effect, because such socializing is not often *part* of the treatment. In other words, it is not the treatment that is causing harm. Other studies have similarly concluded that the deviancy training effect may have been overstated in the crime prevention literature (see Handwerk et al. 2000).

Few studies to date have sought to examine empirically the effect of deviancy training in crime prevention programs. Rorie et al. (2011) report in detail on one after-school program in an effort to isolate the factors associated with deviancy training. A strength of this study is that the authors control for selection in the activities they observe, by examining “within-activity” relationships as well as important covariates. The research team analyzed nearly 400 activities, with 3,000 5-min intervals. The factor that emerged as most important was the structure of the activity, in terms of how well the organization of the activity was defined. The researchers coded “deviancy training” within the activities. They found that peers often reacted to deviance positively (thus reinforcing such behavior), and this is more likely to occur in activities with less structure (see also Gottfredson 2010).

In another study, Gatti et al. (2009) analyze the correlates of harmful effects within a sample of Montréal youths. They specifically examine the effect of variations in juvenile court sanctions on later criminal behavior, and find that the more supervision and intense the intervention (which would result in aggregating deviant peers), the greater the criminogenic effect.

Werch and Owen (2002) carry out a review that was slightly more systematic in its coverage of the literature, focusing on substance abuse programs. Altogether, they review 17 studies with harmful effects. Often, the studies with negative effects also had positive effects, suggesting that programs may not be “bad” or “good” in total. They argue that program theory failure may be one possible reason for harmful effects—something the criminological literature has not considered with respect to programs that harm.

Other mechanisms may be associated with harmful effects. For example, with respect to juvenile justice interventions, the labeling or stigmatizing effect may increase anti-social behavior on the part of treatment recipients. Research supports the notion that those marked as deviant often become more so after such labeling (Lopes et al. 2012; Paternoster and Iovanni 1989; Sampson and Laub 1997). It is also possible that selection effects can partially account for certain harmful outcomes, which points to the importance of considering the *type* of crime prevention program (i.e., universal, selected, or indicated). Psychological research on treatments that harm has argued that programs may backfire for a variety of reasons, including dependency on the treatment (meaning that once the individual is no longer benefiting from the treatment, his or her antisocial behavior increases), poorly trained therapists, and dosage effects (Lillienfeld 2007).

Methodology

To address our second question of interest, we searched for, coded, and analyzed studies that reported harmful effects on crime outcomes. While it would be inaccurate to say that we conducted a full-blown systematic review of harmful effects, we adopted the core principles of a methodologically-informed review, and these particulars are described here.

Sample

Campbell Collaboration systematic reviews served as the basis for the sample of studies used here. Several reasons guided our use of systematic reviews and, specifically, the Campbell Collaboration’s electronic library of systematic reviews on crime and justice

(see http://www.campbellcollaboration.org/reviews_crime_justice/index.php). First, the systematic review is the most rigorous method for reviewing the extant literature on a particular topic (Petticrew and Roberts 2006; Petrosino and Lavenberg 2007). These reviews “essentially take an epidemiological look at the methodology and results sections of a specific population of studies to reach a research-based consensus on a given study topic” (Johnson et al. 2000, p. 35). They use rigorous methods for locating, appraising, and synthesizing evidence from prior studies, and they are reported with the same level of detail that characterizes high quality reports of original research.

A second reason we relied on systematic reviews is that they are not limited to studies published in scientific journals. There is an explicit aim to also search for unpublished studies, which include some governmental and non-governmental agency reports, dissertations and theses, and conference papers, among other sources. This is crucial for addressing publication bias (Wilson 2009). A third reason is that systematic reviews are dedicated to the assembly of the most complete data possible. The reviewer seeks to obtain all relevant studies meeting explicit eligibility criteria, including a minimum level of methodological quality. For the Campbell Collaboration’s Crime and Justice Group, this means rigorous quasi-experimental designs, which incorporate before-and-after measures of crime in treatment and control conditions. This goes a long way toward addressing McCord’s (2003) concern about the ability to monitor the safety of the participants of crime prevention programs.

As for the use of the Campbell Collaboration’s electronic library, it represents the most authoritative source of systematic reviews in the social and behavioral sciences and has the most extensive coverage of criminological interventions (Farrington et al. 2011). The breadth of its coverage is important for the present review, because we are interested in expanding on McCord’s (2003) review.

As of this writing, the Crime and Justice Group has published 34 systematic reviews, and a number of these have been updated recently.² Fifteen of these systematic reviews, ranging from early parent training to boot camps, served as the basis for the sample of studies used here. Of the 19 reviews that were not included, 4 dealt with non-intervention topics, 3 did not report crime outcomes, 1 did not include studies with control conditions, and another did not provide sufficient information on the included studies.³ The other 10 systematic reviews that we excluded focused on the physical setting or place (e.g., city center, public housing community, residential area) rather than the individual (e.g., child, family, offender) as the unit of analysis. (See below for more details on this point.)

Criteria for inclusion of studies

In selecting studies for inclusion, the following criteria were used:

1. There was an outcome measure of delinquency or offending. A study would not be included if it only had outcome measures of risk factors for delinquency or offending (e.g., poor parental supervision, parental conflict).

² Publication dates for these reviews range from 2005 to 2012, with 28 published between 2008 and 2012.

³ As noted by Petrosino et al. (2010, p. 13), “Because the system processing condition is usually the control group in the experiments, it is often not described further.”

2. The effect of the intervention was harmful to members of the treatment group compared to their control group counterparts. Conventional measures of statistical significance had to be reported for effect sizes. For each study, we used either a reported summary outcome (e.g., total crime incidents) or a solitary reported outcome if the study only examined the effects of the intervention on a single crime outcome (e.g., violent crime). Where studies reported multiple effects over time, the latest follow-up was used. In order to identify harmful effects, we reviewed the meta-analysis in each systematic review. Where an effect size for a particular study was statistically significantly negative, we included it in the sample. It is important to point out that we are not equating a single negative effect with evidence of a program causing harm.
3. The child, family, or offender was the focus of the intervention and the unit of analysis was the individual. Our decision to focus exclusively on individual-level studies was an effort to maintain consistency with McCord's (2003) review, which also focused exclusively on individual-level studies.
4. The evaluation design was of high quality methodologically, with the minimum design involving before-and-after measures of crime in treatment and control conditions.

Measures

The following key features of the studies were coded:

Author and date The authors and dates of the most relevant evaluation reports are listed.

Location The place, often the state, as well as the country where the program was implemented are listed.

Published report It is noted whether or not the study was published.

Type of intervention The type of intervention that the treatment group received is identified.

Age at intervention This refers to the initial age of the individuals who received the intervention.

Sex The sex of the individuals who received the intervention is identified.

Context This is defined as the physical setting in which the intervention took place.

Sample size Unless otherwise stated, the figures listed here are initial sample sizes at the start of the intervention.

Dosage This is a measure of the duration of the intervention. Ideally, the number of contacts should be noted, as the intensity of intervention depends both on its duration and on the number and intensity (e.g., length) of contacts.

Intervention format This is a measure of how the intervention was delivered (i.e., individual or group).

Research design Research designs included either experimental or quasi-experimental.

Comparison Where the control group received some kind of treatment, including usual services (e.g., traditional probation, prison services), this is described.

Coding protocol

A coding protocol was established by the research team. The first step involved the researchers meeting to develop the criteria for inclusion of studies and to identify the measures to be coded. Next, studies were coded by one of the researchers. The researchers met and communicated periodically to discuss the coding of all of the studies and resolve any questions. Due to some study features (e.g., sex of participants, dosage) missing from the systematic reviews and the need for more analysis, the original studies were retrieved. We were successful in collecting all but 2 of the studies. These studies were then divided among the research team. In the event of further missing information or the need for clarification about our coding, which was the case for 2 studies, the original study authors were contacted by one of the researchers.

Results

Table 1 presents the total number of studies (n), the total number of independent effect sizes reported (k), and the corresponding results (desirable, null, or undesirable) as organized by the 15 systematic reviews consulted. Altogether, 574 studies (representing 574 crime prevention programs) with 645 independent effect sizes were reviewed.⁴ Of these 645 effect sizes, only 22 or 3.4 % were undesirable or harmful, while a good majority (60.6 %) were null (see Table 1). Each of the 22 harmful effect sizes come from 22 unique studies, meaning that there were not multiple harmful effect sizes from the same program (i.e., for different measures of offending).

It is important to draw attention to two points concerning these data. First, the 645 independent effect sizes originating from the 574 studies are an underestimate; that is, we would expect the number of effect sizes per study to be much higher than 1.1 (645/574). This is a function of our use of systematic reviews rather than primary studies. Authors of systematic reviews are more inclined (and expected) to report only the relevant effect sizes of the studies. Second, we are cautious not to simplistically equate a single undesirable effect size with evidence of a harm-causing study or program. Of the 22 studies with harmful effects (20 of which we were successful in collecting), 10 reported positive or null effects for other outcomes. While it is the case that we are specifically investigating harmful effects on delinquency or offending measures, where other program effects are reported they need to be considered in any characterization of

⁴ All the 574 studies met the criteria for inclusion in the systematic reviews that we consulted, and all the effect sizes were based on treatment versus control comparisons.

Table 1 Systematic reviews consulted and total number of studies and effects

Intervention	<i>n</i>	<i>k</i>	Desirable	Null	Undesirable
Court mandated interventions for DV	10	11	1	9	1
CBT for offenders	57	57	17	40	0
Boot camps	32	43	4	34	5
Drug courts	154	154	71	78	5
Drug substitution	45	48	24	24	0
Early family/parent training	55	55	21	34	0
Second responder	10	15	1	12	2
Mentoring	20	20	7	13	0
Scared Straight	7	7	0	5	2
Anti-bullying	53	82	34	47	1
Self-control	28	31	10	21	0
Serious Juvenile Offender Programs	17	30	3	27	0
Non-custodial employment	8	10	1	9	0
Incarceration-based drug treatment	55	55	27	24	4
Custodial vs. noncustodial sanctions	23	27	11	14	2
Total	574	645	232	391	22

Source: http://www.campbellcollaboration.org/reviews_crime_justice/index.php

DV domestic violence; *CBT* cognitive behavioral therapy

the program (see Weiss et al. 2005; Werch and Owen 2002). Additionally, as stated above, our focus is on the studies that indicated harmful effects. Our ability to draw inferences from those studies to the programs, particularly where the number or proportion of harmful effects is small in a given meta-analysis, is limited. For studies with a small proportion of harmful effects, we also cannot rule out sampling error.

Table 1 also shows that the 22 harmful effects originated from only 8 of the 15 systematic reviews of criminological interventions. No harmful effects were associated with the following 7 interventions: cognitive-behavioral therapy for offenders; drug substitution; early family/parent training; mentoring; self-control programs; programs for serious juvenile offenders; and non-custodial employment programs. Of the 8 interventions associated with harmful effects, Scared Straight had the highest within-group proportion of studies with harmful effects at 28.6 % (or 2 out of 7 studies). This was followed by second responder programs (20 %, or 2 out of 10 studies) and boot camps (15.6 %, or 5 out of 32 studies), with anti-bullying programs at schools having the lowest within-group proportion of studies with harmful effects (1.9 %, or 1 out of 53 studies). At 5 studies apiece, boot camps and drug courts accounted for the largest share or between-group proportion of studies with harmful effects (22.7 %, or 5 out of 22).

We now turn to a closer look at key characteristics of the 22 studies with harmful effects and report on some analyses. Almost all the studies were carried out in the United States ($n=20$), with one apiece conducted in Canada and Taiwan (Bonta et al. 2000; Vaughn et al. (2003); see Table 2). Just over two-thirds (68.2 %) or 15 of the studies can be considered unpublished; that is, they come from dissertations and unpublished government reports. This is very much in keeping with McCord's

Table 2 Key characteristics of studies with harmful effects ($n=22$)

Author/Date	Location	Published (Yes/No)	Intervention type	Dosage	Sample	Age	Sex	Research design
Harrell (1991)	Maryland, USA	N	Court mandated	3 months	348	Adult	Male	Q-Exp
MacKenzie and Souryal (1994)	South Carolina, USA	N	Boot camp	3 months	164	Adult	Male	Q-Exp
MacKenzie and Souryal (1994)	Georgia, USA	N	Boot camp	3 months	218	Adult	Male	Q-Exp
Boyles et al. (1996)	Colorado, USA	N	Boot camp	2 months	735	Juvenile	Male	Q-Exp
Jones and Ross (1997)	North Carolina, USA	Y	Boot camp	3-4 months	700	Adult	Male	Q-Exp
State of New York Department of Correctional Services (NY DCS) (2003)	New York, USA	N	Boot camp	n.a.	5,365	Adult	Mix	Q-Exp
Gilbertson (2009)	Minnesota, USA	N	Drug court	n.a.	32	Adult	n.a.	Q-Exp
Meekins (2003)	Virginia, USA	N	Drug court	12 months	249	Adult	Mix	Q-Exp
Terry (1995)	Florida, USA	N	Drug court	n.a.	1,088	Adult	n.a.	Q-Exp
Scarpitti et al. (2005)	Delaware, USA	N	Drug court	6 months	861	Adult	Mix	Q-Exp
NPC Research (2009)	Maryland, USA	N	Drug court	22.9 months	1,368	Adult	Mix	Q-Exp
Davis and Medina (2001)	New York, USA	N	Second responder	1 visit	406	Adult	Mix	Q-Exp
Hovell et al. (2006)	California, USA	Y	Second responder	1 visit	805	Adult	Mix	Q-Exp
Finckenaue (1982)	New Jersey, USA	Y	Scared Straight	Rap session	82	Juvenile	Male	Experiment
Michigan Department of Corrections (1967)	Michigan, USA	N	Scared Straight	1 tour	60	Juvenile	Male	Experiment
Rosenbluth et al. (2004)	Texas, USA	Y	Anti-bullying	3 months	1,763	Juvenile	Mix	Experiment
Vaughn et al. (2003)	Taiwan	Y	Drug treatment	3 months	628	Adult	Mix	Q-Exp
Siegal et al. (1997)	Ohio, USA	N	Drug treatment	< 3 months	729	Adult	Mix	Q-Exp
Porter (2002)	Pennsylvania, USA	N	Drug treatment	>12 months	513	Adult	Male	Q-Exp

Table 2 (continued)

Author/Date	Location	Published (Yes/No)	Intervention type	Dosage	Sample	Age	Sex	Research design
Gransky and Jones (1995)	Illinois, USA	N	Drug treatment	n.a.	8,399	Mix	n.a.	Q-Exp
MacKenzie and Shaw (1993)	Louisiana, USA	Y	Noncustodial	24 months	543	Adult	n.a.	Q-Exp
Bonta et al. (2000)	Newfoundland, Canada	Y	Noncustodial	> 1 month	154	Adult	Male	Q-Exp

NPC Northwest Pacific Consortium; Q-Exp quasi-experiment; n.a. not available

(2003) finding that most studies with harmful effects are not to be found in the published literature.

The intervention dosage or duration varied widely across the studies. Of the 18 studies that reported this information, the duration of the intervention ranged from a low of 1 month to a high of 24 months, with other studies reporting 1 visit, 1 tour, or even 1 “rap session” in the case of a Scared Straight program (Finckenauer 1982). In terms of demographic characteristics, the majority of the studies were comprised of adult participants (77.3 %, or 17 studies). Of the remaining studies, 4 included juveniles and 1 included participants ages 17 and older (which we refer to as mixed). Not one study had a female-only sample, but 9 had both males and females. The mean sample size for adult-only studies was approximately 830, while the mean sample size for juvenile-only studies was 660.

Table 3 presents data on the context of intervention by the year the study was reported. Studies were grouped by whether they took place outside of the formal justice system (i.e., communities and schools) or as part of the system (i.e., institution, courts, and police). More than half of the 22 studies with harmful effects (59.1 %, $n=13$) took place in criminal justice settings. As shown in Table 3, only 2 of the studies took place prior to 1990. Between 1990 and 1999, 9 of the studies were reported, with the majority of these taking place in criminal justice settings ($n=7$). A full half of the 22 studies were reported between 2000 and 2009, with 4 carried out in criminal justice settings. While it is difficult to say with any certainty what may be driving the recent increase in harmful effects (i.e., from 1990 onwards), it could very well be just a function of the overall growth in evaluations of crime prevention programs during this time (see, e.g., Sherman et al. 2006; Welsh and Farrington 2012b).

As noted above, most of the research on harmful effects in the crime prevention literature has focused on the deviancy training hypothesis. While our data do not allow for a formal test of this hypothesis, the data do allow for a much more thorough assessment than has been reported previously. Table 4 presents a breakdown of the studies with harmful effects by age of the study samples (juvenile or adult) and whether the programs were delivered in a group or individual setting. Three of the 22 studies were excluded from this analysis because they were delivered in both individual and group settings; 1 of these 3 studies also had a mixed sample of juveniles and adults.

Of the 19 studies with harmful effects included in Table 4, almost three-quarters (73.7 %) or 14 were delivered in group settings. The data also show that far more adult sample studies compared to juvenile sample studies took place in group settings (10 vs. 4). This by no means disconfirms the deviancy training hypothesis, which has only been discussed in the context of crime prevention programs for juveniles. It does,

Table 3 Number of studies with harmful effects by context of intervention and year ($n=22$)

Year	Community/school n (%)	CJS n (%)	Total n (%)
<1980	0 (0.0)	1 (7.7)	1 (4.5)
1980–1989	0 (0.0)	1 (7.7)	1 (4.5)
1990–1999	2 (22.2)	7 (53.9)	9 (40.9)
2000–2009	7 (77.7)	4 (30.8)	11 (50.0)
Total	9 (100)	13 (100)	22 (100)

Totals may not add due to rounding
CJS criminal justice system

Table 4 Number of studies with harmful effects by age and type of delivery ($n=19$)

Age	Individual <i>n</i> (%)	Group <i>n</i> (%)	Total <i>n</i> (%)
Adult	5 (100)	10 (71.4)	15 (79.0)
Juvenile	0 (0.0)	4 (28.6)	4 (21.1)
Total	5 (100)	14 (100)	19 (100)

Note: Totals may not add due to rounding

however, raise the possibility that this very mechanism or some other mechanism is operating with crime prevention programs for adults. Moreover, 5 of the 15 adult sample studies were delivered on an individual basis. Based on these findings, it would certainly seem that the explanation for harmful effects of prevention programs is more complex than some of the earlier research has suggested.

Limitations

There were several limitations to this research. One concerns the nature of the sample of studies with harmful effects. Not every study with harmful effects in the crime prevention literature is included here. Moreover, the two most well-known studies with harmful effects—the Cambridge-Somerville Youth Study (McCord 1978) and the Adolescent Transitions Program (Dishion et al. 1999; see also Poulin et al. 2001)—are not among the 22 studies. The reason for this is that neither study was included in the 15 systematic reviews that served as the basis for the sample of studies. We did not judge it prudent to simply add them after-the-fact.

There is, however, reason to believe that the systematic reviews consulted and the studies with harmful effects that were drawn from the reviews are somewhat representative of the literature of individual-based crime prevention programs. Among the 15 systematic reviews consulted, two-thirds (66.7 %) or 10 can be classified as justice system interventions (e.g., boot camps, incarceration-based drug treatment). Of the 22 studies with harmful effects, almost three-fifths (59.1 %) or 13 took place in a juvenile or criminal justice setting; the remainder took place in schools and communities. Also, the studies with harmful effects covered a fairly diverse range of interventions, from anti-bullying programs at schools, to second responder interventions involving police, to the Scared Straight program for juvenile delinquents.

Another limitation of this research concerns the temporal dimension of studies; that is, for longitudinal studies, at what point in time does a harmful effect get classified as a harmful effect? The case in point could very well be the Cambridge-Somerville Youth Study. Had McCord been able to conduct an even longer follow-up and found, for example, that the treatment group no longer had higher rates of violence or death compared to their control counterparts, it is conceivable that the program would no longer continue to be viewed as harmful. There are at least three issues to consider here. One is that limited follow-ups as well as short post-intervention follow-up time periods (e.g., 6 or 12 months) are the norm in evaluations of crime prevention programs (Sherman et al. 2006). This means that the determination of which effect size to use is often not so problematic.

Another issue is that studies with multiple follow-ups may find that program effects change over time; for example, a desirable effect on delinquency at age 15 becoming a

null effect at age 19. This is exactly what happened for the treatment group boys (compared to their control counterparts) in the age-19 follow-up of the Nurse-Family Partnership program (Eckenrode et al. 2010). By using the latest follow-up of the studies, we were able to account for this possibility. A third issue concerns multiple measures of the same outcome at one point in time and with differing results. We were confronted with this very scenario with the 3-year post-intervention follow-up of the Montréal Longitudinal-Experimental Study (Tremblay et al. 1991). The mother-rated measure of her boy's delinquency was found to be higher for the treatment group compared to the control group (hence a harmful effect), while the teacher-rated and self-reported measures of delinquency did not indicate a harmful effect. Because much longer follow-ups of this study have since been carried out (e.g., Boisjoli et al. 2007), which did not show a harmful effect for delinquency, we did not count this short-lived harmful effect. As is often done in the case of meta-analyses, the matter of multiple measures of the same outcome can be dealt with by aggregating the effects.

One other (potential) limitation is the definition of crime prevention that we employed. Following McCord (2003), we adopted a broad view of crime prevention, one which encompasses the full range of techniques, from prenatal home visits to prison sentences. Here, prevention is defined more by its outcome—the prevention of a future criminal event—than its character or approach (e.g., intervening in the first instance, being delivered outside of the justice system; see Welsh and Farrington 2012a). We hesitate to speculate on why McCord used this definition. For us, it captures the full breadth of the “cures” in the main title of her 2003 article: “Cures That Harm.” That is, the cure or the measure to reduce crime involves trying to bring about a social good, by intervening to serve an existing need or improving upon the status quo. Nevertheless, this has become an accepted view of crime prevention in many circles (see, e.g., Sherman et al. 2006).

Discussion and conclusions

What lessons can be drawn from studying harmful effects of crime prevention programs? This was the overarching question that guided the present research. Crucial to its exploration, we were interested in the relevance that harmful effects hold for theory and public policy. Inspired by the scholarship of Joan McCord, a growing body of criminological and psychological research has begun to investigate “cures that harm.” Much of this research has paid particular attention to the deviancy training hypothesis, and has included some analyses of prevention programs for at-risk and delinquent youth. Several narrative reviews of the literature have also been carried out, with some attempting to capture the wider universe of harmful effects of prevention programs in the areas of crime and substance abuse. Curiously, there has been no research that has systematically assessed harmful effects of crime prevention programs.

Based on 15 systematic reviews that included 574 high quality individual-level crime prevention studies with 645 independent effect sizes, we identified a total of 22 harmful effects from 22 unique studies of crime prevention programs. From this sample of studies, we estimate that the base rate of harmful effects arising from individual-based crime prevention programs is approximately 1 out of 26 studies or 3.8 %. For those concerned that these crime prevention programs may cause harm, this seems to be a rather small risk. However, this risk is by no means trivial. As McCord (2003) and

other scholars have argued, the safety as much as the efficacy of crime prevention programs is deserving of careful study. Interestingly, it does not appear that this risk of harmful effects is moderated (or even exacerbated) by several key characteristics of the studies.

Deviancy training

Drawing upon her own empirical analyses, McCord (2002) argued that the most likely means by which social programs resulted in harm was via deviancy training or deviant peer contagion. Her construct theory has not been the subject of much research. In fact, McCord herself stated, “We are a long way from understanding how this [harmful] result came about” (2002, p. 235). It seems to be the case, however, that a fair amount of research assumes a priori that harmful effects result from deviant peer influence (e.g., Cécile and Born 2009). Unfortunately, to date, this theory has not been systematically evaluated and it is not without its detractors (Handwerk et al. 2000; Weiss et al. 2005).

Our findings may offer some insight on the deviancy training effect. We found, much like Dodge et al. (2006), that harmful juvenile crime prevention programs are more often conducted in groups. In our sample, we did not find one study with a harmful effect that was delivered on an individual basis for juveniles. In their overall criticism of this view, Weiss et al. (2005) argued that individual-oriented programs may just happen to be more successful. It also needs to be stressed that a high percentage of effective crime prevention programs for juveniles are delivered in group settings (Greenwood 2006). Interestingly, one-third of the adult studies were delivered on an individual basis (see Table 4). It is more difficult to argue that harmful effects emanating from these individual-based programs and serving adults are the result of deviancy training.

While a number of researchers have reported evidence of deviancy training, it is our view that crime prevention research would be well served by exploring alternative explanations for harmful effects. For example, Greenwood (2006) raised the prospects of negative labeling and ineffective programming. These mechanisms may account for harmful effects even within a group-based program. Gatti et al. (2009) argued that the harmful effects they identified may be explained both by labeling and deviancy training. The negative stigma attached to being involved in the justice system may be more pronounced for particular individuals. According to Gatti et al. (2009, p. 996), those “selected” for intervention are “weakest from a personal and social point of view.” Future research should address whether harmful effects are uniform within particular programs or if there is heterogeneity.

Identifying the precise mechanism(s) by which programs cause harm is not a trivial matter. This issue has implications for both theory and policy. Understanding program elements that are more or less likely to cause harm can help inform program theory and also illuminate the processes by which programs influence future behavior. This information can also be useful for guiding the direction of policy and future programming. For example, if deviancy training is the major mechanism by which harmful effects are produced, then programs should be more individual-oriented or more structured. If, however, other mechanisms are at work, then it becomes incumbent upon program developers to devise programs that minimize components most likely to cause harm.

The deviancy training hypothesis also suggests that (1) programs conducted in groups have similar (negative) effects and (2) all individuals respond to group-based programs in similar ways. This may be a shortsighted approach. Greene (2004, pp. 80–1), after reviewing the literature on social learning and delinquency, argues that “naturally occurring group affiliation processes” result in negative outcomes, but programs that offer more structure with respect to group activities have shown more promise. Gottfredson (2010) and Rorie et al. (2011) agree with this view, based on their finding that programs with more structure and involvement from adults produce more positive results. More structure may impede the socializing that leads to deviancy transmission. In addition, group-based programs that do not mix individuals of varying risk levels would be recommended as opposed to programs that fail to match the antisocial level of participants (Greene 2004). This is analogous to the responsivity principle in the risk-need-responsivity paradigm in correctional treatment (Andrews and Bonta 2010). Research to investigate the mechanism of deviancy training in group-based programs could begin by varying the degree of structure, adult involvement, and individual risk level.

Other explanations

There are several other, more general explanations for how prevention programs may end up producing harmful effects. These include measurement failure, theory failure, and implementation failure (Ekblom and Pease 1995; see also Werch and Owen 2002). In addition, where the proportion of harmful effects on a given intervention was small, we cannot rule out sampling error. These explanations are not exclusively linked with harmful effects. They can just as likely be invoked to explain why a program ended up producing a null effect instead of a desirable effect.

Broadly construed, measurement failure is the failure of an evaluation to detect the “real effect” of a program (Ekblom and Pease 1995, p. 594). While we cannot fully rule out measurement failure as an explanation for 1 or more of the 22 studies with harmful effects, it does seem somewhat remote. For example, all 22 studies used either an experimental or rigorous quasi-experimental design to evaluate program effects, and by all accounts the studies were sufficiently robust. Their overall robustness, that is, having high internal, construct, and statistical conclusion validity, was a condition of their inclusion in the original systematic reviews.

Theory failure is when the “basic idea or mechanism of prevention was unsound” (Ekblom and Pease 1995, p. 594). The evidence suggests that theory failure played a role in several of the crime prevention programs. In the case of Scared Straight, for example, the harmful effects associated with 2 studies are very likely the result of full theory failure. In their systematic review of Scared Straight, which included a total of 7 studies, Petrosino et al. (2004) found a harmful effect overall. The other 5 studies reported null effects on criminal activity, with each in the direction favoring the control group. The authors also found that the programs did not suffer from any problems with implementation; that is, the programs were implemented as designed.

Aside from a loose connection to deterrence, there does not appear to be any theoretical basis to an intervention that involves subjecting youths—some with no prior contact with the juvenile justice system—to confrontations with violent, adult inmates and realistic and graphic depictions of life in prison. In the New Jersey Scared

Straight program, Finckenauer (1982, p. 169) offers the following explanation for the harmful effect: “The project may romanticize the Lifers—and by extension other prison inmates—in young, impressionable minds. Or, the belittling, demeaning, intimidating, and scaring of particular youth may be seen as a challenge; a challenge to go out and prove to themselves, their peers and others that they were not scared.” Finckenauer (1982) referred to this as a “delinquency fulfilling prophecy.”

Yet another piece of evidence that seems to support full theory failure for Scared Straight is that the 2 programs with harmful effects did not investigate other outcomes. So, the scenario of a program with harmful effects coexisting with null or positive effects did not present itself here. This also meets Werch and Owen’s (2002) criteria for program theory failure.

Boot camps may also suffer from theory failure. According to MacKenzie (2006, p. 279), military-style boot camps in juvenile and adult correctional settings “were not developed based on a coherent theoretical model.” With some ties to the social control perspective (i.e., skills training to help control impulsive behavior) and social learning theory (i.e., learning basic life skills to help return to the community), the overriding consideration of boot camps is strict discipline and rigorous physical activities, a product of the “get tough” on crime era (Mackenzie 2006).

The overall ineffectiveness of boot camps in reducing recidivism, as shown in Wilson et al.’s (2008) systematic review involving 32 studies, may offer further evidence of theory failure as an explanation for the harmful effects in 5 boot camp studies. However, it may only be partial theory failure. The reason for this has to do with another explanation offered by the authors. MacKenzie et al. (2001) found that in some juvenile boot camp studies the control groups received enhanced treatment (e.g., drug counseling, basic educational services), whereas juveniles in the boot camps were spending more time on physical activities. This would certainly suggest that there was more at work here than theory failure.

Implementation failure may explain why some of the other crime prevention programs produced harmful effects. The issue here is that “the idea was sound, but it was not properly put into effect” (Ekblom and Pease 1995, p. 594). The harmful effects associated with the 5 drug court studies are very likely a result of implementation failure rather than theory failure. This is because of the strong and substantial evidence of the effectiveness of drug courts. In their systematic review of drug courts, which included 154 studies, Mitchell et al. (2011; see also Mitchell et al. 2012a) found that recidivism was reduced by an average of 12 %.

Implementation failure is also likely the reason for the harmful effects of the 4 incarceration-based drug treatment studies. Here again, a systematic review, by Mitchell et al. (2012b), found strong and substantial evidence of this intervention’s effectiveness in reducing criminal recidivism. In both these cases, however, the lack of detail about implementation in the original studies precludes us from making an even stronger statement about implementation failure as an explanation for the harmful effects. Nevertheless, in the absence of measurement and theory failure and other plausible alternative explanations, implementation failure seems rather robust. From a policy perspective, this may not be such bad news. It stands to reason that trying to identify (and possibly even correct) implementation problems during the course of a program is far more straightforward, not to mention desirable, than having to contend with a program whose “basic idea or mechanism” is unsound.

Some priorities for policy

What is a policymaker or practitioner to do? They certainly do not need this research to remind them of their oath to do no harm. Instead, this research points to a number of priorities or general principles that may go some way toward lessening the prospects of programs causing harm as well as advancing crime prevention policy.

First, good public policy is based on sound theoretical understanding of the phenomena it addresses. As Green (2000, p. 125) reminds us, “empirical evidence alone is insufficient to direct practice, and that recourse to the explanatory and predictive capability of theory is essential to the design of both programmes and evaluations.” In the absence of sound theory, social programs stand little chance to bring about social good and may even cause harm. The present research identified a couple of examples of the latter. Some may not be so surprised by this state of affairs, what with these programs being derived from a “get tough” or punitive ideology.

Second, group-based programs for youths should continue to be thoroughly scrutinized for the possibility of peer contagion or deviancy training effects. This should not be viewed as a general indictment of these programs. Instead, we view this as a health warning of sorts. Our findings, along with the works of others (Handwerk et al. 2000; Weiss et al. 2005), suggest that crime prevention programs for youths can still be carried out in group settings under certain circumstances without the emergence of deviancy training effects. Policymakers and practitioners will need to attend to the complex interplay of the nature of the intervention, the risk level of the participating youths, and the setting (e.g., community, school, clinic, institution), among other key factors. We echo the recommendations of McCord (2002, 2003) and other scholars (e.g., Dodge et al. 2006; Gottfredson 2010) in calling for more targeted research in this area.

Lastly, programs need to be rigorously monitored and evaluated. Whether there is the prospect of theory failure, implementation failure, or deviancy training, McCord’s (2003) call for monitoring the safety of crime prevention programs rings true. A high quality evaluation design that is also closely monitored—along the lines of a process study or implementation evaluation—will allow for problems to be addressed or the program to be shut down. Active vigilance is in order. We suspect that Joan McCord would have had it no other way.

Acknowledgments We are grateful to David Wilson and the anonymous reviewers for helpful comments.

References

References marked with an asterisk (*) are studies with harmful effects included in the review.

- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct* (5th ed.). New Providence: Matthew Bender.
- Boisjoli, R., Vitaro, F., Lacourse, E., Barker, E. D., & Tremblay, R. E. (2007). Impact and clinical significance of a preventive intervention for disruptive boys: 15-year follow-up. *British Journal of Psychiatry*, *191*, 415–419.

- *Bonta, J., Wallace-Capretta, S., & Rooney, J. (2000). A quasi-experimental evaluation of an intensive rehabilitation supervision program. *Criminal Justice and Behavior*, 27, 312–329.
- *Boyles, C. E., Bokenkamp, E., & Madura, W. (1996). *Evaluation of the Colorado juvenile regimented training program*. Golden: Colorado Department of Human Services, Division of Youth Corrections.
- Braga, A. A., & Weisburd, D. L. (2012). The effects of focused deterrence strategies on crime: A systematic review and meta-analysis of the empirical evidence. *Journal of Research in Crime and Delinquency*, 49, 323–358.
- Braga, A. A., Papachristos, A. V., & Hureau, D. M. (2012). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly*. doi:10.1080/07418825.2012.673632.
- Cécile, M., & Born, M. (2009). Intervention in juvenile delinquency: Danger of iatrogenic effects? *Children and Youth Services Review*, 31, 1217–1221.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston: Houghton Mifflin.
- Cullen, F. T. (2013). Rehabilitation: Beyond nothing works. In M. Tonry (Ed.), *Crime and justice in America, 1975–2025* (pp. 299–376). Chicago: University of Chicago Press.
- *Davis, R. C., & Medina, J. (2001). *Results from an elder abuse prevention experiment in New York City*. National Institute of Justice Research in Brief. Washington: National Institute of Justice, U.S. Department of Justice.
- Dishion, T. J., & Dodge, K. A. (2005). Peer contagion in interventions for children and adolescents: Moving toward an understanding of the ecology and dynamics of change. *Journal of Abnormal Child Psychology*, 33, 395–400.
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist*, 54, 755–764.
- Dodge, K. A., Dishion, T. J., & Lansford, J. E. (Eds.). (2005). *Deviant peer influences in programs for youth: Problems and solutions*. New York: Guilford.
- Dodge, K. A., Dishion, T. J., & Lansford, J. E. (2006). Deviant peer influences in intervention and public policy for youth. *Social Policy Report*, 20, 1–19.
- Eckenrode, J., Campa, M., Luckey, D. W., Henderson, C. R., Cole, R., Kitzman, H., et al. (2010). Long-term effects of prenatal and infancy nurse home visitation on the life course of youths: 19-year follow-up a randomized trial. *Archives of Pediatrics and Adolescent Medicine*, 164, 9–15.
- Eklblom, P., & Pease, K. (1995). Evaluating crime prevention. In M. Tonry & D. P. Farrington (Eds.), *Building a safer society: Strategic approaches to crime prevention* (pp. 585–662). Chicago: University of Chicago Press.
- Farrington, D. P., & Welsh, B. C. (2006). A half century of randomized experiments on crime and justice. In M. Tonry (Ed.), *Crime and justice: A review of research* (Vol. 34, pp. 55–132). Chicago: University of Chicago Press.
- Farrington, D. P., Weisburd, D., & Gill, C. E. (2011). The Campbell Collaboration Crime and Justice Group: A decade of progress. In C. J. Smith, S. X. Zhang, & R. Barberet (Eds.), *Routledge handbook of international criminology* (pp. 53–63). New York: Routledge.
- *Finckenauer, J. O. (1982). *Scared Straight and the panacea phenomenon*. Englewood Cliffs: Prentice-Hall.
- Gatti, U., Tremblay, R. E., & Vitaro, F. (2009). Iatrogenic effect of juvenile justice. *Journal of Child Psychology and Psychiatry*, 50, 991–998.
- *Gilbertson, T. (2009). *2008 DWI court evaluation report*. Bemidji: Criminal Justice Department, Bemidji State University.
- 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, 229–243.
- *Gransky, L. A., & Jones, R. J. (1995). *Evaluation of the post-release status of substance abuse program participants*. Chicago: Illinois Criminal Justice Information Authority.
- Green, J. (2000). The role of theory in evidence-based health promotion practice. *Health Education Research*, 15, 125–129.
- Greene, M. B. (2004). Implications of research showing harmful effects of group activities with anti-social adolescents. In *Persistently safe schools: The national conference of the Hamilton Fish Institute on school and community violence* (pp. 73–83). Washington, DC: George Washington University.
- Greenwood, P. W. (2006). Promising solutions in juvenile justice. In K. A. Dodge, T. J. Dishion, & J. E. Lansford (Eds.), *Deviant peer influences in programs for youth: Problems and solutions* (pp. 278–295). New York: Guilford.
- Greenwood, P. W., & Welsh, B. C. (2012). Promoting evidence-based practice in delinquency prevention at the state level: Principles, progress, and policy directions. *Criminology and Public Policy*, 11, 491–513.
- 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, 223–238.

- *Harrell, A. (1991). Evaluation of court-ordered treatment for domestic violence offenders. Final report. Washington, DC: National Institute of Justice, U.S. Department of Justice.
- Hough, M. (2010). Gold standard or fool's gold? The pursuit of certainty in experimental criminology. *Criminology and Criminal Justice*, 10, 11–22.
- *Hovell, M. F., Seid, A. G., & Liles, S. (2006). Evaluation of a police and social services domestic violence program: Empirical evidence needed to inform public health policies. *Violence Against Women*, 12, 137–159.
- Johnson, B. R., De Li, S., Larson, D. B., & McCullough, M. (2000). A systematic review of the religiosity and delinquency literature: A research note. *Journal of Contemporary Criminal Justice*, 16, 32–52.
- *Jones, M., & Ross, D. L. (1997). Is less better? Boot camp, regular probation and rearrest in North Carolina. *American Journal of Criminal Justice*, 21, 147–161.
- Lillienfeld, S. O. (2007). Psychological treatments that cause harm. *Perspectives on Psychological Science*, 2, 53–70.
- Lopes, G., Krohn, M. D., Lizotte, A. J., Schmidt, N. M., Vasquez, B. E., & Bernberg, J. G. (2012). Labeling and cumulative disadvantage: The impact of formal police intervention on life chances and crime during emerging adulthood. *Crime and Delinquency*, 58, 456–488.
- MacKenzie, D. L. (2006). *What works in corrections: Reducing the criminal activities of offenders and delinquents*. New York: Cambridge University Press.
- MacKenzie, D. L. (2012). Preventing future criminal activities of delinquents and offenders. In B. C. Welsh & D. P. Farrington (Eds.), *The Oxford handbook of crime prevention* (pp. 466–486). New York: Oxford University Press.
- MacKenzie, D. L. (2013). First do no harm: A look at correctional policies and programs today: The 2011 Joan McCord Prize lecture. *Journal of Experimental Criminology*, 9, 1–17.
- *MacKenzie, D. L., & Shaw, J. W. (1993). The impact of shock incarceration on technical violations and new criminal activities. *Justice Quarterly*, 10, 463–487.
- *MacKenzie, D. L., & Souryal, C. (1994). *Multi-site evaluation of shock incarceration: Executive summary*. Washington, DC: National Institute of Justice, U.S. Department of Justice.
- MacKenzie, D. L., Wilson, D. B., & Kider, S. B. (2001). Effects of correctional boot camps on offending. *Annals of the American Academy of Political and Social Science*, 578, 126–143.
- McCord, J. (1978). A thirty-year follow-up of treatment effects. *American Psychologist*, 33, 284–289.
- McCord, J. (2002). Counterproductive juvenile justice. *Australian and New Zealand Journal of Criminology*, 35, 230–237.
- McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *Annals of the American Academy of Political and Social Science*, 587, 16–30.
- *Meekins, B. J. (2003). *Deterrence in the drug court setting: Case study and quasi experiment. Unpublished doctoral dissertation*. Alexandria: University of Virginia.
- *Michigan Department of Corrections (1967). A six month follow-up of juvenile delinquents visiting the Ionia Reformatory. Research Report No. 4. Lansing, MI: Michigan Department of Corrections.
- Mitchell, O., Wilson, D. B., Eggers, A., & MacKenzie, D. L. (2011). Drug courts' effects on criminal offending for juveniles and adults. *Campbell Collaboration*. doi:10.4073/csr.2012.4.
- Mitchell, O., Wilson, D. B., Eggers, A., & MacKenzie, D. L. (2012a). Assessing the effectiveness of drug courts on recidivism: A meta-analytic review of traditional and non-traditional drug courts. *Journal of Criminal Justice*, 40, 60–71.
- Mitchell, O., Wilson, D. B., & MacKenzie, D. L. (2012b). The effectiveness of incarceration-based drug treatment on criminal behavior: A systematic review. *Campbell Collaboration*. doi:10.4073/csr.2012.18.
- *NPC Research. (2009). *Baltimore City Circuit Court adult treatment court and felony diversion initiative: Outcome and cost evaluation*. Portland: NPC Research.
- Paternoster, R., & Iovanni, L. (1989). The labeling perspective and delinquency: An elaboration of the theory and an assessment of the evidence. *Justice Quarterly*, 6, 359–394.
- Petrosino, A., & Lavenberg, J. (2007). Systematic reviews and meta-analyses: Best evidence on 'what works' for criminal justice decision makers. *Western Criminology Review*, 8, 1–15.
- Petrosino, A., Turpin-Petrosino, C., & Buehler, J. (2004). "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency. *Campbell Collaboration*. doi:10.4073/csr.2004.2.
- Petrosino, A., Turpin-Petrosino, C., & Guckenburg, S. (2010). Formal system processing of juveniles: Effects on delinquency. *Campbell Collaboration*. doi:10.4073/csr.2010.1.
- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden: Blackwell.
- *Porter, R. (2002). *Breaking the cycle: Technical report*. New York: Vera Institute of Justice.

- Poulin, F., Dishion, D. J., & Burraston, B. (2001). 3-year iatrogenic effects associated with aggressive high-risk adolescents in cognitive-behavioral preventive interventions. *Applied Developmental Science, 5*, 214–224.
- Rhule, D. M. (2005). Take care to do no harm: Harmful interventions for youth problem behavior. *Professional Psychology: Research and Practice, 36*, 618–625.
- Rorie, M., Gottfredson, D. C., Cross, A., Wilson, D., & Connell, N. (2011). Structure and deviancy training in after-school programs. *Journal of Adolescence, 34*, 105–117.
- *Rosenbluth, B., Whitaker, D. J., Sanchez, E., & Valle, L. A. (2004). The Expect Respect Project: Preventing bullying and sexual harassment in US elementary schools. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 211–233). Cambridge: Cambridge University Press.
- Rosenthal, R. (1979). The ‘file drawer problem’ and tolerance of null results. *Psychological Bulletin, 86*, 638–641.
- Sampson, R. J. (2010). Gold standard myths: Observations on the experimental turn in quantitative criminology. *Journal of Quantitative Criminology, 26*, 489–500.
- Sampson, R. J., & Laub, J. H. (1997). A life-course theory of cumulative disadvantage and the stability of delinquency. In T. P. Thornberry (Ed.), *Developmental theories of crime and delinquency*. New Brunswick: Transaction.
- *Scarpitti, F. R., Butzin, C. A., Saum, C. A., Gray, A. R., & Leigey, M. E. (2005). *Drug court offenders in outpatient treatment: Final report to National Institute of Drug Abuse*. Newark: University of Delaware.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sherman, L. W. (2003). Preface: Misleading evidence and evidence-led policy: Making social science more experimental. *Annals of the American Academy of Political and Social Science, 589*, 6–19.
- Sherman, L. W. (2007). The power few: Experimental criminology and the reduction of harm. *Journal of Experimental Criminology, 3*, 299–321.
- Sherman, L. W., Farrington, D. P., Welsh, B. C., & MacKenzie, D. L. (Eds.). (2006). *Evidence-based crime prevention* (rev. ed.). New York: Routledge.
- *Siegal, H. A., Wang, J., Falck, R. S., Rahman, A. M., & Carlson, R. G. (1997). *An evaluation of Ohio's prison-based therapeutic community treatment programs for substance abusers: Final report*. Dayton: School of Medicine, Wright State University.
- *State of New York Department of Correctional Services Division of Parole (NY DCS). (2003). *The fifteenth annual shock legislative report*. New York: Department of Correctional Services and the Division of Parole.
- *Terry, W. C., III. (1995). *Repeat offenses of the first year cohort of Broward County, Florida's drug court*. Miami: Florida International University.
- Tremblay, R. E., McCord, J., Boileau, H., Charlebois, P., Gagnon, C., Le Blanc, M., et al. (1991). Can disruptive boys be helped to become competent? *Psychiatry, 54*, 148–161.
- *Vaughn, M. S., Deng, F., & Lee, L. J. (2003). Evaluating a prison-based drug treatment program in Taiwan. *Journal of Drug Issues, 33*, 357–384.
- Weisburd, D. (2010). Justifying the use of non-experimental methods and disqualifying the use of randomized controlled trials: Challenging the folklore in evaluation research in crime and justice. *Journal of Experimental Criminology, 6*, 209–227.
- Weisburd, D., & Hinkle, J. (2012). The Importance of randomized experiments in evaluating crime prevention. In B. C. Welsh & D. P. Farrington (Eds.), *The Oxford handbook of crime prevention* (pp. 446–465). New York: Oxford University Press.
- Weiss, B., Caron, A., Ball, S., Tapp, J., Johnson, J., & Weisz, J. R. (2005). Iatrogenic effects of group treatment for antisocial youth. *Journal of Consulting and Clinical Psychology, 73*, 1036–1044.
- Welsh, B. C., & Farrington, D. P. (2012a). Science, politics, and crime prevention: Toward a new crime policy. *Journal of Criminal Justice, 40*, 128–133.
- Welsh, B. C., & Farrington, D. P. (Eds.). (2012b). *The Oxford handbook of crime prevention*. New York: Oxford University Press.
- Werch, C. E., & Owen, D. M. (2002). Iatrogenic effects of alcohol and drug prevention programs. *Journal of Studies on Alcohol, 63*, 581–590.
- Wilson, D. B. (2009). Missing a critical piece of the pie: Simple document search strategies inadequate for systematic reviews. *Journal of Experimental Criminology, 5*, 429–440.
- Wilson, D. B., MacKenzie, D. L., & Mitchell, F. N. (2008). Effects of correctional boot camps on offending. *Campbell Collaboration*. doi:10.4073/csr.2003.1.

Brandon C. Welsh, Ph.D., is a Professor in the School of Criminology and Criminal Justice at Northeastern University and a Senior Research Fellow at the Netherlands Institute for the Study of Crime and Law Enforcement in Amsterdam. His research interests include the prevention of delinquency and crime and evidence-based policy. His latest book is *Experimental Criminology: Prospects for Advancing Science and Public Policy* (Cambridge University Press, 2013, with Anthony Braga and Gerben Bruinsma).

Michael Rocque, Ph.D., is the Director of Research for the Maine Department of Corrections and an adjunct faculty member in the Department of Sociology, University of Maine, Orono. His research interests include life course criminology, race and justice, and crime prevention. His work has appeared in the *Journal of Criminal Justice*, *Journal of Criminal Law and Criminology*, *Deviant Behavior*, *Victims and Offenders*, and *International Journal of Offender Therapy and Comparative Criminology*.