

The ANNALS of the American Academy of Political and Social Science

<http://ann.sagepub.com/>

Cures That Harm: Unanticipated Outcomes of Crime Prevention Programs

Joan McCord

The ANNALS of the American Academy of Political and Social Science 2003 587: 16

DOI: 10.1177/0002716202250781

The online version of this article can be found at:

<http://ann.sagepub.com/content/587/1/16>

Published by:



<http://www.sagepublications.com>

On behalf of:



<http://www.aaps.org>
American Academy of Political and Social Science

Additional services and information for *The ANNALS of the American Academy of Political and Social Science* can be found at:

Email Alerts: <http://ann.sagepub.com/cgi/alerts>

Subscriptions: <http://ann.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ann.sagepub.com/content/587/1/16.refs.html>

>> [Version of Record](#) - May 1, 2003

[What is This?](#)

Cures That Harm: Unanticipated Outcomes of Crime Prevention Programs

By
JOAN MCCORD

This review highlights the importance of recognizing the possibility for doing harm when intentions are good. It describes several examples showing that well-planned and adequately executed programs provide no guarantee for safety or efficacy. The author concludes with recommendations for scientifically credible evaluations to promote progress in the field of crime prevention.

Keywords: interventions; treatment; crime prevention; random assignment; evaluations; harm

The *New York Times* published an article on Thursday, 4 April 2002 announcing that “a trade group representing British pharmaceutical companies publicly reprimanded Pfizer for promoting several medicines for unapproved uses and marketing another drug before it received government approval” (p. C5). The reprimand was justified because the drugs had not been appropriately tested for safety. Pfizer risked causing harm. No such reprimand could possibly occur in the fields of social intervention.

Researchers, practitioners, and policy makers have begun to understand that evidence is required to identify effective programs to reduce crime. Yet they typically couple the desire for evidence with an inappropriately nar-

Joan McCord, a professor of criminal justice at Temple University, earned a Ph.D. in sociology from Stanford University. Her research has focused on social environments that are conducive to various types of crimes and on the impacts of a variety of kinds of social interventions designed to reduce crime. She has coauthored or edited eleven books and authored more than one hundred articles in the field of criminology. For her contributions to research, she has received the Durkheim Prize from the International Society of Criminology and the Sutherland Award from the American Society of Criminology. Recently, the Society for Research on Adolescence awarded its Social Policy Best Journal Article Award for one of her coauthored articles.

NOTE: The author thanks David Weisburd and Anthony Petrosino for their helpful comments on an earlier draft of this article.

DOI: 10.1177/0002716202250781

row focus. They ask, Does the program work or not? This question is too narrow because it fails to recognize that some treatments cause harm. Intervention programs may, for example, increase crime or the use of drugs. They may decrease the punitive impact of sanctions available to the criminal justice system. They may, perhaps, result in reductions in the ability to cope with life—or even in premature death. 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.

No public reservoir of data permits evaluating whether a given type of program meets even minimum requirements to provide benefits and avoid harm either to recipients of the social programs or to the communities from which they come. Yet social harm is costly to the public, perhaps even more costly than physical harm.

Reluctance to recognize that good intentions can result in harm can be found in biased investigating and reporting. Many investigators fail to ask whether an intervention has had adverse effects, and many research summaries lack systematic reporting of such effects (Sherman et al. 1997).

What has been called a publication bias appears when analyses show that a higher proportion of studies that reinforce popular opinions than those that do not get into peer-reviewed journals (Dickersin and Min 1994; Easterbrook et al. 1991; Scherer, Dickersin, and Langenberg 1994). In summarizing the results of studies evaluating publication bias, Colin Begg (1994) reported that “most studies of the issue have consistently demonstrated that positive (statistically significant) studies are more likely to be published” (p. 401).

One reason for what appears to be a code of silence about adverse effects is fear that all social programs will be tainted by the ones that are harmful. That fear, perhaps justified in some quarters, would be like blocking publication of potentially damaging effects of Celebrex, thalidamide, or estrogen because the publication could slow experimental work in disease prevention. Social programs deserve to be treated as serious attempts at intervention, with possibly toxic effects, so that a science of intervention can prosper.

What follows is a discussion of some social programs that have been carefully evaluated using experimental designs with random assignment to a treatment and a comparison group. They have been found to have harmful effects, and for this reason, they are important experiments. Knowledge that well-designed, carefully implemented social programs can produce unwanted results should set a solid foundation for insisting that all social programs should be coupled with evaluations that have scientific credibility.

The Cambridge-Somerville Youth Study

The Cambridge-Somerville Youth Study was a carefully designed, adequately funded, and well-executed intervention program. Furthermore, a scientifically credible research design played a central role in its construction.

Richard Clark Cabot funded, designed, and, until his death, directed the Cambridge-Somerville Youth Study. As a professor of clinical medicine and social ethics at Harvard, Cabot had made a mark in medicine by showing how to differentiate typhoid fever from malaria. His etiological study of heart disease was widely recognized as an important contribution to the field. He had introduced social services to Massachusetts General Hospital and had been president of the National Conference on Social Work. Not surprisingly, in turning to the problem of crime, Cabot insisted on using a scientific approach, one that aimed to alleviate the probable causes of crime but also one that would permit adequate tests of the results of intervention.

Social programs deserve to be treated as serious attempts at intervention, with possibly toxic effects, so that a science of intervention can prosper.

Cabot's beliefs about the causes of crime derived in part from the work of William Healy and Augusta Bronner, prominent researchers who codirected the Judge Baker Foundation (later known as the Judge Baker Guidance Centre) in Boston. Healy and Bronner reviewed four thousand delinquent cases, half from Chicago and half from Boston. Having discovered that less than 10 percent of the delinquents in their study had come from good homes, Healy and Bronner (1926) concluded that "where to place a large measure of responsibility, where to direct a strong attack in treatment and for prevention of delinquency stands out with striking clearness" (p. 129).

Cabot hypothesized that even rebellious youth from ghastly families "may conceivably be steered away from a delinquent career and toward useful citizenship if a devoted individual outside his own family gives him consistent emotional support, friendship, and timely guidance" (Allport 1951, vi). The Cambridge-Somerville Youth Study would test this hypothesis.

The study began with a matched case design. Staff hired by the youth study solicited names of boys younger than ten who were living in the congested urban environments of Cambridge and Somerville, Massachusetts. To avoid stigmatizing the program, scout leaders as well as the police contributed to the pool of names.

Laboriously, the staff gathered information about the boys, their families, and the neighborhoods of their homes. Each boy was matched to another of similar age, social background, somatotype, and temperament. A toss of a coin determined

which member of each matched pair would be placed into the treatment group and which into the control group.

When a match was identified and the coin had been tossed, a counselor visited the home of the treatment boy. These caseworkers visited the homes as frequently as weekly, when that seemed necessary, but the average frequency was twice a month. Treatment lasted an average of five and one-half years.

The logic of the study required being convinced that the treatment and control groups would have turned out similarly but for the introduction of treatment. Therefore, the groups were compared after a reduction of caseloads due to wartime gas restrictions had taken place in 1942. After the reduction, 253 matched pairs of boys remained in the program. No biases were discovered in the comparisons.

No reliable differences were discovered in comparisons of age, intelligence, whether referral to the youth study had been as "difficult" or "average," or the delinquency prediction scores assigned by the selection committee on the basis of the boys' family histories and home environments. No reliable differences appeared in comparisons regarding the boys' physical health as rated by the doctor after a medical examination, in mental health, in social adjustment, in acceptance of authority, or in social aggressiveness as reflected by teachers' descriptions of the boys. Nor were reliable differences found in ratings of adequacy of the home, disruption of the home, delinquency in the home, adequacy of discipline, standard of living, occupational status of the father, social status level of the elementary school attended by the boy (a measure based on the occupational levels of fathers whose children attended the school), or quality of the neighborhood in which the boys resided. Thus, the randomization within matched pairs had succeeded in producing two groups of boys who were substantially similar prior to the beginning of the treatment program.

During the period of treatment, counselors (most of whom had professional degrees in social work) provided friendly guidance to the boys, counseled parents, assisted the families in a variety of ways, and referred the boys to specialists when that seemed advisable. Boys in the treatment group were tutored, taken to a variety of sports events, and encouraged to participate in the woodwork shop provided by the youth study. Counselors encouraged the boys to join community youth groups and helped them get jobs. Many were sent to summer camps to take them away from the heat of the city.

Counselors were not permitted to accompany the boys to court. Nor were they permitted to include boys from the control group for any of their activities. Of course, boys in the control group received whatever services were provided by other organizations.

When the program terminated in 1945, more than half the treatment boys had been tutored in academic subjects, more than one hundred received medical or psychiatric attention, almost half had been sent to summer camps, and most of the boys had participated with their counselors in such activities as swimming, visits to local athletic competitions, and woodwork in the project's shop. The boys and their parents called on the social workers for help with such problems as illness and

unemployment. They talked with their counselors about their hopes and ambitions as well as about their fears and defeats.

Although a discouraging number of boys in the treatment group were known to have broken the law, at the close of treatment, many boys identified as maladjusted when they entered the program had made fairly good adjustments. Had improvement from prediction been accepted as the measure of success, the program might have been judged effective.

To determine whether the improved adjustment should be attributed to treatment, interviewers tracked down 148 boys who had been in the control group. The interviewers gathered information from the boys, their families, and their school principals. Dr. Helen Witmer was brought into the program to help in its evaluation. She classified each boy among the 148 pairs in terms of adjustment. Disconcertingly, the results indicated that almost equal numbers of the control and the treatment group did better than had been anticipated at the beginning of the project. (See Powers and Witmer 1951 for a more complete description of the program and its early evaluation.)

Additional disappointment came in 1948 from the Massachusetts Department of Probation. Court records showed that a slightly larger number of boys in the treatment group had been in court, 96 versus 92, and they had been charged with a slightly larger number of offenses, 264 versus 218.

Gordon Allport, president of the Board of Directors for the Ella Lyman Cabot Foundation, called for patience. He believed that the program might have prepared the boys to benefit from experience. If so, treatment effects might appear as the youth matured.

Between 1975 and 1981, when the boys were reaching middle age, my research assistants and I retraced the 253 matched pairs who had remained in the program after the cut in 1942. We located 98 percent of them. Questionnaires sent to men from the treatment group asked how, if at all, the program had helped them. Two-thirds of the respondents claimed that the program was helpful, with most of these men amplifying their judgments by specifying ways in which the project or the counselors had improved their lives. These testimonials included claims that the program had helped the men become law-abiding citizens, that it had helped to provide a better understanding of people, and that it had provided evidence that there were "people around who care." Many mentioned that the program had kept them "off the streets," that they were helped by having someone with whom they could talk, and that the counselors had affected their values. Some noted that the program had put them on the right track. Others mentioned the friendships encouraged or the talents acquired. With these subjective endorsements in hand, we sought objective evidence of the program's effects.

We tracked court records both in Massachusetts and in the states to which the men had migrated. We tracked mental hospital records and records from facilities for treatment of alcoholism. We obtained death records to confirm deaths when this was reported, and we searched death records for men who had not been found.

Comparisons between the treatment and control groups showed that for the majority of pairs ($n = 150$), treatment had no measured effect on the objective out-

comes. Nevertheless, for the 103 pairs who had different outcomes, those who had been in the treatment program were more likely to have been convicted for crimes indexed by the Federal Bureau of Investigation as serious street crimes. Those who had been in the treatment program had died an average of five years younger. And those who had been in the treatment program were more likely to have received a medical diagnosis as alcoholic, schizophrenic, or manic depressive (McCord 1978, 1981, 1992).

In 1945, counselors had identified thirty-eight boys as having received the most benefit from the program. Among this select group, twenty-two appeared neither better nor worse than their matches in the control group. Four of the men turned out better than their matches, but twelve turned out worse. Thus, even among those whom the staff believed it had helped most, the objective evidence failed to show that the program had been beneficial.

One might argue that these results had nothing to do with the treatment program. Two comparisons suggest that this argument is wrong.

The first is that adverse treatment effects increased with increased intensity and duration of treatment. That is, the treatment program appeared to reflect a dose response. Boys whose counselors more frequently visited them and those in the treatment program the longest were most likely to fare badly as compared with their matched mates in the control group.

The second is that adverse effects occurred only among boys whose families had cooperated with the program. Families were divided into those who presented problems of cooperation and those who did not. Counselors had dictated reports about each of their interactions with the boys or the families, so most of the case records included several hundred pages. Cases were considered to have shown problems of cooperation if the counselor reported such difficulties or if the case record was exceptionally short (fewer than twenty-five pages), indicating little interaction.

Among the pairs in which the treatment family was uncooperative, the control and treatment boys were equally likely to turn out badly. Among the pairs in which the treatment family was cooperative, however, there were twenty-seven pairs in which the treatment boys turned out better but fifty-two pairs in which the treatment boys turned out worse. These comparisons strongly suggest that the treatment itself had been harmful.

To evaluate effects of the various treatment approaches, I computed an adverse odds ratio by dividing the number of pairs in which the treatment boy did worse than his match by the number of pairs in which the treatment boy did better than his match for each of the major emphases of the treatment program. Adverse odds ratios less than 1 indicate benefits of the treatment program. Conversely, ratios greater than 1 indicate harmful effects of the treatment program.

The odds ratio for bad outcomes for an emphasis on encouraging the boy to participate in community youth groups such as Boy Scouts and YMCA was 1.75 (35:20). That for an emphasis on providing academic help was 1.91 (42:22). The odds ratio for an emphasis on personal problems was 3.5 (28:8). And that for an

emphasis on family problems was 3.75 (30:8). No emphasis seemed to have produced benefits from treatment.

Treatment in the Cambridge-Somerville Youth Study had specifically included summer camp. The camps selected for placement were not designed for troublesome kids. They catered to a general population, one for which summer camping offered an alternative to city heat and boredom as well as the pleasures of outdoor activities.

*Without appropriate equivalent comparisons
in which both efficacy and safety are evaluated,
we cannot know which treatments ought
to be considered beneficial.*

In part because I had developed a theory that would predict increased deviance through close association with peers one wanted to impress, I focused on effects of summer camp (Dishion, McCord, and Poulin 1999). The construct theory of motivation suggests that people construct their motives through the way they perceive choices and that these perceptions are influenced by perceived actions of their associates (McCord 1997, 1999, 2000). At summer camp, misbehaving boys would have unsupervised time during which they would be likely to brag about deviance. A bragging effect would be particularly noticeable among those sent to camp more than once. After the first summer, these boys would have known what camp was like and be in a position to estimate the effects of their reported daring (whether or not these reports were factual).

Among the 253 matched pairs assessed for follow-up, 125 of the treatment boys had been sent to summer camp, and 128 were not. The odds ratio for bad outcomes among those not sent to summer camp was 1.12 (28:25), that for the 59 boys sent to summer camp once was 1.33 (16:12), and that for the 66 boys sent to summer camp at least twice was 10.0 (20:2). In short, none of the treatment approaches showed measurable benefits, and some, particularly repeated placement in summer camps, resulted in harm.

I will summarize with the following list:

1. The Cambridge-Somerville Youth Study was carefully planned.
2. It was based on knowledge that poor families in disorganized neighborhoods were at high risk for crime.
3. Counselors had been trained to carry out their roles, and weekly conferences ensured that they were doing so.

4. Counselors integrated services provided by other available agencies with their own.
5. The program included youth with good as well as bad prognoses so that participation was not stigmatizing.
6. The youth study aimed to change many features of the environment, providing the boys with prosocial guidance, social skills, and healthful activities.
7. The program gave medical assistance and tutoring as well as guidance to both parents and youth.
8. Clients, for the most part, were satisfied with the program.
9. The program lasted five and one-half years, covering the period when the boys were between the ages of 10.5 and 16.
10. The program could be scientifically evaluated because its founder insisted that evaluation was central to the advance of social intervention practices.

Had there been no control group, evaluators might have concluded that the program was beneficial because so many of the treatment boys were better adjusted than anticipated. Or because two-thirds reported beneficial effects for themselves, evaluators might have judged that the program was effective. But these judgments would have been contrary to objective evidence that the program resulted in adverse outcomes for many of the participants.

Let me emphasize again the fact that the Cambridge-Somerville Youth Study was effective. The intervention had lasting effects. These effects were not beneficial.¹ The important legacy of the program, however, is its contribution to the science of prevention. Because the design supports scientifically credible conclusions, it showed that social interventions can have long-term effects. The results also serve to remind anyone willing to heed the warning that we do not yet know how to ensure benefits for youth in need of assistance.

Other Counterproductive Programs

The Cambridge-Somerville Youth Study is not alone in showing that sensible ideas and adequate implementation may produce interventions that fail to achieve their beneficial goals. The following sections describe some others.

Court volunteers

Many courts in the United States encourage volunteer counselors to work with delinquents. Few of these receive adequate evaluation. An exception occurred when Martin Gold, who was director of the Program on Children, Youth, and Family Life at the University of Michigan Institute for Social Research, arranged to evaluate Volunteers in Probation. The program had already won community respect.

Police, caseworkers, or judges could assign probationers to the Volunteers in Probation program. Participation required consent from both the juvenile and his or her guardian. The consent form requested participation in a study involving Volunteers in Probation. Random assignment took place after this consent was obtained, with two out of three being assigned to the program and one of three to a

control group. Those in the control group received the ordinary services of the court, whereas those in the participation group were assigned to group counseling, individual counseling, and tutoring services provided by the volunteers. Evaluations occurred after six months and again after twelve months.

Both self-reports and official records showed that participation in the program inhibited a decline in criminality. Those assigned to the control group and those who had been assigned to the volunteer program but had not participated in it decreased their rates of crime. Those who participated in the volunteer program, however, increased the number of crimes they reported committing. Their court records, too, showed increases in crime as measured by the number of their police contacts (Berger et al. 1975).

Berger et al. (1975) summarized, "While we found some ways that the volunteer service was delivered that seem superior to other ways, none of these proved superior to providing no volunteer service at all" (p. VIII-2). Surprised and disappointed by the results of their study, Berger et al. cautioned,

To those who may feel that other such programs, perhaps their own, are so much superior or so different from this program that our findings and recommendations are irrelevant to them, we urge caution. The staff responsible for this program has reasons good enough for them to feel that their program was effective when this study began, and without this study might still have no reason to feel otherwise. If there is anything that such a study as this one demonstrates, it is the danger of relying exclusively on faith in good works in the absence of systematic data. (Pp. VIII-1–VIII-2)

Group interaction training

Several studies have reported deficiencies in the social skills of delinquents. Hoping to reduce delinquency, many schools developed programs designed to increase the social skills of potential delinquents by giving them practice in discussing issues with well-adjusted peers. Typically, adult leaders guide the discussions. The programs have been called Positive Peer Culture, Peer Culture Development, Peer Group Counseling, and Guided Group Interaction. Several of the programs claim to be highly successful. Few have been evaluated using scientifically credible designs.

In 1982-1983, Gary Gottfredson (1987) arranged to have students in public schools of Chicago randomly selected for inclusion in either the treatment or the control group of a Guided Group Interaction program. Positive leaders, negative leaders, troublesome children, and average children were included in the pool. Fifty-one percent of both the treatment group and the control group were male, Caucasians were approximately equally distributed between the groups, and the groups were equivalent in terms of the prestige of parental occupations, prior police contacts, and age. School tardiness, attachment to parents, self-reported delinquency, and waywardness were used as measures of outcome.

Overall, the results for elementary school children showed no effects. For the high school students, however, the Guided Group Interaction program tended to

increase misbehavior and delinquency. Gottfredson (1987) summarized the posttreatment comparisons: “the present results lend no support to any claim of benefit of treatment. . . . For the high school students, the effects appear predominantly harmful” (p. 708).

A somewhat different approach toward training young adolescents to have increased social skills has backfired in a program administered by the Oregon Social Learning Center. There, aggressive youngsters were randomly assigned to one of four groups: a teen training group that encouraged self-regulation and socialized behavior; a parental training group that encouraged parents to track their youngsters' behavior and to praise them for positive deeds, both, or one in which tapes and booklets substituted for group interaction. Whereas the parental training group (without peer training) seemed to show benefits, both groups assigned to peer training turned out worse than the no-interaction controls (Dishion and Andrews 1995).

Activities programs

Because of the poverty in which so much delinquency is embedded, many observers have concluded that delinquency might be reduced if alternative recreation were available. The Social Options for Teenagers Like You (S.O.F.T.L.Y.) program in Australia was designed as an activities program to provide healthful recreation to delinquent adolescents. In addition, the program was designed

to develop socially relevant skills, develop an awareness of options, teach skills to create further options, teach decision-making, planning and organizational skills (being at the same time aware of the effects of the choice on self and others), and reduce recidivism. (Dufty and Richards 1978, ii)

The program consisted in group activities guided by peer group leaders trained by a supervisor to attend to the participating teenagers' interests. Weekly meetings provided support to the leaders.

Normally, groups met twice a week. Attempts were made to include parents in the meetings, a process facilitated by rotating meeting places among participants' homes. During the first weekly meeting, the group planned the activity to be carried out during the second meeting. Peer groups lasted between ten and twelve weeks.

The experimental group included ten peer groups with four to seven participants in each. Although forty-six teenagers were originally selected, only thirty-nine participants took part in both baseline and follow-up evaluation. A control group of teenagers was matched on sex, age, offending history during six months prior to the initial interview, guardianship, race, nationality of parental figures, work involvement of parental figures, and intellectual capacity.

Assessments were carried out for the experimental group just before the groups were formed and again six months later. For the comparison group, assessments were carried out when a match was identified and then six months after this identification.

The evaluation included measures of school and work involvement as well as delinquent activity. Reliable differences were not found for the former.

Court records identified a greater number of offenders among the treatment group during the first three months following completion of the intervention. Both groups decreased their rates of offending, but only the control group showed a significant decrease in the number of offenses committed. Dufty and Richards (1978) concluded that “this means that S.O.F.T.L.Y. as it currently operates has a detrimental effect on the ‘delinquently inclined’ by increasing recidivism once the intervention ceases” (p. 42). As a consequence of the evaluation, the S.O.F.T.L.Y. program was disbanded.

Scared Straight

Inmates designed a program, popularly known as Scared Straight, on an assumption that delinquency could be prevented by giving wild youngsters a taste of what it would be like to be imprisoned. The project started in Rahway Prison in New Jersey, where its endorsement by judges helped to make a convincing film that popularized the program.

Without scientifically respectable evaluations, Scared Straight projects were adopted in thirty-eight states. Congress held hearings about the program because researchers were skeptical. Miller and Hoelter (1979) found the town from which thirteen of seventeen youngsters in the film had come. They learned that some of the teenagers in the film claimed to have committed crimes to prove they were not scared.

Finally, careful research was carried out, with random assignment to San Quentin’s Squires Program or to a control group. Twelve months later, 81 percent of the experimental group and 67 percent of the control group had been arrested (Lewis 1983). Other scientifically credible evaluations, too, have shown that attempts to scare teenagers into better behavior is not a successful enterprise (Petrosino, Turpin-Petrosino, and Buehler 2002).

Summary and Conclusions

I have described five types of programs that seemed promising but had harmful effects. Evidence about two of these—those involving court volunteers and those providing healthful group activities—appear in what has been called the fugitive literature. That is, despite solid research designs, the results have not been published. Evidence about adverse effects from social programs is hard to find in part because of a strong bias against reporting adverse effects of social programs. Authors of studies that fail to produce evidence of beneficial outcomes sometimes do not bother to submit their reports for publication. But also, those who do submit for publication tend to receive delays or rejections attributable to the unpalatable message they convey.

Many people seem to be willing to believe favorable results of inadequate evaluation designs. Some accept testimonials from clients who express their appreciation of a program. Against the claim that these provide valid evidence of effect, it should be noted that each of the programs described above would have been counted as successful by this criterion. Yet the clients would have been better off had they not participated in the programs.

Some argue that without comparison groups, measures taken before and after intervention can be used for valid evaluations. But changes over time occur for a variety of reasons, many of which are not documented. If changes are favorable and are more likely to occur in the absence of a program, the program should not be considered beneficial. The Cambridge-Somerville Youth Study might have been considered beneficial had improvements over prediction been accepted as the measure of outcome.

*[E]ach of the harmful programs described
above had been considered beneficial
prior to its evaluation.*

Often, one finds resistance to scientifically credible evaluations on the grounds that one ought not deprive some clients of the benefits given to others. Yet each of the harmful programs described above had been considered beneficial prior to its evaluation. Without appropriate equivalent comparisons in which both efficacy and safety are evaluated, we cannot know which treatments ought to be considered beneficial.

I have read several final reports of intervention programs that describe outcomes that are significantly worse than those in the comparison but include in the executive summary only results favorable to the program, often adding that the size of the sample precludes obtaining significant differences favoring treatment.

When results of the Cambridge-Somerville study were first published (and they were published only on the condition that a critical article would be coupled with its publication), I received threatening phone calls and notes. When I gave talks about these results, in many audiences, people shouted ugly names at me.

Researchers typically fail to consider whether social programs have had adverse effects, looking only for favorable results of treatment. Government agencies sponsor intervention programs with no provision for adequate evaluation. These are problems for the advancement of social well-being.

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.

Clearly, social programs can have enduring effects. Although some popular interventions have harmful effects, of course, other intervention programs benefit their clients. Without scientifically credible evaluations, we cannot learn which programs are beneficial and which are harmful.

It is not enough to evaluate a program once. As noted by Weisburd and Taxman (2000), "The strength of experimental designs in specifying treatment impacts for specific populations does not in itself overcome the weaknesses associated with single site research studies" (p. 316).

Even when replications suggest that a particular type of program is effective, we should not assume that the program will work under new conditions. Historical changes, for example, in the definitions of crime or availability of drugs or of employment might alter the outcome of particular interventions. Demographic differences such as age, sex, or ethnicity might affect whether an intervention is effective. Different places, with different practices (e.g., regarding day care, medical coverage, or education), might reflect the influence of unmeasured variables on the relationship between interventions and outcomes. As Peter Grabosky (1996) noted in his review of unintended consequences of crime prevention strategies, "What works in Wollongong might fail on Palm Island" (p. 39).

Canada bears many similarities to the United States. Nevertheless, the Center for Children & Families in the Justice System wisely recognized that programs effective in some environments might not be effective in different environments. It brought the promising multisystemic therapy (Henggeler, Melton, and Smith 1992; Henggeler et al. 1993) from the United States to Ontario, Canada. The Canadian program involved a multisite design with random assignment to treatment and comparison groups. The comparison groups received the usual treatments in each of the four sites involved in the study. Program fidelity was monitored. Survival curves for convictions of 407 youth at six months, 363 at one year, 239 at two years, and 115 at three years give no indication of benefit from the program. Alison Cunningham (2002), director of research and planning for the project wrote, "Because the control group has the same outcomes as the MST [multisystemic therapy] recipients, it is unsafe to conclude that the two American studies are sufficient evidence to justify the wide-spread adoption of MST in Canada" (p. 11).

Social programs can cause crime as well as reduce it. They also can increase illness and reduce the ability of clients to cope with life's challenges. Effects of criminal justice interventions on education, mental health, and job performance deserve attention. A practice that decreases crime but increases alcoholism or mental illness might not be considered a net gain either by the clients or by the community that supported the program.

Potentially harmful effects of drugs have been recognized, and drug companies are required to keep track of reports of problems with the medications they advertise and sell. These can be subject to periodic review. Similar standards might be

embraced for social programs. Recognizing that programs can have harmful effects may be critical to acceptance of experimental designs for evaluating social interventions.

Clearly, if social practice is to be improved, continuing evaluation should be an integral part of social interventions. Whenever possible, these evaluations should employ random assignment of similar people to either treatment or comparison groups. Always, the outcome should be measured in ways that do not rely on the typically favorable biases of clients, program providers, and sponsors. The evaluations should, of course, include a check for evidence of adverse effects as well as benefits.

We do not know the dimensions of variation that affect social programs. Careful collection of data to document the process of treatments and their effects should become as essential in the field of criminology as it is in the field of highway or airline safety.

It would be extremely useful to have not only a data repository that provides systematic reviews of high-quality research, as will the Campbell Collaboration (Farrington and Petrosino 2001), but also one that collects information about particular programs in specific venues. If evaluation becomes an expected part of program administration, and all well-designed programs and their evaluations contribute toward such a data repository, knowledge about the safety and effectiveness of social programs would begin to accumulate, and informed decisions could be made.

Note

1. Discussion of possible causes for these effects can be found in McCord (1978, 1981) and Dishion, McCord, and Poulin (1999).

References

- Allport, Gordon. 1951. Foreword. In *An experiment in the prevention of delinquency: The Cambridge-Somerville Youth Study*, edited by E. Powers and H. Witmer. New York: Columbia University Press.
- Begg, Colin B. 1994. Publication bias. In *The handbook of research synthesis*, edited by Harris Cooper and Larry V. Hedges. New York: Russell Sage.
- Berger, R. J., J. E. Crowley, M. Gold, J. Gray, and M. S. Arnold. 1975. *Experiment in a juvenile court: A study of a program of volunteers working with juvenile probationers*. Ann Arbor: Institute for Social Research, University of Michigan.
- Cunningham, Alison. 2002. *One step forward: Lessons learned from a randomized study of multisystemic therapy in Canada*. London, Canada: Centre for Children and Families in the Justice System.
- Dickersin, Kay, and Y.-I. Min. 1994. Publication bias: The problem that won't go away. *Annals of the New York Academy of Sciences* 703:135-46.
- Dishion, Thomas J., and David W. Andrews. 1995. Preventing escalation in problem behaviors with high-risk young adolescents: Immediate and 1-year outcomes. *Journal of Consulting and Clinical Psychology* 63 (4): 538-48.
- Dishion, Thomas J., Joan McCord, and François Poulin. 1999. When interventions harm: Peer groups and problem behavior. *American Psychologist* 54 (9): 1-10.

- Dufty, B. J., and W. Richards. 1978. Evaluation of S.O.F.T.L.Y. Unpublished manuscript, Australian Institute of Criminology, Canberra.
- Easterbrook, P. J., J. A. Berlin, R. Gopalan, and D. R. Matthews. 1991. Publication bias in clinical research. *Lancet* 337:867-72.
- Farrington, David P., and Anthony Petrosino. 2001. The Campbell Collaboration Crime and Justice Group. *Annals of the American Academy of Political and Social Science* 578:35-49.
- Gottfredson, Gary D. 1987. Peer group interventions to reduce the risk of delinquent behavior: A selective review and a new evaluation. *Criminology* 25 (3): 671-714.
- Grabosky, Peter N. 1996. Unintended consequences of crime prevention. *Crime Prevention Studies* 5:25-56.
- Healy, William, and Augusta F. Bronner. 1926. *Delinquents and criminals: Their making and unmaking*. New York: Macmillan.
- Henggeler, Scott W., Gary B. Melton, and Linda A. Smith. 1992. Family preservation using multisystemic therapy: An effective alternative to incarcerating serious juvenile offenders. *Journal of Consulting & Clinical Psychology* 60:953-61.
- Henggeler, Scott W., Gary B. Melton, Linda A. Smith, Sonja K. Schoenwald, and J. H. Hanley. 1993. Family preservation using multisystemic therapy: Longterm follow-up to a clinical trial with serious juvenile offenders. *Journal of Child & Family Studies* 2:283-93.
- Lewis, Roy V. 1983. Scared Straight—California style. *Criminal Justice and Behavior* 10 (2): 284-89.
- McCord, Joan. 1978. A thirty-year follow-up of treatment effects. *American Psychologist* 33 (3): 284-89.
- . 1981. Consideration of some effects of a counseling program. In *New directions in the rehabilitation of criminal offenders*, edited by Susan E. Martin, Lee B. Sechrest, and Robin Redner. Washington, DC: National Academy of Sciences.
- . 1992. The Cambridge-Somerville Study: A pioneering longitudinal-experimental study of delinquency prevention. In *Preventing antisocial behavior: Interventions from birth through adolescence*, edited by Joan McCord and Richard E. Tremblay. New York: Guilford.
- . 1997. He did it because he wanted to . . . In *Motivation & delinquency*. Vol. 44 of *Nebraska symposium on motivation*, edited by D. Wayne Osgood. Lincoln: University of Nebraska Press.
- . 1999. Understanding childhood and subsequent crime. *Aggressive Behavior* 25:241-53.
- . 2000. A theory of motivation and the life course. In *Social dynamics of crime and control: New theories for a world in transition*, edited by Susanne Karstedt and Kai-D Bussmann. Portland, OR: Hart.
- Miller, Jerome G., and Herbert H. Hoelzer. 1979. *Prepared testimony: Oversight on Scared Straight*. Washington, DC: Government Printing Office.
- The New York Times*. 2002. 4 April, p. C5.
- Petrosino, Anthony, Carolyn Turpin-Petrosino, and John Buehler. 2002. *The effects of Scared Straight and other juvenile awareness programs on delinquency*. Issue 3 of *Cochrane library*. Oxford, UK: Update Software.
- Powers, Edwin, and Helen Witmer. 1951. *An experiment in the prevention of delinquency: The Cambridge-Somerville Youth Study*. New York: Columbia University Press.
- Scherer, Roberta W., Kay Dickersin, and Patricia Langenberg. 1994. Full publication of results initially presented in abstracts: A meta-analysis. *Journal of the American Medical Association* 272:158-62.
- Sherman, Lawrence W., Denise C. Gottfredson, Doris L. MacKenzie, John E. Eck, Peter Reuter, and Shawn D. Bushway. 1997. *Preventing crime: What works, what doesn't, what's promising*. Washington DC: U.S. Department of Justice, National Institute of Justice.
- Weisburd, David, and Faye S. Taxman. 2000. Developing a multicenter randomized trial in criminology: The case of HIDTA. *Journal of Quantitative Criminology* 16:315-40.