doi:10.1093/bjc/azv033 BRIT. J. CRIMINOL. (2016) **56**, 141–160
Advance Access publication 14 June 2015

EXAMINING THE IATROGENIC EFFECTS OF THE CAMBRIDGE-SOMERVILLE YOUTH STUDY: EXISTING EXPLANATIONS AND NEW APPRAISALS

Steven N. Zane, Brandon C. Welsh* and Gregory M. Zimmerman

Criminology has paid increasing attention to the prospect that prevention programmes can cause harm. The Cambridge-Somerville Youth Study, a delinquency prevention experiment of 506 boys that began in 1939, provides some of the earliest evidence of programmatic iatrogenic effects. A series of hypotheses were advanced by Joan McCord and other scholars to explain these unintended effects. Drawing upon this scholarship, related research and developmental theory, this article examines the leading explanations and offers new appraisals of iatrogenic effects of crime prevention programmes. The research suggests that there is not a grand explanation, and we encourage a more nuanced perspective for understanding iatrogenic effects of crime prevention programmes. Implications for policy and practice are discussed.

Keywords: Cambridge-Somerville Youth Study, delinquency prevention, iatrogenic effects, developmental theory

Introduction

The foundation of this article is firmly grounded in the premise that interventions attempting 'to modify human behavior … have the power to bring about unintended, harmful consequences' (Rhule 2005: 621). This knowledge has the ability to move us beyond the 'effective' versus 'not effective' distinction in the 'what works' literature, and toward a third possibility—that a prevention programme may cause harm. While many scholars and practitioners have observed that crime prevention programmes must first 'do no harm' (MacKenzie 2013), it is unlikely that this Hippocratic Oath can be upheld without an understanding of *why* some programmes cause harm.

In the spirit of this view, this article sets out to examine the leading explanations and offer new appraisals of iatrogenic effects of crime prevention programmes. We seek to accomplish this with reference to Joan McCord's work on the Cambridge-Somerville Youth Study (heretofore the CSYS), a longitudinal-experimental study that began in 1939 in neighbouring towns of Boston, Massachusetts. Situating an examination of iatrogenic effects with reference to the CSYS is appropriate, both because of the long-term prospective, longitudinal-experimental design of the study and the empirically examined explanations for the study's iatrogenic effects.

McCord's influential article on the CSYS, published in 1978, reported on the results of the 30-year follow-up (mean age = 45 years) of the delinquency prevention experiment. Findings indicated that the programme produced harmful effects. Her research

Steven N. Zane, School of Criminology and Criminal Justice, Northeastern University, Churchill Hall, 360 Huntington Avenue, Boston, MA 02115, USA; *Brandon C. Welsh, School of Criminology and Criminal Justice, Northeastern University, Churchill Hall, 360 Huntington Avenue, Boston, MA 02115, USA; Netherlands Institute for the Study of Crime and Law Enforcement, De Boelelaan 1077a, 1081 HV Amsterdam, the Netherlands; b.welsh@neu.edu; Gregory M. Zimmerman, School of Criminology and Criminal Justice, Northeastern University, Churchill Hall, 360 Huntington Avenue, Boston, MA 02115, USA.

was comprehensive, investigating and detecting iatrogenic effects for a wide range of important life-course outcomes, including criminal offending, physical and mental health, family stability, employment and alcohol abuse. Of the 15 criminal outcome comparisons between the treatment and control groups, none favoured the treatment group and one favoured the control group; of the 15 health outcome comparisons, none favoured the treatment group and 4 favoured the controls and of the 13 comparisons of family, work and leisure time, none favoured the treatment group and 2 favoured the controls (McCord 1978).

To compare outcomes across a common metric, McCord (1978) grouped subjects by whether or not they experienced any 'undesirable outcomes', which included FBI index crimes (all of which were felonies), treated alcoholism, serious mental illness and death. The results indicated that 42 per cent of treatment group participants experienced undesirable outcomes compared to 32 per cent of controls, a statistically significant difference ($p = 0.02$). McCord (1980) also discovered that the iatrogenic effects were more pronounced when the intervention was more frequent, longer in duration and involved more than one counsellor. While these differences were often small in magnitude, they were all statistically significant (McCord 1978).

Based on these findings, McCord became a pioneering force for increasing attention to, and disarming hostility toward, the possibility of harmful effects arising from crime prevention programmes. The following best captures her position:

> Researchers typically fail to consider whether social programs have adverse effects, looking only for favorable results of treatment. . . . Yet providers of social services do not have a right to harm their clients. Nor do most providers wish to do so. *But the social climate that buries evidence of harm is powerful. That social climate must be changed.* (McCord 2003: 27–8, emphasis added)

Surprisingly, since McCord's influential article, there has been no published research that has attempted to bring together the rich body of scholarship on the CSYS as well as research on iatrogenic effects more generally. We are well aware of the concerns that criminology is becoming far too concerned with present day events and is overlooking the insight that can be gleaned from classic studies, including those with longitudinal designs like the CSYS (see Laub 2004). It is our position that the CSYS still holds relevance to understanding how and why crime prevention programmes cause harm.

We begin with a brief discussion of the background of the CSYS and McCord's early hypotheses for the observed iatrogenic effects of the delinquency prevention experiment. This is followed by a broader discussion of the implications of McCord's work for iatrogenic effects. Specifically, we highlight recent explanations for iatrogenic effects of crime prevention programmes—most notably peer deviancy training, theory failure, implementation failure and heterogeneous treatment effects. We conclude by discussing concrete implications for research, policy and practice.

Background

The CSYS began in 1939 at a time of tremendous innovation to prevent delinquency outside of the formal justice system. The programme was developed by Richard Clarke Cabot, a medical doctor and professor of social ethics and clinical medicine at Harvard University (McCord 1992). Inspired by studies of juvenile delinquency

during the 1930s, including disconcerting recidivism results published by S. Glueck and E. T. Glueck (1934), Cabot set out to study the development of criminal behaviour (McCord 1992). The CSYS was conceived as a treatment programme to prevent delinquency (Powers and Witmer 1951), but also, in the spirit of Cabot's medical background, to test the efficacy of the treatment by using a rigorous experimental design. As noted by Cabot (1940: 143), 'treatment and research objectives [were] closely intertwined.'

The approach to the treatment itself was firmly grounded in developmental insights. For Cabot, the family unit was the foundation of society, and social problems such as juvenile delinquency invariably had roots in this foundation (Cabot 1940). The treatment programme paired youth with adults who could provide positive influences in their lives and 'supplement but not replace what would normally be a satisfactory parent-child relationship' (Cabot 1940: 143). The intent of the treatment was to 'determine whether boys would model themselves after a positive figure who was simply warmly and humanly interested in them' (Bergin 1963: 245).

The study began with 650 boys, aged 5–13 years (median age = 10.5 years), from the two Boston area towns of Cambridge and Somerville. To prevent stigmatization—a chief concern when designing the intervention—both youth with and without behavioural problems (rated as 'difficult' and 'average', respectively) were recommended by local schools, welfare agencies, churches and police (McCord 1978). Following physical examinations, interviews (also of family members) and detailed case histories, the boys were placed into 325 matched pairs and one member of each pair was chosen at random (on the toss of a coin) to be in the treatment group (Powers and Witmer 1951: 7). Boys were matched on a wide range of variables (e.g. grade placement, physical health, neighbourhood, occupational status of father) with the aim of producing 'diagnostic twins' (Cabot 1940: 146). Ultimately, the investigators were interested in achieving greater comparability between the treatment and control groups—in addition to that provided by random assignment.

From 12 November 1937 to 13 May 1939, treatment youth were assigned to counsellors who would support character and personality development through mentoring and provide a 'friendly interest in the boys' problems' (Cabot 1940: 143). Nineteen counsellors (15 men and 4 women) were employed by the study, and the average counsellor had a caseload of 34 boys (Powers and Witmer 1951). While eight counsellors were social workers, there were no training qualifications beyond a 'warm, outgoing' disposition (Powers and Witmer 1951: 92).

The treatment group received individual counselling and home visits. Keeping in close contact with the police, counsellors talked to the boys, took them on trips, engaged them in recreational activities, tutored them in reading and arithmetic, encouraged them to participate in the YMCA and in summer camps, played games with them at the project's centre, encouraged them to attend church and gave advice and general support to the boys' families (W. McCord and J. McCord 1959). While the treatment generally followed Cabot's vision of 'directed friendship', the manner in which this was accomplished varied: '[e]ach counselor was left largely to his own resources' in determining the treatment approach for a particular youth (Powers 1949: 82). The control group received no services.

In 1942, shortly following the United States' entry into World War II, the loss of financial resources (owing to gas shortages) as well as some of the young men who were

counsellors (owing to military service) caused the programme to be scaled back to 253 matched pairs of boys. The programme, ending on 31 December 1945 (Powers 1949), lasted an average of 5 years, during which time treatment youth received an average of two visits per month (McCord 1978).

Follow-ups, carried out in 1948 (Powers and Witmer 1951) and 1956 (W. McCord and J. McCord 1959), indicated that the programme had a null effect on official offending. The first follow-up suggested that the treatment boys committed more offences up to 1948, but the significance of this result was not tested (Powers and Witmer 1951: 326). The later follow-up suggested that the control boys committed more crimes, but the difference was not statistically significant (W. McCord and J. McCord 1959: 92).

The next follow-up was carried out by Joan McCord in 1978, 30 years post-intervention (mean age = 45 years). It is important to note that this follow-up was influenced by major prospective longitudinal studies of the day, including West and Farrington's (1977) Cambridge Study in Delinquent Development, which involves 411 London boys born in the 1950s. The CSYS fits squarely in the global tradition of longitudinal research on offending over the life-course (see also Moffitt *et al.* 2001; Vitaro *et al.* 2013). On the matter of the 30-year follow-up, records were located for 94.9 per cent or 480 of the 506 participants (of whom 48 had died) and interviews were conducted with (or questionnaires distributed to) 347 of them. Comparisons between the treatment and control groups indicated that the treatment group had not fared better on *any* measured outcome, and actually fared worse on seven outcomes. Compared to the controls, treatment group men were more likely to (1) commit more than one crime (among those who committed at least one crime); (2) suffer symptoms of alcoholism; (3) manifest signs of mental illness; (4) die at a younger age; (5) suffer from at least one stress-related disorder, especially high-blood pressure or heart trouble; (6) have occupations with lower prestige and (7) report their work as unsatisfying (McCord 1978).

Various critiques of the study's findings have been advanced over the years. For example, as Short (1954) suggested, it is possible that the observed iatrogenic effects were artefacts of measurement bias. That is, measurement failure could have jeopardized internal, construct or statistical conclusion validity, thereby biasing the strength and significance of estimated coefficients (see Welsh and Rocque 2014). Additionally, Vosburgh and Alexander (1980) were concerned that the study involved implicit self-selection in the treatment design. Counsellors made a series of treatment decisions (such as the frequency of treatment and the services provided) *after* random assignment, thereby threatening internal validity for certain comparisons. Weiss and colleagues (2005) suggested that construct validity was problematic. Specifically, they objected to the inclusion of schizophrenia and bipolar disorder in the measure of undesirable outcomes. While criminal offending and alcoholism may be valid measures of negative outcomes affected by the CSYS, Weiss *et al.* (2005) argued that mental illness could not be influenced by a social intervention programme and therefore should not have been included in the outcome construct.

McCord had her own ideas about why the programme produced harmful results, proposing four hypotheses in her 1978 article. In later works (McCord 1980; 1981), she investigated if the data supported any of these hypotheses. This is the focus of the next section.

McCord's Hypotheses

Value conflict hypothesis

"Interaction with adults whose values are different from those of the family milieu may produce later internal conflicts that manifest themselves in disease and/or dissatisfaction." (McCord 1978: 289)

The value conflict hypothesis drew upon research on 'hard core unemployables'. Padfield and Williams (1973) suggested that when certain 'unemployables' were finally employed, it exacerbated their problems rather than solving them because it increased their conflict with middle-class American values. Based on this research, McCord (1981) suggested that the CSYS might have harmed participants by creating a conflict between the values of the counsellors and those of the participants. In this case, the programme's goal of character development would have been at odds with the values in the treatment boys' working-class neighbourhoods of Cambridge and Somerville, making it difficult for the treatment group boys to incorporate middle-class values into their lives.

To assess this hypothesis, McCord (1981: 399) compared various measures of academic achievement, assuming that higher achievement would reflect 'middle-class achievement values'. The only observed statistically significant difference involved high school graduation (the treatment group was less likely to graduate), and there were no significant differences in how often participants mentioned education when discussing their children. Men were also asked more generally about qualities they admired, but again did not differ significantly. Overall, McCord (1981: 400) concluded: 'none of the comparisons indicates that treatment had increased the achievement orientation attributed to holding middle-class values.'

Dependency hypothesis

"Agency intervention may create dependency upon outside assistance. When this assistance is no longer available, the individual may experience symptoms of dependency and resentment." (McCord 1978: 289)

According to this hypothesis, the treatment group may have become dependent on the programme, which resulted in maladjustment when they no longer had access to programme resources. For example, the treatment group boys may have become attached to the positive adult relationships provided by the programme and suffered when these relationships ended at programme termination. In support of this hypothesis, some have argued that welfare programmes may harm recipients by creating a dependency on handouts and preventing the development of self-efficacy (Kane 1987; Kalil and Kunz 1999).

To assess this hypothesis, McCord used measures of 'dependency' constructed from questionnaires collected at the 30-year follow-up. These included how often men asked others for help, how often men visited their parents, whether men were active in clubs and how men spent their leisure time. No significant differences across treatment and control group men emerged for any of these measures.

Labelling hypothesis

"[T]hrough receiving the services of a 'welfare project,' those in the treatment program may have justified the help they received by perceiving themselves as needing help." (McCord 1978: 289)

McCord hypothesized that the CSYS harmed participants by creating a self-perception that they 'needed treatment' because they had been selected for help. Placing youth in the treatment group may have created a negative self-image that undermined any positive effects of the treatment. This possibility—'labelling'—was an explicit concern of CSYS founders, who made it a point to include 'average' as well as 'difficult' boys in the treatment and control groups (Cabot 1940). Despite the precautions taken, McCord compared measures of self-confidence, feelings of competence and psychosomatic disorders across the treatment and control groups to assess the potential for labelling effects. The two groups did not differ significantly in responses, and McCord concluded that the data did not support a labelling explanation (McCord 1981: 401).

Failed expectations hypothesis

"The treatment program may have generated such high expectations that subsequent experiences tended to produce symptoms of deprivation." (McCord 1978: 289)

Lastly, McCord (1981) reasoned that a social intervention could harm participants by creating unrealistic expectations regarding life possibilities. Failed expectations, in turn, would create a sense of disillusionment and diminish satisfaction with everyday life events, thereby precipitating subsequent negative outcomes.

McCord investigated this hypothesis by comparing measures of life satisfaction. Pertaining to marriage, McCord (1981: 401) found that a higher proportion of treatment group men had been separated or divorced than control group men ($p < 0.05$). Additionally, for current marriages, only 47 per cent of treatment participants, compared to 65 per cent of controls, were coded as exhibiting 'warmth' toward their spouses during interviews, a statistically significant difference. The treatment group was also significantly less likely to report being satisfied with work and with life. Ultimately, McCord (1981: 405) concluded: 'The Cambridge-Somerville Youth Study seems to have raised the expectations for its clients without also providing the means of increasing satisfactions. The resulting disillusionment seems to have contributed to the probability of having an undesirable outcome.' Despite this conclusion, relatively little theoretical attention has been paid to whether failed expectations and subsequent disillusionment explain why other crime prevention programmes cause harm (but see Fabricatore *et al.* 2007, in the context of obesity treatment).

Peer Deviancy Training

Contagion as an explanation of the CSYS's iatrogenic effects

Almost two decades after she presented and investigated these initial hypotheses, McCord arrived at a new explanation: the iatrogenic effects of the CSYS were due, in part, to peer contagion effects among a subgroup of treatment youth (Dishion *et al.* 1999). Peer contagion involves a mutual-influence relationship between peers, wherein

the behaviour and emotions of each peer influences the other in ways that potentially cause harm or otherwise undermine the developmental process (Dishion and Tipsord 2011). 'Contagion' is used to describe this process due to its resemblance to the contagious transmission of disease (Loftin 1986).

Peer contagion can occur by a variety of causal mechanisms (Cécile and Born 2009). In the context of the CSYS, McCord proposed the 'peer deviancy training' hypothesis (Dishion *et al.* 1999; McCord 2003). This mechanism of peer contagion involves 'the interpersonal dynamic of mutual influence during which youth respond positively to deviant talk and behavior . . . characterized by give-and-take exchanges between friends that promote deviant actions (e.g., past stories of deviant acts, suggestions for future behavior, what ifs) and elicit positive responses, such as laughter' (Dishion and Tipsord 2011: 190). In other words, peers who communicate about deviancy—real or imagined—reinforce antisocial behaviours and emotions and provide motivation for future deviant behaviour.

Notably, however, peer contagion is not only a matter of normative influence but also a function of opportunities for deviance in unstructured environments (see, e.g. Osgood *et al.* 1996; Osgood and Bridell 2006). That is, socialization may interact with opportunity via 'unstructured socializing' (Haynie and Osgood 2005). This idea comports with recent work by Weerman and colleagues (2013), which suggests that peer deviancy training may be more salient in conjunction with socializing (e.g. hanging out), being in public (i.e. engaging in unstructured activity) and being unsupervised (see also Anderson and Hughes 2009; Bagwell and Schmidt 2011).

To assess whether deviancy training could explain the iatrogenic effects of the CSYS, McCord considered summer camp participation by the treatment group boys (Dishion *et al.* 1999). Roughly half were sent to summer camp (125/253), and approximately half of these boys went to camp more than once (66/125). In theory, these camps allowed for a great deal of unstructured socializing, representing an ideal environment for deviancy training to take place.

Re-analyzing the 30-year follow-up data, McCord (2003) found that for boys who were sent to summer camp only once (n = 59), the odds ratio predicting undesirable outcomes was 1.33, which was significantly higher than the odds ratio of 1.12 among treatment group boys who did not attend summer camp. The truly stark result, however, involved treatment group boys who attended camp more than once (n = 66). For this group, the odds ratio for undesirable outcomes was 10.0, meaning that participants were ten times as likely to experience undesirable outcomes as their matched pairs. McCord (2002: 235) concluded: 'I strongly suspect that the boys from the Youth Study tended to bond together, encouraging one another's deviant values.' McCord's construct theory of motivation—which argues that youth respond to situational cues in constructing motivation based on the way they perceive the choices of others—provided a theoretical explanation for why deviancy training takes place among high-risk youth in unsupervised settings (McCord 2003; 2004).

Since McCord's research, a robust literature on peer deviancy training has emerged (Poulin *et al.* 2001; Dishion 2000; Dishion and Dodge 2005; Dodge *et al.* 2006; Dishion and Tipsord 2011). As a whole, the empirical literature suggests that unrestricted interaction between low- and high-risk youth contributes to higher rates of deviant behaviour among low-risk youth (Haynie and Osgood 2005; Dishion and Tipsord 2011). These effects can be observed as early as the first grade of elementary school (Gifford-Smith *et al.* 2005) and persist over time (Dishion and Tipsord 2011).

Several meta-analyses and systematic reviews also provide support for the peer contagion hypothesis. For example, Lipsey and colleagues (2000) conducted a review of juvenile delinquency intervention programmes and found that the most successful programmes were those in which delinquent youth did not have the opportunity for unstructured interaction with deviant peers. In addition, a recent review of (both juvenile and adult) programmes with iatrogenic effects indicated that almost three-quarters (14/19) of programmes with harmful effects occurred in group settings, as compared to individual settings (Welsh and Rocque 2014). Despite these findings, whether deviancy training actually undermines group interventions and causes harm remains the subject of much controversy (Rhule 2005).

Challenges to establishing peer deviancy effects

Some argue that group-based interventions are a cost-effective alternative to individualized treatment (Kaminer 2005; Greenwood 2006) and only cause harm under certain conditions, e.g., in the absence of parental involvement and adequate supervision (Dishion *et al.* 1999; Kaminer 2005; Rhule 2005; Dishion and Tipsord 2011). Others question the causal relationship between group treatment and iatrogenic effects (see Weiss *et al.* 2005). The contention, which is the most notable objection to peer deviancy effects, is that negative group treatment effects are not actually treatment effects but rather represent selection effects under which deviant peers associate due to homophily (Arnold and Hughes 1999; Weiss *et al.* 2005). Failing to account for selection effects may lead one to conclude that deviant peers are driving personal delinquency (i.e. socialization), when in reality personal delinquency precedes deviant peer associations (see Steglich *et al.* 2010). This is especially true if high-risk youth are selected for group treatment, since these youth are likely to affiliate with deviant peers both in and out of the treatment context.

In the case of the CSYS, the selection of treatment group boys for summer camp was based on counsellor discretion rather than on random assignment (Arnold and Hughes 1999; Gottfredson 2010), making it impossible to distinguish selection from socialization effects. To be fair, McCord appeared eminently aware of this potential objection: 'One caution of [the peer-deviancy] interpretation is that youth self-selected into summer camp experiences; because their matched controls did not make a similar selection, the intervention group may be biased toward the deviance in an unknown way' (Gifford-Smith *et al.* 2005: 261). But, this quandary need not plague all studies of group interventions, since research designs (e.g. focusing on the transmission of specific types of crime; see Bayer *et al.* 2004; Mennis and Harris 2011) and statistical techniques (see Weerman 2011; Young 2011) can separate selection and socialization effects.

These arguments aside, even if peer aggregation does lead to iatrogenic effects in the group treatment context, it may not follow that deviancy training is the responsible causal mechanism (Weiss *et al.* 2005). Critics argue that while deviancy training may be correct as a more general explanation for delinquency—that is, outside of treatment—peer influence during group treatment is minimized via restricted exposure time. This leaves room for theorizing about other causal mechanisms responsible for iatrogenic effects in group-based crime prevention programmes.

Theory, Implementation and Heterogeneous Treatment Effects

While the explanations for the CSYS findings discussed above focus on aspects of the treatment that may have had unintended, harmful consequences, we should not overlook potential explanations that focus on the treatment programme itself: the underlying programme theory (theory failure), the way the programme was designed (design failure), and the way it was implemented (implementation failure). Each of these represents a commonly considered explanation for null effects, but they also provide an important perspective from which we can interpret iatrogenic effects.

Theory failure

The first step to planning an effective treatment programme involves a solid theoretical foundation on which to base the treatment (Rossi *et al.* 2004). Most importantly, this should include a theory of programme change: how will the treatment elicit change in the treatment group? The importance of programme theory has led to developments in the evaluation literature such as logic models, which describe how a programme works (or is supposed to work) in terms of inputs, resources, activities, outputs, outcomes and long-term impact (Mertens and Wilson 2012). In the context of crime prevention, theory failure is likely responsible for the ineffectiveness of Scared Straight programmes, boot camps and intensive supervision probation (Welsh and Rocque 2014; see also Sherman 2007).

In the case of the CSYS, the question is whether it was based on a coherent theoretical approach to preventing delinquency. And, if a sound theory was absent, what was the implicit logic of the programme? As others have noted, interventions tend to have some theoretical basis—whether it is articulated or not (Pawson and Tilley 1997; Mears 2007). For crime prevention efforts to be successful, it is necessary for this theoretical basis to be rationally related to the development of criminal offending. It is thus important to assess the rationale underlying the CSYS and examine whether partial theory failure could be responsible for its long-term iatrogenic effects.

Friendly mentoring is not a therapeutic treatment

Cabot's vision for the CSYS was to use friendly mentoring from positive role models to supplement the parent-child relationship and provide positive guidance for character and personality development (Cabot 1940). Cabot believed that 'the thing that "keeps any of us straight" is "the contagion of the highest personalities whom we have known"', and that providing an 'ego-ideal' to troubled youth would effectively prevent delinquency (Short 1954). While this may seem intuitively plausible, others have since suggested that it does not present an effective therapeutic model to combat delinquency (Tremblay 2005). As MacKenzie (2013: 3) observed, 'since the McCord study, we have learned a great deal about what is effective in reducing criminal activities of delinquents and offenders. "Friendly understanding" programmes do not address the criminogenic deficits of these individuals and, thus, are doomed to failure.'

Elsewhere, the CSYS has been criticized for not presenting a clear therapy model. Weiss *et al.* (2005: 1038) suggested that the CSYS treatment programme—consisting mostly of case management and mentoring—would 'not be considered a potentially

efficacious treatment' by current standards in clinical psychology. Even in its day, it was observed that the CSYS could not 'be truly said to have been what most of us would term psychotherapy' (Bergin 1963: 245). In the end, friendly mentoring may not have represented a therapeutically sound treatment model, echoing an early suggestion that the lesson of the CSYS findings is that 'good will is not sufficient' (Wrenn 1952: 215).

The compensatory model is a critical error

Despite the important observation that the CSYS did not have a clear therapeutic approach, the treatment is perhaps best described as individual-directed—because it focussed explicitly on the individual child as opposed to the child's social environment—and compensatory—because it sought to compensate for perceived deficits in the child's development. However, both of these aspects of the treatment approach may lead to partial theory failure.

First, individual-level treatment may not present an effective approach to preventing delinquency without being grounded in a broader ecological approach (Short 1954). As Lundman (2001) suggests, such an individual treatment model is premised on the theoretical assumption that delinquency is an individualized phenomenon as opposed to the result of social forces. Yet this assumption may be too narrow to deal effectively with the complex social forces (school, peer and neighbourhood factors) in a child's life (Shirk *et al.* 2000).

Second, the compensatory approach to reducing delinquency may itself be 'a critical error' (McCord 1992: 203). The underlying theory of the CSYS seemed to be that at-risk youths needed positive attention from adults because such attention had been largely missing from their lives. Although parental attention (such as supervision and monitoring) may reduce the likelihood of delinquency for a young child, it may have the opposite effect for an at-risk or delinquent child who has already experienced these deficits:

> "Why did the treatment have harmful effects? Part of the reason, it seems to me, has been the compensatory model on which the treatment was based. Cabot—and many others—have assumed that an appropriate treatment would undo deficits in backgrounds of people at high risk for developing problems. This can be a critical error. A child rejected by parents may not be best served by someone else who tries to take the role of a parent. Such a strategy might result in an exaggerated sense of loss; it might produce expectations for or dependence on assistance. . . . [A] child who has not been supervised may become *more* antisocial if he is placed under close supervision." (McCord 1992: 203)

Because the compensatory model was grounded in assumptions similar to Hirschi's (1969) control theory (e.g. providing conventional bonds), the failure of the compensatory model provided 'grounds for doubting the adequacy of control theory as an explanation for crime' (McCord 1992: 204). Consequently, it may not be surprising that McCord (2004) later developed the construct theory of motivation to explain delinquency more generally—a social learning rather than control theory of criminal offending.

Implementation failure

Even if the theoretical foundations of the CSYS were not responsible for the programme's failure, it is possible that the implementation of Cabot's vision was flawed.

Implementation failure occurs when a sound theory is not properly executed. As others have noted, implementation failure is perhaps not as serious a problem as theory failure because it does not call the basic idea or mechanism behind the programme into question (Welsh and Rocque 2014). But, the literature suggests that poorly trained therapists and dosage effects are the most consistent explanations for iatrogenic effects (Lilienfeld 2007).

Early observers of the CSYS considered the theoretical foundations of the programme robust but the actual clinical approach—the implementation of the treatment—unfocussed (Lowry 1940). Perhaps most importantly, there was no formal training of CSYS counsellors to ensure treatment integrity, and, as a result, the treatment delivered by counsellors was often unclear and inconsistent. As Barlow (2010: 16) explained, the implementation of the CSYS involved 'instructing 10 therapists with no formal training to do whatever they thought best over a minimum of five sessions per year for up to five years with pre delinquent boys.' Such instructions do not provide programme fidelity or yield a clear treatment model that can be easily evaluated (Vosburgh and Alexander 1980; Barlow 2010). Indeed, W. McCord and J. McCord (1959: 32) observed that the CSYS may have failed to produce positive results because the counsellors did not have 'intimate, intensive' relationships with most of the treatment boys as dictated by Cabot's treatment model.

In addition, dosage effects may have contributed to implementation failure (Mears 2007). Just as medical treatment requires proper dosage (where too little is ineffective and too much is harmful), social intervention programmes necessitate a specific treatment amount. As discussed above, McCord (1978) found that treatment frequency varied considerably among youth and that negative outcomes were associated with more frequent treatment. While this seems contrary to a traditional dosage effect, an alternative explanation would be that while a moderate amount of intervention is more harmful than little-to-no intervention, frequent intervention (not present in the CSYS) would have been more effective. Given that the most frequent home visits in the CSYS were twice per month, it does not appear that any CSYS participants received treatment that could be characterized as intensive (W. McCord and J. McCord 1959: 32).

To date, the most effective treatment prevention programmes for delinquent youth demonstrate a strong programme model (theory) and fidelity to that model (implementation) (Rhule 2005). It is possible that the CSYS was lacking in one, or both, of these elements.

Heterogeneity of treatment effects

While these design and implementation issues warrant critical attention, there may be an even more plausible explanation for why the outcomes were iatrogenic rather than null: the variability of treatment and resulting heterogeneity of treatment effects. Such concerns have been raised in other contexts. In psychotherapy, Bergin (1963) argued that many (if not most) null results can actually be 'accounted for by the mutually cancelling effects of two different kinds of therapists having provided treatment—one kind apparently promotes positive change and the other kind promotes negative change' (Bergin 1963: 247). On closer observation, Bergin observed that therapist and client characteristics often interact with treatment to produce wide variation in outcomes. According to Bergin, this variation is greater for treatment groups than for

controls. For example, based on a careful review of seven unrelated studies with null results, Bergin (1966) found a significant difference in the variability of change across treatment and control groups, despite the average null effect. While treatment and control groups shared 'no change' and 'spontaneous change' (positive) effects over time, the treatment group was subject to an additional 'double-edged effect', what he called 'therapeutic change' and 'deterioration' (Bergin 1966: 237). This theory was named 'the deterioration effect' for its provocative suggestion that psychotherapy could actually harm some clients (Barlow 2010).

Importantly, this finding did not call into question psychotherapy itself, as the potential for a deterioration effect is accompanied by the potential for an improvement effect. Instead, this finding suggested that therapy is 'powerful and, like any discovery in nature, it can have beneficial or harmful effects' (Bergin 1970: 301), implying a need to understand what makes different treatments effective, ineffective and even harmful (Bergin 1966).

Outside of psychotherapy, it has been suggested that the very concept of 'average treatment effect' implies systematic variation in treatment effects (Gaines and Kuklinski 2011). For example, in their initial follow-up of the CSYS, Powers and Witmer (1951) suggested that the null effect of the programme seemed to be masking variation in effects across different groups of youth in the study. Specifically, the most at-risk youth appeared to have the most negative outcomes, while youth with strong family and community support seemed to benefit from the programme. It may be that more heterogeneous samples have the effect of 'increasing the variance in subject response to intervention and reducing statistical power to detect truly powerful effects on some subgroups but not others' (Sherman 2007: 304). As McCord (1993: 413–4) suggested, 'Heterogeneity in samples can mask relationships if the direction of relationships within subsamples are opposite in direction.'

Sample heterogeneity is closely related to theory and implementation failure. For example, the CSYS may have been ineffective as a whole due to theory or implementation failure, and the findings of harm may have represented iatrogenic effects of a subset of the study. In other words, a small subgroup could have experienced a deterioration effect—perhaps the summer camp youth due to peer deviancy training—that would push the results of the otherwise ineffective programme—due to theory and/or implementation failure—into the red, producing an average iatrogenic effect for the treatment group.

Even with this understanding, the peer deviancy thesis would not tell the whole story. Following the logic of Bergin (1963), had the CSYS been more effective for other subgroups, this should have cancelled out the deterioration effects of the summer camp group. Given the heterogeneity of the CSYS programme, there are many subgroup analyses that can be performed, and the matched-pair design allows for this to be done in a highly rigorous manner. For all discordant pairs, we have grounds to investigate cases where the treatment men did better (improvement effect) and where the treatment men did worse (deterioration effect).

Another possible implication of heterogeneous effects is that the quantitative, nomothetic approach of seeking generalizable findings should be complemented with a more idiographic approach (such as case study) that focuses on how individual cases are affected by treatment. While a nomothetic approach tends to treat variation as a function of sampling and focuses on average between-subject effects, qualitative

supplementation may increase understanding of the complex variation in outcomes by focusing on intra-subject variability (Barlow 2010). This speaks directly to Pawson and Tilley's (1997) 'scientific realist' approach to evaluation, which calls for a greater understanding of the mechanisms, contexts and conditions associated with observed effects. In the words of Pawson and Tilley (1994: 292, emphasis in original), 'we need to know why and in what circumstances programmes affect potential subjects *before* we can begin to say if they 'work'.'

For the CSYS, treatment may have helped some, hurt others, and had no effect for many more. It is necessary to examine patterns of variability of effects in the data through subgroup analysis as well as through the qualitative assessment of outcomes. The CSYS provides an important lesson for evaluating crime prevention programmes today: 'meaningful discussion of iatrogenic effects must consider these sources of variation (e.g., implementation quality, program focus and strategies, sample and subgroup differences) in the youths' responses to intervention' (Rhule 2005: 621).

Discussion and Conclusions

This article reviewed the proffered explanations for the iatrogenic effects of the Cambridge-Somerville Youth Study. In addition to McCord's hypotheses, others have weighed in on the subject, providing a rich body of scholarship upon which to draw. It also presented an opportunity for us to draw upon a wider body of research that has investigated iatrogenic effects of social interventions and apply this knowledge to the study. The many stand-alone and interdependent explanations for the CSYS's iatrogenic effects illustrate the difficulty of ascertaining why the programme was harmful. We proposed that the heterogeneity of treatment and treatment effects might have resulted in small overall positive or negative effects via subgroups that were especially helped or harmed. Ultimately, this suggests that there is not a grand explanation for the programme's iatrogenic effects, a lesson that may apply to crime prevention efforts more generally.

The oft-ignored possibility of iatrogenic effects of social interventions carries important implications for research, policy and practice. First, while it would be easy enough to dismiss a study with null or negative effects and simply 'move on', studying why a social intervention did not work—or even caused harm—is a critical component of connecting sound developmental theory with prevention science (Poulin *et al.* 2001). In our immediate context, it has been observed that 'crime policy on the whole lacks a rational foundation' (Mears 2007: 679). While one important response to this problem has been a focus on 'what works' in preventing crime, this may ignore important lessons that can be gleaned from programmes that fail to work or even cause harm. The focus on positive results is too narrow, 'because it fails to recognize that some treatments cause harm', such as increased offending, increased substance abuse or a general inability to 'cope with life' (McCord 2003: 17). This can lead to an under-appreciation of what is truly at stake in designing social interventions.

A second, related implication is that prevention programmes that are not grounded in sound criminological theory are unlikely to succeed and may do more harm than good. Determining the theoretical basis for why some programmes cause harm represents an important consideration when designing future social interventions, both to diminish risk of harm and to allow for a more nuanced understanding of intervention effects. To

take a more recent example, Scared Straight was developed on the theoretical premise that 'delinquency could be prevented by giving wild youngsters a taste of what it would be like to be imprisoned' (McCord 2003: 26). Without any scientifically credible evidence that such a programme would actually work, Scared Straight was adopted across the United States in 38 states—due largely to endorsement by judges and politicians along with a film that popularized the 'get tough' approach. In their systematic review of Scared Straight evaluations, however, Petrosino and colleagues (2003) found harmful effects for youth who participated in the programme. As McCord (2003) observed, if the theory behind a social intervention is erroneous—here, that at-risk youth can be 'scared' into non-deviant behaviour—then there is little reason to expect such programmes to work. In the case of Scared Straight, exposing youth to adult prisoners seems to have romanticized the prisoners' lives and thereby encouraged further offending, creating a 'delinquency fulfilling prophecy' (Finckenauer 1982: 169).

In a similar vein, evaluations of boot camp correctional programmes have shown them to be generally ineffective in reducing re-offending (Wilson *et al.* 2005). While aspects of boot camp programmes may borrow from social learning and social control theories, the central component is militaristic: strict discipline and rigorous physical activity are utilized to rehabilitate offenders. This get tough approach may be politically appealing, but it does not represent a 'coherent theoretical model' for reducing crime and delinquency (MacKenzie 2006: 279). Too often, prevention and intervention programme designs are based on popular but erroneous notions that 'draw loosely if at all on mainstream criminological theories' (Mears 2007: 671). Yet it is clear that 'in the absence of sound theory, social programs stand little chance to bring about social good and may even cause harm' (Welsh and Rocque 2014: 262).

These concerns are not limited to punitively oriented programmes. Another important implication for designing crime prevention programmes is that aggregating high-risk participants in groups—like the CSYS youth who attended summer camp—may have unintended consequences. Group interventions are a popular approach in crime prevention, because they represent a cost-effective alternative to individualized treatment (Greenwood 2006; Kaminer 2005). However, the possibility of increasing offending by grouping delinquent peers should not be taken lightly. One example is the Adolescent Transitions Program, by Poulin and colleagues (2001). Unlike Scared Straight and correctional boot camps, the intervention appeared to be based on a theoretically coherent approach to reducing delinquency and other antisocial behaviours: to promote self-monitoring, pro-social goal-setting and communication skills while developing peer environments supportive of such pro-social behaviour. The authors found that, compared to controls, treatment group youth were significantly more likely to have higher rates of self-reported smoking and teacher-reported delinquency three years later. Moreover, these iatrogenic outcomes were especially stark for treatment youth with initially low levels of delinquency. The authors concluded that grouping high-risk youth appeared to undermine the treatment and cause long-term deleterious effects, leading them to advocate that social interventions should rigorously assess the risks posed by a proposed design before implementation due to the possibility of unintended, iatrogenic effects.

Yet another key implication is that, in addition to sound theory, incorporating experimental designs into crime prevention programmes may be necessary to evaluate their true impact (McCord 2003). Despite the failure of the CSYS to prevent delinquency,

McCord (1992) insisted that there were several reasons why it should still be considered a success, including the importance of random assignment to assess the validity of findings. After all, the treatment group did fare better than expected. It was only in comparison to the control group that the treatment group appeared harmed (McCord 1978). As Sherman (2007: 300) observed, if only experimental conditions can reliably detect (or rule out) iatrogenic effects, then the advancement of experimental criminology is tied to the social obligation to prevent harm:

> The potential for any program to cause harm, no matter how sensible it may appear in theory, remains the primary ethical justification for experimental criminology … Absent any other widely accepted means of creating unbiased estimates of treatment effects …, the only way we can be sure that treatments do not harm people is to subject those practices to randomized controlled experiments in field settings. One promise of such tests is that they will guide democratic societies in deciding what *not* to do, as one way to reduce human suffering.

In addition to these ethical considerations, experiments 'may be the best way to test theories of the causes of crime' due to the elimination of various confounding influences (Tremblay and Craig 1995: 153; see also McGloin and Thomas 2013). As McCord's later research attests (e.g. McCord 1991*a*; 1991*b*; 1994), part of the value of the CSYS is that, as a longitudinal-experimental design, it allows for the testing of causal hypotheses within the context of the natural history of the development of offending (Farrington 2006; Farrington *et al.* 2010).

One other key implication for research and policy is the importance of transparency of evaluation results. A recent review of 15 Campbell Collaboration systematic reviews of crime prevention programmes found 22 harmful effects from 22 unique studies out of a pool of 574 experimental and quasi-experimental studies—approximately 3.8 per cent. While this may seem small, such a risk of harm is 'by no means trivial' (Welsh and Rocque 2014: 258), particularly if crime prevention must first 'do no harm'. More disturbing is that two-thirds of these harm-causing studies were unpublished, substantiating McCord's concern that evidence of harm is often 'buried' (McCord 2003: 28).

The resistance to publishing null or negative findings is so commonplace today that unexpected results have been referred to as 'Type-III' errors by some statisticians (Barlow 2010: 17). Such strong resistance should be jettisoned by the research community and transparency of findings must become a greater priority. As Rhule (2005: 622) has suggested, 'the research and clinical community should commend those who share the negative results of their own interventions, recognizing such a disclosure as a service to the field.' In one of her final publications, McCord (2003) proposed that evaluation results—regardless of the findings—be made available through a centralized data bank to combat publication bias and to enhance transparency. This has begun with the Campbell Collaboration's library of systematic reviews.

In order for future crime prevention efforts to be successful, we must learn from the lessons of past programmes. One such lesson is that well-intentioned social interventions can produce iatrogenic effects. While the CSYS is not the only crime prevention programme to produce iatrogenic effects, it does represent one of the most enduring examples of how well-intentioned programmes can do harm: 'No study in the history of criminology has ever demonstrated such clear, unintended, criminogenic effects of a program intended to prevent crime. To this day, it is 'exhibit A' in discussions with legislators, students, and others skeptical of the value of evaluating government

programs of any sort, let alone crime prevention programs' (Sherman 2005: 124). We have observed that crime prevention research tends to focus on positive results. While choosing to study programmes that cause harm may be contrary to this focus, doing so is necessary for sound policy and practice. We could not present the stakes any better than did McCord (2003: 17): '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.'

References

Anderson, A. L. and Hughes, L. A. (2009), 'Exposure to Situations Conducive to Delinquent Behavior', *Journal of Research in Crime and Delinquency*, 46: 5–34.

Arnold, M. E. and Hughes J. N. (1999). 'First Do No Harm: Adverse Effects of Grouping Deviant Youth for Skills Training', *Journal of School Psychology*, 37: 99–115.

Bagwell, C. L. and Schmidt, M. E. (2011), *Friendships in Childhood and Adolescence*. Guilford Press.

Barlow, D. H. (2010), 'Negative Effects from Psychological Treatments: A Perspective', *American Psychologist*, 65: 13–20.

Bayer, P., Hjalmarsson, R. and Pozen, D. (2004), 'Building Criminal Capital Behind Bars', *Quarterly Journal of Economics*, 124: 105–47.

Bergin, A. E. (1963), 'The Effects of Psychotherapy: Negative Results Revisited', *Journal of Counseling Psychology*, 10: 244–9.

——. (1966), 'Some Implications of Psychotherapy Research for Therapeutic Practice', *Journal of Abnormal Psychology*, 71: 235–46.

——. (1970), 'The Deterioration Effect: A Reply to Braucht', *Journal of Abnormal Psychology* 75: 300–2.

Cabot, P. S. deQ. (1940), 'A Long-Term Study of Children: The Cambridge-Somerville Youth Study', *Child Development*, 11: 143–51.

Cécile, M. and Born, M. (2009), 'Intervention in Juvenile Delinquency: Danger of Iatrogenic Effects?', *Children and Youth Services Review*, 31: 1217–21.

Dishion, T. J. (2000), 'Cross-Setting Consistency in Early Adolescent Psychopathology: Deviant Friendships and Problem Behavior Sequelae', *Journal of Personality*, 68: 1109–26.

Dishion, T. J. and Dodge, K. A. (2005), 'Peer Contagion in Interventions for Children and Adolescents: Moving Towards an Understanding of the Ecology and Dynamics of Change', *Journal of Abnormal Child Psychology*, 33: 395–400.

Dishion, T. J. and Tipsord, J. M. (2011), 'Peer Contagion in Child and Adolescent Social and Emotional Development', *Annual Review of Psychology*, 62: 189–214.

Dishion, T. J., McCord, J. and Poulin, F. (1999), 'When Interventions Harm: Peer Groups and Problem Behavior', *American Psychologist*, 54: 755–64.

Dodge, K. A., Dishion, T. J. and Lansford, J. E. (2006), 'Deviant Peer Influences in Intervention and Public Policy for Youth', *Social Policy Report*, 20: 1–20.

Fabricatore, A. N., Wadden, T. A., Womble, L. G., Sarwer, D. B., Berkowitz, R. I., Foster, G. D. and Brock, J. R. (2007), 'The Role of Patients' Expectations and Goals in the Behavioral and Pharmacological Treatment of Obesity', *International Journal of Obesity*, 31: 1739–45.

Farrington, D. P. (2006), 'Key Longitudinal-Experimental Studies in Criminology', *Journal of Experimental Criminology*, 2: 121–41.

Farrington, D. P., Loeber, R. and Welsh, B. C. (2010), 'Longitudinal-Experimental Studies', in A. R. Piquero and D. Weisburd, eds., *Handbook of Quantitative Criminology*, 503–18. Springer.

Finckenauer, J. O. (1982), *Scared Straight and the Panacea Problem*. Prentice-Hall.

Gaines, B. J. and Kuklinski, J. H. (2011), 'Experimental Estimation of Heterogeneous Treatment Effects Related to Self-Selection', *American Journal of Political Science*, 55: 724–36.

Gifford-Smith, M., Dodge, K. A., Dishion, T. J. and McCord, J. (2005), 'Peer Influence in Children and Adolescents: Crossing the Bridge from Developmental to Intervention Science', *Journal of Abnormal Child Psychology*, 33: 255–65.

Glueck, S. and Glueck, E. T. (1934), *One Thousand Juvenile Delinquents*. Harvard University Press.

Gottfredson, D. C. (2010), 'Deviancy Training: Understanding How Preventative Interventions Harm', *Journal of Experimental Criminology*, 6: 229–43.

Greenwood, P. W. (2006), *Changing Lives: Delinquency Prevention as Crime-Control Policy*. University of Chicago Press.

Haynie, D. L. and Osgood, D. W. (2005), 'Reconsidering Peers and Delinquency: How Do Peers Matter?' *Social Forces*, 84: 1109–30.

Hirschi, T. (1969), *Causes of Delinquency*. University of California Press.

Kalil, A. and Kunz, J. (1999), 'Self-Esteem, Self-Efficacy, and Welfare Use', *Social Work Research*, 23: 119–26.

Kaminer, Y. (2005), 'Challenges and Opportunities of Group Therapy for Adolescent Substance Abuse: A Critical Review', *Addictive Behaviors*, 30: 1765–74.

Kane, T. J. (1987), 'Giving Back Control: Long-Term Poverty and Motivation', *Social Service Review*, 61: 405–19.

Laub, J. H. (2004), 'The Life Course of Criminology in the United States', *Criminology*, 42: 1–26.

Lilienfeld, S. O. (2007), 'Psychological Treatments That Cause Harm', *Perspectives on Psychological Science*, 2: 53–70.

Lipsey, M. W., Wilson, D. B. and Cothern, L. (2000), *Effective Intervention for Serious Juvenile Offenders*. Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice.

Loftin, C. (1986), 'Assaultive Violence as a Contagious Social Process', *Bulletin of the New York Academy of Medicine*, 62: 550–5.

Lowry, L. G. (1940), 'Special Comment: The Cambridge Somerville Youth Study', *American Journal of Orthopsychiatry*, 19: 611–2.

Lundman, R. L. (2001), *Prevention and Control of Juvenile Delinquency*. Oxford University Press.

MacKenzie, D. L. (2006), *What Works in Corrections: Reducing the Criminal Activities of Offenders and Delinquents*. Cambridge University Press.

——. (2013), 'First Do No Harm: A Look at Correctional Policies and Programs Today', *Journal of Experimental Criminology*, 9: 1–17.

McCord, J. (1978), 'A Thirty-Year Follow-Up of Treatment Effects', *American Psychologist*, 33: 284–9.

——. (1980), 'The Treatment that Did Not Help', *Social Action and the Law*, 5: 85–7.

——. (1981), 'Consideration of Some Effects of a Counseling Program', in S. E. Martin, L. B. Sechrest and R. Redner, eds., *New Directions in the Rehabilitation of Criminal Offenders*, 394–405. National Academy Press.

——. (1991*a*), 'Family Relationships, Juvenile Delinquency, and Adult Criminality', *Criminology*, 29: 397–417.

——. (1991*b*), 'The Cycle of Crime and Socialization Practices', *Journal of Criminal Law & Criminology*, 82: 211–28.

——. (1992), 'The Cambridge-Somerville Youth Study: A Pioneering Longitudinal Experimental Study of Delinquency Prevention', in J. McCord and R. E. Tremblay, eds., *Prevention of Antisocial Behavior: Interventions from Birth through Adolescence*, 196–206. Guilford Press.

——. (1993), 'Descriptions and Predictions: Three Problems for the Future of Criminological Research', *Journal of Research in Crime and Delinquency*, 30: 413–26.

——. (1994), 'Family Socialization and Antisocial Behavior: Searching for Causal Relationships in Longitudinal Research', in E. G. M. Weitekamp and H. -J. Kerner, eds., *Cross-National Longitudinal Research on Human Development and Criminal Behavior*, 177–88. Kluwer.

——. (2002), 'Counterproductive Juvenile Justice', *Australian and New Zealand Journal of Criminology*, 35: 230–7.

——. (2003), 'Cures that Harm: Unanticipated Outcomes of Crime Prevention Programs', *Annals of the American Academy of Political and Social Science*, 587: 16–30.

——. (2004), 'Toward a Theory of Criminal Responsibility', in J. McCord, ed., *Beyond Empiricism: Institutions and Intentions in the Study of Crime*, 147–76. Transaction Publishers.

McCord, W. and McCord, J. (1959), *Origins of Crime: A New Evaluation of the Cambridge-Somerville Youth Study*. Columbia University Press.

McGloin, J. M. and Thomas, K. J. (2013), 'Experimental Tests of Criminological Theory', in B. C. Welsh, A. A. Braga and G. J. N. Bruinsma, eds., *Experimental Criminology: Prospects for Advancing Science and Public Policy*, 15–42. Cambridge University Press.

Mears, D. P. (2007), 'Towards Rational and Evidence-Based Policy', *Journal of Criminal Justice*, 35: 667–82.

Mennis, J. and Harris, P. (2011), 'Contagion and Repeat Offending Among Urban Juvenile Delinquents', *Journal of Adolescence*, 34: 951–63.

Mertens, D. M. and Wilson, A. T. (2012), *Program Evaluation Theory and Practice: A Comprehensive Guide*. Guilford Press.

Moffitt, T. E., Caspi, A., Rutter, M. and Silva, P. A. (2001), *Sex Differences in Antisocial Behavior*. Cambridge University Press.

Osgood, D. W. and Bridell, L. (2006), 'Peer Effects in Juvenile Justice', in K. A. Dodge, T. J. Dishion and J. Lansford, eds., *Deviant Peer Influences in Programs for Youth: Problems and Solutions*, 141–61. Guilford Press.

Osgood, D. W., Wilson, J. K., O'Malley, P. M., Bachman, J. G. and Johnston, L. D. (1996), 'Routine Activities and Individual Deviant Behavior', *American Sociological Review*, 61: 635–55.

Padfield, H. and Williams, R. (1973), *Stay Where You Were: A Study of Unemployables in Industry*. Lippincott.

Pawson, R. and Tilley, N. (1994), 'What Works in Evaluation Research?' *British Journal of Criminology*, 34: 291–306.

——. (1997), *Realistic Evaluation*. Sage.

Petrosino, A., Turpin-Petrosino, C. and Buehler, J. (2003), '"Scared Straight" and Other Juvenile Awareness Programs for Preventing Juvenile Delinquency: A Systematic Review of Randomized Experimental Evidence', *Annals of the American Academy of Political and Social Science*, 589: 41–62.

Poulin, F., Dishion, T. J. and Burraston, B. (2001), '3-Year Iatrogenic Effects Associated with Aggregating High-Risk Adolescents in Cognitive Behavioral Preventative Interventions', *Applied Developmental Science*, 5: 214–24.

Powers, E. (1949), 'An Experiment in the Prevention of Delinquency', *Annals of the American Academy of Political and Social Science*, 261: 77–88.

Powers, E. and Witmer, H. L. (1951), *An Experiment in the Prevention of Delinquency: The Cambridge-Somerville Youth Study*. Columbia University Press.

Rhule, D. M. (2005), 'Take Care to Do No Harm: Harmful Interventions for Youth Problem Behavior', *Professional Psychology: Research and Practice*, 36: 618–25.

Rossi, P. H., Lipsey, M. W. and Freeman, H. E. (2004), *Evaluation: A Systematic Approach*, 7th edn. Sage.

Sherman, L. W. (2005), 'The Use and Usefulness of Criminology, 1751–2005: Enlightened Justice and Its Failures', *Annals of the American Academy of Political and Social Science*, 600: 115–35.

——. (2007), 'The Power Few: Experimental Criminology and the Reduction of Harm', *Journal of Experimental Criminology*, 3: 299–321.

Shirk, S., Talmi, A. and Olds, D. (2000), 'A Developmental Psychopathology Perspective on Child and Adolescent Treatment Policy', *Development and Psychopathology*, 12: 835–55.

Short, J. F. (1954), 'An Experiment in the Prevention of Delinquency: The Cambridge-Somerville Youth Study by Edwin Powers; Helen Witmer', *American Journal of Sociology*, 59: 587–8.

Steglich, C., Snijders, T. A. B. and Pearson, M. (2010), 'Dynamic Networks and Behavior: Separating Selection from Influence', *Sociological Methodology*, 40: 329–93.

Tremblay, R. E. (2005), 'Towards an Epigenetic Approach to Experimental Criminology', *Journal of Experimental Criminology*, 1: 397–415.

Tremblay, R. E. and Craig, W. M. (1995), 'Developmental Crime Prevention', *Crime and Justice*, 19: 151–236.

Vitaro, F., Brendgen, M., Giguère, C. -É. and Tremblay, R. E. (2013), 'Early Prevention of Life-Course Personal and Property Violence: A 19-Year Follow-Up of the Montreal Longitudinal-Experimental Study (MLES)', *Journal of Experimental Criminology*, 9: 411–27.

Vosburgh, W. W. and Alexander, L. B. (1980), 'Long-Term Follow-Up as Program Evaluation: Lessons from McCord's 30-Year Follow-Up of the Cambridge-Somerville Youth Study', *American Journal of Orthopsychiatry*, 50: 109–24.

Weerman, F. (2011), 'Delinquent Peers in Context: A Longitudinal Network Analysis of Selection and Influence Effects', *Criminology*, 49: 253–86.

Weerman, F. M., Bernasco, W., Bruinsma, G. J. N. and Pauwels, L. J. R. (2013), 'When is Spending Time with Peers Related to Delinquency? The Importance of Where, What, and With Whom', *Crime & Delinquency*. doi:10.1177/0011128713478129.

Weiss, B., Caron, A., Ball, S., Tapp, J., Johnson, M. and Weisz, J. R. (2005), 'Iatrogenic Effects of Group Treatment for Antisocial Youth', *Journal of Consulting and Clinical Psychology*, 73: 1036–44.

Welsh, B. C. and Rocque, M. (2014), 'When Crime Prevention Harms: A Review of Systematic Reviews', *Journal of Experimental Criminology*, 10: 245–66.

West, D. J. and Farrington, D. P. (1977), *The Delinquent Way of Life*. Heinemann.

Wilson, D. B., MacKenzie, D. L. and Mitchell, F. N. (2005), 'Effects of Correctional Boot Camps on Offending', *Campbell Collaboration Library of Systematic Reviews*. doi:10.4073/csr.2005.6.

Wrenn, C. G. (1952), 'An Experiment in the Prevention of Delinquency: The Cambridge-Somerville Youth Study by Edwin Powers; Helen Witmer', *Annals of the American Academy of Political and Social Science*, 279: 214–5.

Young, J. T. N. (2011), 'How Do They 'End up Together'? A Social Network Analysis of Self-Control, Homophily, and Adolescent Relationships', *Journal of Quantitative Criminology*, 27: 251–73.