J Exp Criminol (2007) 3:353–375
DOI 10.1007/s11292-007-9040-2

Does incarceration-based drug treatment reduce recidivism? A meta-analytic synthesis of the research

Ojmarrh Mitchell · David B. Wilson · Doris L. MacKenzie

Published online: 28 September 2007

Abstract This research synthesized results from 66 published and unpublished evaluations of incarceration-based drug treatment programs using meta-analysis. Incarceration-based drug treatment programs fell into five types: therapeutic communities (TCs), residential substance abuse treatment (RSAT), group counseling, boot camps specifically for drug offenders, and narcotic maintenance programs. We examined the effectiveness of each of these types of interventions in reducing post-release offending and drug use, and we also examined whether differences in research findings can be explained by variations in methodology, sample, or program features. Our results consistently found support for the effectiveness of TC programs on both outcome measures, and this finding was robust to variations in method, sample, and program features. We also found support for the effectiveness of RSAT and group counseling programs in reducing re-offending, but these programs' effects on drug use were ambiguous. A limited number of evaluations assessed narcotic maintenance or boot camp programs; however, the existing evaluations found mixed support for maintenance programs and no support for boot camps.

Keywords Corrections · Drug offenders · Drug treatment · Meta-analysis · Offenders

While there is clear evidence that a substantial proportion of incarcerated offenders are drug dependent, the effectiveness of common incarceration-based drug abuse

Studies included in this meta-analysis are indicated with an asterisk

O. Mitchell (✉)
Division of Criminal Justice, University of Cincinnati, 600 Dyer Hall, PO Box 210389, Cincinnati, OH 45221-0389, USA
e-mail: ojmarrh.mitchell@uc.edu

D. B. Wilson
Administration of Justice, George Mason University, Manassas, VA, USA
e-mail: dwilsonb@gmu.edu

D. L. MacKenzie
Department of Criminology and Criminal Justice, University of Maryland, College Park, MD, USA
e-mail: dmackenzie@crim.umd.edu

programs in reducing recidivism is ambiguous. In the United States of America, research suggests that between 40% and 45% of inmates are drug dependent. For example, a 2002 study of US jail inmates found that nearly half (45%) were substance dependent and another 23% reported substance abuse symptoms (Karberg and James 2005). Similarly, findings from the Arrestee Drug Abuse Monitoring (ADAM) program revealed that the median rate of drug dependence among incarcerated arrestees in 39 US cities was 39% (Zhang 2003). High levels of drug dependence also have been found in samples from England and Canada. For example, a study of incarcerated arrestees in five English cities found that 45% of inmates were drug dependent at one point in their lives and 33% were drug dependent at the time of interview (Bennett 1998). Likewise, a sample of Canadian federal prisoners found that the rate of drug dependence was between 31% and 43% (Pernaen et al. 2002).

In the absence of effective abuse treatment, it is likely that a high proportion of these drug-dependent offenders will re-offend. Research consistently indicates that drug-using probationers/parolees recidivate at higher rates than do non-drug users (see e.g., Bureau of Justice Statistics 1995; Gendreau et al. 1996; Horney et al. 1995; MacKenzie et al. 1999). Thus, the period of incarceration is a potentially critical opportunity to intervene in this cycle of drug use and crime. However, the two most common types of drug abuse interventions in US prisons and jails, self-help 12-step and drug education programs, have not been demonstrated to reduce recidivism, in existing reviews of the literature (Pearson and Lipton 1999; Lipton et al. 1992).

In recognition of the potential of incarceration-based drug treatment, the limitations of existing programs, and the rising need for effective substance abuse interventions, US Congress created the Residential Substance Abuse Treatment (RSAT) Program under the Violent Crime Control and Law Enforcement Act of 1994 [Bureau of Justice Assistance (BJA) 2005]. This program provides financial assistance to local and state correctional facilities to develop and implement incarceration-based drug treatment programs. Between fiscal years 1996 and 2003, more than $400 million have been provided to the states under this program; and as of July 2004, 300 RSAT programs were in operation (BJA 2005).

This tremendous investment in incarceration-based drug treatment programs and their proliferation present a host of important questions that need to be addressed, such as: Are incarceration-based drug treatment programs effective in reducing recidivism? Does the existing body of research indicate that participation in RSAT-funded programs reduces recidivism? Are there particular types of drug treatment that are especially effective or ineffective? Approximately how effective are these programs? What characteristics differentiate effective programs from ineffective programs? This research attempts to answer these questions by synthesizing existing evaluations of incarceration-based drug treatment programs using systematic, quantitative review techniques.

Common incarceration-based drug treatment programs and their effectiveness

Many correctional interventions could be interpreted as "incarceration-based drug treatment." We defined incarceration-based drug treatment programs as interventions

that were specifically targeted at substance abusers, intended to reduce future substance abuse and other criminal behaviors, and were based in a prison, jail, or analogous facility for juvenile offenders. Given this definition, program evaluations fell into five types: (1) therapeutic community (TC) programs, (2) RSAT-funded programs (most of which were based on the TC model), (3) group counseling programs, (4) boot camps for drug offenders, and (5) narcotic maintenance programs.

TCs are the most intensive of all incarceration-based drug treatment programs. While the components of TCs vary, there are several common components. TCs house participants in a distinct treatment unit away from the anti-social influences of the general correctional environment to create an atmosphere focused on rehabilitation and reformation. Another core component of the TC model is that participants are instrumentally involved in running the TC; for example, participants lead treatment sessions, monitor other residents for rule compliance, maintain the treatment unit, and resolve disputes. TCs have a confrontational nature in which both staff and participants openly challenge anti-social behaviors and attitudes; yet, participants also are supportive of each others' struggles to make pro-social reformations. Many TCs share a philosophy that sees drug abuse as symptomatic of a broader personality disorder, and, consequently, many TCs focus on the larger disorder and not just drug abuse, per se (Kooyman 1993). Further, not only are TCs intense in nature but they are also long in duration—typically lasting at least 6 months but usually around 12 months.

RSAT-funded programs share several components with the TC model. Like TC programs, RSAT-funded programs usually separate program participants from the general population; they tend to be intensive and have a long duration (6 to 12 months for prison-based programs, or at least 3 months for jail-based programs). Within these parameters, local corrections agencies control the content of such programs. Our review of the extant literature reveals that most of the RSAT-funded programs are based on the TC model. Thus, in this review, RSAT-funded programs are essentially TCs, but not all TCs are RSAT-funded programs.

Counseling programs use various combinations of drug education, self-help 12-step programming, life skills training, and cognitive and behavioral skills training. The key commonality among these programs is their reliance on group-based counseling approaches, in which substance abuse and other common problems are discussed among peers in an effort to solve these problems. Typically, counseling programs are multi-faceted, including several of the above-mentioned components; however, some counseling programs heavily emphasize one of these components. And still other counseling programs include both group and individual counseling.

Boot camps are modeled after military basic training. Inmates participate in rigorous exercise regimens, learn military drill and ceremony, wear uniforms, and take on challenge courses (timed obstacle courses). Boot camps are highly structured. From the moment residents wake in the morning until lights out they are constantly engaged in scheduled activities. Boots camps also involve considerable confrontation, but, unlike TC programs, confrontations most often occur between correctional staff and inmates—with drill instructors disciplining any deviation from established codes of conduct. Boot camp proponents claim that the harsh nature of boot camps serves as a deterrent to future criminal conduct, and the

content of these programs instills self-discipline, which also leads to reduced recidivism (Wilson and MacKenzie, 2006).

Narcotic maintenance programs [i.e., methadone and levo-alpha-acetyl-methadol (LAAM) maintenance] are pharmacological approaches. These programs attempt to reduce the harms associated with opiate dependency (e.g., disease transmission, criminal activity) by prescribing synthetic opioid medication. Unlike heroin and other strong opiates, these medications do not produce a euphoric high; instead, methadone and LAAM block the euphoric high produced by opiate use and suppress opiate withdrawal symptoms. Incarceration-based narcotic maintenance programs link released inmates to community-based programs for continued care. The success of incarceration-based maintenance programs appears to be heavily dependent on their ability to place released offenders into community-based programs.

Each of the above types of drug interventions has the potential to reduce drug use and other criminal behaviors. However, existing systematic reviews of this body of research have found that only participation in TC programs was consistently associated with reductions in recidivism. In particular, Pearson and Lipton (1999) systematically reviewed experimental and quasi-experimental research assessing the effectiveness of corrections-based drug abuse programs in reducing recidivism conducted in any country and completed between 1968 and 1996, inclusive. Their search revealed 30 studies meeting their eligibility criteria. These authors found that six of the seven TC evaluations indicated that participants exhibited lower recidivism than non-participants; the overall mean weighted r effect size was 0.133 (p=0.025). By contrast, the mean effect size was not statistically significant and near zero for both boot camp and group counseling programs. Although Pearson and Lipton's review found too few studies of methadone maintenance, drug education, cognitive-behavioral, and 12-step programs to draw strong conclusions, it characterized the existing evidence of these programs' effectiveness as being promising.

In many regards, this systematic review is an extension of the work by Pearson and Lipton. Like that earlier work, this synthesis systematically and comprehensively reviews the effects of incarceration-based drug interventions on post-treatment drug use and other types of criminal behaviors, using meta-analytic procedures. The primary difference between their work and the current systematic review is that this research project uses a more current time frame (1980 through 2004). We believe that this difference is salient for two reasons: (1) more recent evaluations of drug treatment interventions may be more generalizable to current correctional practices; and, (2) numerous evaluations of incarceration-based drug treatment programs have been conducted since 1996, particularly evaluations of RSAT programs.

Method

Eligibility criteria

The scope of this review was published or unpublished experimental and quasi-experimental evaluations of incarceration-based drug treatment programs for juveniles and adults that utilized a comparison group. The eligibility criteria for this review were that: (1) the evaluation assessed an intervention that was administered in a correctional

facility; (2) the intervention specifically targeted substance users; (3) the evaluation used an experimental or two-group quasi-experimental research design that included a no-treatment or minimal treatment comparison group; (4) the evaluation reported a post-release outcome measuring offending or drug use; (5) the intervention was conducted between 1980 and 2004, inclusive; and, (6) the evaluation had to report enough information for us to calculate an effect size.

Regarding the first eligibility criterion, our operational definition of "correctional facilities" included only jails and prisons, and analogous facilities for juveniles. Interventions conducted at half-way houses or community-based residential facilities were not included. It is worth noting that this criterion excluded a small number of notable evaluations. Specifically, programs designed to be alternatives to incarceration, such as those reported in Dynia and Sung (2000) and Knight and Hiller (1997), were excluded by this criterion.

The second criterion restricted the focus of this review to evaluations that specifically targeted drug users. Therapeutic interventions conducted in correctional facilities that were generally available to offenders regardless of an offender's drug history were not included. For instance, Shaw and MacKenzie (1992) evaluated the effects of a boot camp program on a sub-sample of drug-using offenders; however, this evaluation was excluded because the boot camp program was not specifically targeted at drug users. By contrast, Zhang (2000) evaluated a boot camp program specifically geared towards drug users—this evaluation was included in this review. This criterion was necessary, because this review is concerned with incarceration-based *drug treatment*; without this criterion, the present review would become a review of incarcerated-based treatment *comprised* of drug users.

The third criterion specified that all included evaluations must have a comparison/control group that received no treatment or minimal treatment. Thus, we excluded quasi-experiments that involved comparisons of two or more interventions that were roughly comparable or whose comparability in terms of effectiveness in reducing recidivism was in dispute (i.e., treatment–treatment comparisons or dose–response evaluations). For example, the comparison group in the study by Swartz et al. (1996) was constructed by dividing program participants into four groups based upon length of program participation. Evaluations utilizing such comparison groups were not included in this systematic review. Furthermore, we did not include evaluations in which the comparison group was comprised predominantly or solely of dropouts from the intervention of interest. For instance, evaluations such as those by Field (1985, 1989) and Berggren and Svard (1990) that used program drop-outs as the comparison group were excluded from this systematic review.

The fourth and fifth criteria are largely self-explanatory. We chose to exclude studies conducted before 1980, because we questioned the generalizability of findings from such studies to current correctional contexts. It is important to emphasize, however, that all evaluations needed to report a *post-release* measure of recidivism. This criterion excluded a few notable evaluations, such as those by Shewan et al. (1996) and Dolan et al. (2003), which reported in-prison outcomes.

The last criterion excluded evaluations that did not report enough information for us to calculate an effect size. This criterion was necessary for practical purposes. Unfortunately, a handful of otherwise eligible evaluations (e.g., those by Schippers et al. 1998; Guerin 2002) were ruled ineligible on the basis of this last criterion.

Search strategy

The goal of the search strategy was to identify all studies, published or unpublished, that met the above eligibility criteria. The search began by our conducting a computerized keyword search of bibliographic databases. In particular, we conducted a search of the following databases: PsychLit, MedLine, NCJRS, Criminal Justice Abstracts, Dissertation Abstracts, Sociological Abstracts, Social Science Citation Index, SocioFile, Conference Papers Index, UnCover, C2 SPECTR, and CINAHL, as well as Google internet searches. The keywords used were: drug treatment, substance abuse treatment, drug counseling, therapeutic community(ies), methadone maintenance, boot camp(s), offenders, residential substance abuse treatment, RSAT, inmates, incarceration, incarcerated, prison, evaluation, outcome evaluation, and recidivism. These keywords were used in various combinations.

We also searched for eligible evaluations by carefully reading existing reviews of unfamiliar research. In particular, we reviewed the reference lists of existing syntheses to identify potentially eligible evaluations. Likewise, many of the eligible studies reviewed the work of similar studies; these studies were also assessed for eligibility. Additionally, we reviewed the *Digest of Research on Drug Use and HIV/AIDS in Prisons* (Flanagan et al. 2004), which abstracts much of the "grey" literature, particularly European grey research.

Further, we searched websites of several prominent research organizations. Specifically, we searched for relevant research reports on the following websites: Correctional Service Canada's research publications page; the Home Office; RAND Drug Policy Research Center; The Urban Institute's crime/justice research page; and, Vera Institute of Justice's publications page.

Finally, we hand-searched the titles/abstracts of articles published between 1999 and 2004 in the following journals: *Journal of Substance Abuse Treatment*, *International Journal of Offender Therapy and Comparative Criminology*, *Journal of Drug Issues*, *The Prison Journal*, *Crime & Delinquency*, and *Journal of Offender Rehabilitation*. We chose to hand-search these journals because they have a strong track record of publishing relevant studies, and many of these journals were not indexed well by the computerized databases we utilized.

All studies that could not be ruled ineligible, based on a preliminary review of the title and abstract, were retrieved. That is, we reviewed each title/abstract looking for clear evidence that the study violated one or more of the eligibility criteria. For example, the first eligibility criterion was that all studies evaluated an intervention administered in a correctional facility; thus, if a title/abstract clearly indicated that the study assessed a program that was not administered in a correctional facility, then the study was not retrieved for further scrutiny. We read closely the retrieved studies to determine final eligibility status.

It is important to note that the unit of analysis in this meta-analysis is the evaluation—a contrast between one treatment group and one comparison group. We distinguish *studies*, documents reporting the results of an evaluation or evaluations, from *evaluations*, specific contrasts between a treatment and comparison group. This distinction is necessary, as several studies reported the results of more than one evaluation; that is, evaluations were nested within studies.

Coding of effect sizes and moderator variables

The coding forms employed in this review were structured hierarchically, so that we could recognize explicitly the nested nature of evaluations within studies. Any number of effect sizes could be coded from each evaluation using these forms. Further, the coding forms captured key features of the nature of the treatment, research sample, research methodology, outcome measures, and direction and magnitude of observed effects.

We coded at least one effect size from each evaluation. We utilized the odds ratio effect size for dichotomous outcomes, as this type of effect size is the most appropriate for such outcomes (Lipsey and Wilson 2001). Indicators of criminal behavior based on non-dichotomous scales were coded using the standardized mean difference effect size. Effect sizes were coded in a manner such that *positive effect sizes indicated that the treatment group had a more favorable outcome than the comparison group did* (i.e., less re-offending or drug use). Odds ratio and standardized mean difference effect sizes were combined using the method developed by Hasselblad and Hedges (1995). Specifically, mean difference effect sizes were transformed onto the odds ratio effect size scale (see also Sánchez-Meca et al. 2003).

It is important to note that many authors in this body of research report their findings by contrasting participants who completed the treatment ("graduates"), participants who failed to complete the treatment ("failures"), and the comparison group. Findings reported in this manner clearly open the door for selection bias, and, thus, such findings are difficult to interpret (i.e., was the observed reduction in recidivism due to the treatment or selection bias). Whenever we encountered results reported in this fashion, we re-calculated recidivism rates for the treatment group by combining graduates with failures. We then compared this overall treatment group to the comparison group. We believed that this practice would reduce the likelihood that selection bias would influence the results of this meta-analysis.

Our analyses of these effect sizes utilized the statistical approach outlined by Lipsey and Wilson (2001) and Cooper and Hedges (1994). In particular, we used the inverse variance method and assumed that the true treatment effects varied as a function of both measured (i.e., coded features of each evaluation) and unmeasured differences between evaluations. This analytic approach required statistical independence of effect sizes. Several types of statistical dependencies were evident in the coded evaluations. First, it was common for evaluators to measure and analyze more than one indicator of criminal behavior (e.g., re-arrest, re-conviction, drug use) or analyze the same indicator of criminal behavior at multiple post-program follow-ups (e.g., 6 months, 12 months). Second, it was also common for authors to report findings from the same sample of research participants in multiple studies.

We employed several procedures to attain statistical independence. As a first-step, all evaluations were cross-checked against one another to ensure that each evaluation was included no more than once in each data set. Next, we created two data sets, one for drug use outcomes and another for re-offending outcomes. Some evaluations reported more than one measure of drug use or re-offending:, instead of simply averaging these multiple measures, we selected particular outcomes of interest to maximize comparability of outcomes. In particular, for re-offending measures, we selected effect sizes

that were: (1) general (i.e., covered all offense types as opposed to being offense specific), (2) based on re-arrests, (3) dichotomous, and (4) followed sample members for 12 months. We decided to select such effect sizes because this type of outcome measure was the most commonly reported outcome measure. For evaluations that did not report such an outcome, we coded the available outcome that most closely resembled this preferred outcome measure. Thus, the re-offending data set primarily contained outcome measures concerning re-arrest rates in the first 12-months after release. Similarly, for the drug use data set, we selected effect sizes that were: (1) general (i.e., concerning any illicit drug use), (2) dichotomous, and, (3) followed sample members for 12 months. Consequently, the drug relapse data set primarily contained outcomes concerning drug relapse in the first 12 months after release.

Results

Description of eligible studies and evaluations

Our search strategy uncovered 233 potentially eligible studies. We were able to obtain copies of 229 of these studies. Of the retrieved studies, 53 unique studies met our eligibility criteria. These 53 unique studies reported the results of 66 independent evaluations, as several studies contained multiple evaluations. Of these evaluations, 58 had been conducted in the United States of America, three had been conducted in Australia, three had been conducted in Canada, one had been conducted in the UK, and one evaluation had been conducted in Taiwan. Approximately half of the evaluations (32) were coded from studies published as journal articles or book chapters, and the other 34 evaluations were coded from unpublished technical reports and government documents. In regards to date of publication, two-thirds of the evaluations were coded from studies made available after 1996—the latest date eligible for inclusion in Pearson and Lipton's (1999) review; thus, the vast majority of the evaluations included in the current research were not included in the earlier review.

The bulk of the evaluations assessed either TC or counseling programs. of the 66 evaluations, 30 examined the effectiveness of TC programs, and another 25 evaluations were of counseling programs. Relatively few of the evaluations were of narcotic maintenance programs or boot camp programs designed for drug offenders: five and two evaluations, respectively. And primary type of treatment modality could not be reliably coded for four evaluations.

Mean odds ratio for re-offending and drug use outcomes

Of the 66 evaluations, 65 reported at least one re-offending outcome measure. One odds ratio was an outlier in the negative direction; this odds ratio was "Winsorized" to the 5th percentile.[1] Approximately 83% of the 65 odds ratios for offending

[1] The outlier came from an evaluation of a narcotics maintenance program (Magura et al. 1993, female sample). The logged odds ratio for this contrast was -4.61, indicating that the comparison group had a lower rate of post-release offending than the treatment group did. No other logged odds ratio had an absolute value of 2 or more. Alternative analyses that excluded this outlier yielded substantively similar results.

indicated that the treatment group re-offended less than the comparison group did. The random-effects mean odds ratio was 1.37, with a 95% confidence interval (CI) of 1.24 to 1.51. This finding indicates that, in general, participation in incarceration-based drug treatment programs was associated with a reduction in post-release offending. A more intuitive sense of this effect size can be gained by translating this effect size into percentages. As a matter of convenience, we assumed a 50% rate of re-offending for the comparison group. Given this assumption, the overall means odds ratio translates into a 42% re-offending rate for the treatment group.[2] The distribution of the odds ratios for offending exhibited considerably more variability than expected by sampling error alone (Q=551.84, df=64, p<0.001). This finding suggested that features of the treatment programs, research methodology, and/or characteristics of the sample might be systematically related to effect size. Subsequent analyses tested this possibility.

Examination of the effect for drug use produced a slightly smaller overall mean effect size that was not statistically significant. The random-effects mean odds ratio for the 20 evaluations on drug use outcomes was 1.28 (95% CI of 0.92-1.78), and there was more variability than expected by sampling error alone (Q=197.97 df=19, p<0.001). Thus, with respect to drug relapse, the evidence suggests great variability in effects across evaluations, and, overall, we cannot conclude that these programs are effective in reducing post-release drug use, although the evidence is positive. Moreover, the finding of great variability in effects suggests that there may be program or method features that are associated with drug relapse effects; for example, perhaps participation in programs with long treatment durations do, in fact, facilitate post-release reductions in drug use.

Moderator analyses

The above analyses indicated that the effect size distributions displayed more variability than expected by chance alone. This finding suggested that there may be important differences in research methodology, sample, and/or interventions that may account for some of this variability in effect size. We tried to capture important differences between evaluations by coding information from each evaluation. These moderator analyses, however, were confronted with three major issues. First, our ability to code many relevant features was limited by the quality of the descriptions provided by the primary authors. Second, the moderator analyses had limited statistical power. Post-hoc power analyses revealed that the power to detect a small effect (i.e., a difference in logged odds ratio of 0.20) in these moderator analyses ranged from approximately 0.10 to 0.40. This limited statistical power means that only contrasts with large effects were likely to be statistically significant. To combat this issue, we interpreted as statistically significant any contrast that had a probability of occurring by chance alone of less than 10% (i.e., p<0.10), and, instead of relying solely on *statistical significance*, we discussed *substantively significant* effects. We defined "substantively significant" relationships based on the

[2] Alternatively, we assumed a 35% re-offending rate for the comparison group, which was the average 12-month re-offending rate. This effect size suggested a 28% rate of re-offending for the treatment group.

magnitude of the difference between categories in terms of the *logged* odds ratio. Specifically, if a moderator variable's categories differed by a logged odds ratio of 0.20 or more, and each category had a least five evaluations, then we considered such differences as substantively significant. Similarly, if one category of a moderator had fewer than five evaluations, and the difference between categories was 0.40 or more, then we considered such differences as substantively significant. Third, all of the moderator analyses were essentially bivariate. Unfortunately, the limited number of available effect sizes, and the sparseness of the data sets utilized, made multivariate data analysis highly problematic, and the results of such analyses were very sensitive to small alternations (e.g., excluding one observation). As a result, these bivariate findings were vulnerable to spuriousness. Given these challenges, the following moderator analyses are offered as suggestive.

The first moderator variable examined was primary type of intervention. Table 1 displays the random-effects mean odds ratio effect size for each type of treatment program for re-offending and drug use outcomes. Analysis of these means indicated that treatment effectiveness on both types of outcome measures varied considerably by primary type of intervention (re-offending: Q=6.36 with df=3, p=0.09. This analysis excludes the "ambiguous" category, drug use: Q=4.29 with df=2, p=0.12). The mean odds ratio for TC programs on re-offending outcomes was 1.38, with a 95% CI of 1.17 to 1.62. As this CI did not include 1.00, this mean odds ratio is statistically significant at the 0.05 level. Similarly, the mean odds ratio for the 12 evaluations of TC programs on drug use outcomes was 1.41 (95% CI 0.95-2.05, $p<$ 0.10). These findings indicated that TC participants generally had statistically lower rates of re-offending and drug use after release from incarceration than did non-participants.

Ten of the TC evaluations were RSAT-funded programs (not shown in Table 1). These ten evaluations yielded a mean odds ratio of 1.39 (95% CI 1.10–1.76) for the re-offending outcomes. Only four of the ten RSAT evaluations assessed a drug use outcome; two of these evaluations indicated that participants in RSAT-funded programs had lower drug use rates than non-participants, and the other two evaluations found just the opposite. Not surprisingly, given these conflicting results, the overall mean odds ratio from these four evaluations was 1.02 (95% CI 0.48–2.15), which indicates no difference. Thus, the existing evidence indicated that RSAT-funded programs were associated with statistically significant reductions in re-

Table 1 Odds ratio by type of treatment and type of outcome

	Offending			Drug use		
Type of program	Mean ES	95% CI	k[a]	Mean ES	95% CI	k
TC	1.38*	1.17–1.62	30	1.41[#]	0.95–2.09	12
Counseling	1.50*	1.25–1.79	25	0.78	0.35–1.73	3
Narcotic maintenance	0.84	0.54–1.29	5	1.95	0.87–4.40	4
Boot camp	1.10	0.62–1.96	2	0.56	0.15–2.06	1
Ambiguous	1.59	0.98–2.58	3	—	—	–

[a] Number of odds ratios
*p<0.05
[#] p<0.10

offending, but the limited available evidence was decidedly mixed in regards to drug use outcomes.

Likewise, evaluations of counseling programs indicated that these programs were associated with statistically significant reductions in re-offending but that these programs' effects on drug use were ambiguous. In particular, the 25 evaluations of counseling programs exhibited the largest mean odds ratio on re-offending outcomes (1.50, 95% CI 1.25–1.79). Once again, the available evidence is scant in regards to drug use, with only three evaluations reporting drug use outcomes. Two of the three effect sizes were greater than, 1 indicating that program participants had less post-release drug use; however, one effect size was substantially less than 1. Combined, these three odds ratios yielded a mean odds ratio of 0.77 (95% CI 0.35–1.67).

A particularly small number of evaluations assessed the effectiveness of incarceration-based narcotic maintenance or boot camp programs for drug offenders. The handful of extant evaluations of narcotic maintenance programs indicated that participation in maintenance programs was associated with substantially significant reductions in drug use but not re-offending. All four of the odds ratios for drug use outcomes indicated that participation in narcotic maintenance programs produced lower post-release drug use than not-participating; this random-effect mean is substantively large, 1.94, but it was not statistically significant (p=0.11). By contrast, three of the four odds ratios for offending indicated that participants were more likely to re-offend than were non-participants. The two available evaluations of boot camps indicated that these programs were not associated with statistically significant reductions in either re-offending or drug use. Both of the evaluations measuring re-offending outcomes found that boot camp participants had slightly lower rates of re-offending than non-participants did (i.e., both evaluations had logged odds ratios less than 0.14).

As further evidence of the importance of primary type of treatment, we found that the association between odds ratio and several moderator variables depended on whether the odds ratio came from the evaluation of a TC or of a counseling program. Therefore, we conducted separate moderator analyses for TC and counseling programs in a series of parallel analyses. The odds ratios concerning boot camp, narcotic maintenance, and ambiguous programs were set aside for these analyses.

Table 2 examines variation in the odds ratios for offending by coded methodological features. One of the most important findings from Table 2 is that the vast majority of the mean odds ratios were statistically significant, which suggests that the finding that TC participants had lower rates of re-offending than did non-participants was largely robust to methodological variations. The first moderator variable, "overall method quality," was a four-point ordinal measure of the internal validity of each evaluation. This four-point categorization was similar to the University of Maryland's Scientific Methods Scale (see Farrington et al. 2006). The lowest level of method quality was "weak quasi-experimental designs." These studies utilized a comparison that lacked comparability to the treatment group before the intervention. The next level of method quality, "standard quasi-experiment," was assigned to evaluations characterized by research designs that used a comparison group that was slightly different from the treatment group on important observed variables before the intervention. "Rigorous quasi-experiments" were characterized as evaluations involving treatment and comparison groups that were highly

Table 2 Odds ratio for offending, by methodological features: therapeutic communities

Coded feature	Mean ES	95% CI	k^a
Overall method quality[+]			
Weak quasi-experiment	1.03	0.79–1.35	6
Standard quasi-experiment	1.40*	1.17–1.69	13
Rigorous quasi-experiment	1.39*	1.09–1.77	9
Experimental design	1.90*	1.22–2.97	2
Type of comparison group[†]			
Randomly assigned	1.90*	1.25–2.88	2
Eligible but not referred to treatment	1.40*	1.20–1.65	17
Eligible but declined treatment	0.79	0.54–1.16	2
Historical comparison group	1.00	0.67–1.50	2
Offenders from other jurisdiction/facility	1.20	0.74–1.93	2
Combination of the above	1.54*	1.16–2.05	5
Overall attrition apparent			
No	1.31*	1.12–1.54	23
Yes	1.48*	1.01–2.16	5
Differential attrition apparent			
No	1.37*	1.17–1.61	23
Yes	1.31	0.85–2.01	4
Published[†]			
No	1.16[#]	1.00–1.36	18
Yes	1.69*	1.39–2.05	12

[a] Number of odds ratios
*$p<0.05$
[#] $p<0.10$
[†] Difference between means is statistically significant at $p<0.05$
[+] Difference between means is statistically significant at $p<0.10$

comparable in important observed variables (e.g., age, gender, prior criminal history, prior drug use) or as evaluations that employed slightly different treatment and comparison groups but also used multivariate analyses that controlled for pre-existing differences in important variables. The highest level of method quality, "experimental designs," randomly assigned research participants to conditions and did not have attrition problems.[3]

Most of the available evaluations were methodologically weak. Of the 30 TC evaluations, 63% were rated as weak quasi-experiments (20%) or as standard quasi-experiments (43%). Nine evaluations (30%) were rated as rigorous quasi-experiments, and two (7%) were rated as experimental designs. Thus, in absolute terms, this body of research was rated as methodologically weak; yet, in relative terms, these evaluations were comparable in methodological rigor to other criminal justice research. In particular, Weisburd et al. (2001) rated the methodological rigor of a sample of 308 evaluations of criminal justice interventions using a scale very similar to the one used in our research. Those authors reported that 73% of evaluations were either weak (level 2) or standard (level 3) quasi-experiments and that 24% were either rigorous

[3] We coded two types of attrition problems: total and differential. Total attrition problems were defined as overall attrition of 20% or greater, or if the primary authors indicated that attriters differed substantially from non-attriters. Differential attrition problems were defined similarly; that is, differential attrition of 20% or greater, or if the primary authors indicated that attrition substantially reduced the comparability of the treatment and comparison group.

quasi-experiments (level 4) or randomized experiments (level 5); in comparison, the corresponding percentages for our research were 63% and 37%.

A common concern when one is dealing with a body of research that is generally methodologically weak is that this research may systematically overestimate treatment effects. In other words, perhaps methodologically weak evaluations are more likely to find treatment effects than are more rigorous evaluations. The research of Weisburd et al. (2001) research found precisely this pattern of results. Our synthesis of TC evaluations, however, reveals a different pattern of results. Table 2 indicates that the mean odds ratios for the three highest levels of method quality were all statistically significant, indicating that the effectiveness of TC programs was not confined to only methodologically weak evaluations. Further, the methodological quality variable was statistically associated with effect size; all of the three higher levels of rigor had mean effect sizes that were statistically larger than the lowest level of rigor, and evaluations with the highest level of rigor differed statistically from the three lower levels. Thus, evaluations with the highest level of rigor produced the largest treatment effects.

Two additional methodological factors, type of comparison group and publication status, had a statistically significant relationship to effect size. Evaluations that formed the comparison group via random assignment had the largest odds ratios, whereas, evaluations that utilized individuals who declined to participate in TC treatment had the lowest odds ratios. Post-hoc analyses indicated that evaluations composed of decliners had statistically smaller odds ratios than did the other comparison groups, and evaluations that utilized comparison groups formed by randomization had a larger mean odds ratio than did either evaluations that utilized a comparison group of decliners, or offenders from other jurisdictions/facilities. This set of findings, once again, suggests that evidence of the effectiveness of TC programs is not confined to methodologically suspect evaluations (i.e., evaluations that utilized comparison group composed of decliners or offenders from other jurisdictions/facilities or time periods).

Another methodological characteristic that had a statistically significant association with effect size was publication status. Specifically, evaluations coded from published studies exhibited statistically larger effect sizes than evaluations from unpublished studies. This finding is an indication of publication bias in evaluations of TC programs. As a further test for the presence of publication bias in these evaluations, we conducted specialized statistical tests for publication bias. Specifically, we conducted both the Begg and Mazumdar (1994) and the Egger et al. (1997) tests for publication bias. The more statistically powerful Egger et al. method found evidence of publication bias. Given this finding, we conducted Duvall and Tweedie's (2000) "trim and fill" method for accounting for publication bias. This procedure added seven effect sizes to the distribution, which lowered the mean random effects odds ratio to 1.21, with a 95% CI of 1.08 to 1.36 (Q=252.90, df=36, p<0.001); this publication bias-adjusted mean odds ratio translates into a re-offending rate of approximately 45% for the treatment group, if we continue to make the convenient assumption that 50% of the comparison group re-offends.

Table 3 presents the results of a similar bivariate analysis between the odds ratios for re-offending and sample/treatment characteristics for the odds ratio from evaluations of TC programs. Three sample characteristic were coded: age group

Table 3 Odds ratios for offending, by sample/treatment features: therapeutic communities

Coded feature	Mean ES	95% CI	k^a
Age group of sample			
Adults	1.37*	1.18–1.60	27
Juveniles	1.47	0.89–2.43	2
Gender composition of sample			
All female	1.65*	1.14–2.39	6
Mixed (male and female)	1.23	0.84–1.79	4
All male	1.36*	1.13–1.64	18
Offender type			
Non-violent offenders	1.49*	1.24–1.79	15
Mixed (violent and non-violent)	1.28*	1.02–1.62	9
Mandatory aftercare			
No	1.31*	1.07–1.59	14
Yes	1.51*	1.16–1.95	9
Treatment location			
Prison	1.35*	1.16–1.56	27
Jail	1.56#	0.94–2.60	3
Length of treatment			
90 days or more	1.45*	1.26–1.68	22
Fewer than 90 days	1.15	0.79–1.67	3
Strictly voluntary treatment			
No	1.32*	1.08–1.61	8
Yes	1.57*	1.35–1.84	16
Program maturity			
New program (less than 1 year)	1.34*	1.10–1.64	14
Developing program (1–3 years)	1.18	0.79–1.77	4
Established program (3+ years)	1.45*	1.15–1.83	11

[a] Number of odds ratios
*$p<0.05$
#$p<0.10$

(juvenile or adult), gender composition of sample, and type of offender (non-violent or a mixture of violent and non-violent offenders). Five treatment characteristics were coded: the presence of a mandatory aftercare component, treatment location (prison or jail), length of treatment (more than 90 days or fewer than 90 days), nature of participation (strictly voluntary vs at least some non-voluntary participation), and program maturity. These analyses found that none of the coded sample or treatment characteristics displayed a statistically or substantively significant relationship with effect size. Perhaps the most striking finding from Table 3 is that nearly all of the mean odds ratios were statistically significant, which suggests that evidence of the effectiveness of TC programs is largely robust to sample and program variations. Stated differently, participation in TC programs were associated with lower rates of re-offending, regardless of the age group, gender composition, offender type, use of coercion to compel participation, and several other factors.

A parallel set of analyses was conducted for counseling programs (Tables 4 and 5). Once more, most evaluations were methodologically weak. Nearly three-quarters (72%) of evaluations were rated as either "weak" or "standard" quasi-experiments, which again is fairly typical of criminal justice research, as revealed by Weisburd et

Table 4 Odds-ratios for offending , by methodological features: counseling programs

Coded feature	Mean ES	95% CI	k^a
Overall method quality			
Weak quasi-experiment	1.82*	1.24–2.66	8
Standard quasi-experiment	1.49*	1.08–2.06	10
Rigorous quasi-experiment	1.33	0.86–2.06	5
Experimental design	1.09	0.52–2.30	2
Type of comparison group			
Randomly assigned	1.09	0.51–2.32	2
Eligible but not referred for treatment	1.63*	1.21–2.20	11
Historical comparison group	1.32	0.81–2.16	5
Offenders from other jurisdiction/facility	1.54	0.83–2.86	3
Combination of the above	1.30	0.62–2.70	2
Overall attrition apparent			
No	1.63*	1.28–2.07	18
Yes	1.15	0.77–2.22	6
Differential attrition apparent†			
No	1.68*	1.37–2.05	20
Yes	0.71	0.43–1.16	3
Published			
No	1.51*	1.07–2.18	9
Yes	1.49*	1.15–1.94	16

[a] Number of odds-ratios

* $p<0.05$

† Difference between means is statistically significant at $p<0.05$.

al. (2001). This lack of methodological rigor was particularly problematic, as evaluations rated higher on this scale exhibited non-statistically significant mean odds ratios. In particular, evaluations rated as "rigorous" quasi-experiments or "experimental designs" exhibited mean effect sizes of 1.33 (95% CI 0.86–2.06) and 1.09 (95% CI 0.52–2.30), respectively, neither of which was statistically significant. While the statistical test comparing the mean odds ratios for the various levels of methodological rigor was not statistically significant, substantively, this finding suggests that the strongest evidence of the effectiveness of counseling programs in reducing re-offending came from methodologically weak evaluations—a finding that parallels that of Weisburd et al. (2001). The only methodological variable that had a statistically significant association with the magnitude of the odds ratios was differential attrition. Evaluations that had more than 20% differential overall attrition had a mean odds ratio that was statistically less than those without differential attrition problems.

The moderator distinguishing published from unpublished studies found no difference in the magnitude of the effect for counseling programs. This finding comports with specialized tests for publication bias (i.e., the Begg and Mazumdar, and Egger et al. publication bias tests). These findings converged in concluding that there was no evidence of publication bias among evaluations of counseling programs.

In regards to coded sample features, two of the three moderator variables were statistically related to effect size (Table 5). Evaluations of samples of adults had a

Table 5 Odds-ratios for offending, by sample and treatment features: counseling programs

Coded feature	Mean ES	95% CI	k^a
Age group of sample†			
Adults	1.53*	1.41–1.67	19
Juveniles	1.16	0.92–1.46	3
Gender composition of sample†			
All female	2.94*	1.74–4.97	3
Mixed (male and female)	1.01	0.69–1.48	6
All male	1.67*	1.26–2.21	11
Offender type			
Non–violent offenders	1.48*	1.18–1.86	11
Mixed (violent and non-violent)	1.26#	0.98–1.62	12
Primary emphasis of counseling†			
Cognitive behavioral therapy	1.12	0.74–1.68	5
12-step/drug education	1.54*	1.00–2.36	5
Multifaceted counseling	1.21	0.85–1.72	7
Unknown emphasis	2.14*	1.54–2.97	8
Mandatory aftercare			
No	1.45*	1.14–1.85	20
Yes	1.82*	1.10–3.03	4
Treatment location			
Prison	1.56*	1.20–2.04	16
Jail	1.42#	1.00–2.01	9
Length of treatment			
90 days or more	1.44#	0.98–2.12	10
Fewer than 90 days	1.58*	1.12–2.23	11
Strictly voluntary treatment+			
No	1.07	0.60–1.92	4
Yes	1.75*	1.26–2.44	14
Program maturity			
New program (less than 1 year)	1.08	0.82–1.41	8
Developing program (1–3 years)	1.43	0.70–2.92	2
Established program (3+ years)	1.79*	1.36–2.37	9

[a] Number of odds-ratios
* $p<0.05$
# $p<0.10$
† Difference between means is statistically significant at $p<0.05$.
+ Difference between means is statistically significant at $p<0.10$.

statistically larger mean odds ratio than evaluations that used samples of juvenile. The gender composition moderator was also statistically related to odds ratio. In particular, evaluations that were composed solely of female offenders had a larger mean odds ratio than evaluations of samples of men or of mixed genders; in fact, post-hoc contrasts indicated that all three mean odds ratios statistically differed from one another.

In addition to the moderator variables coded from evaluations of TC programs, for the evaluations of counseling programs we also coded primary emphasis of each program. Unfortunately, many evaluations were not described in detail by the primary authors, and, as a consequence, our ability to code program variation was limited. We were able to categorize 17 of the 25 programs evaluated into three types: cognitive behavioral therapy, 12-step programming/drug education, and multifaceted (i.e., programs that combined the above approaches without a clearly discernible

primary emphasis); the remaining eight evaluations did not supply enough information to be reliably classified and are listed as "Unknown emphasis" in Table 5. The evaluations classified into the unknown emphasis category exhibited the largest mean odds ratio (2.14, with 95% CI of 1.54–2.97), whereas, evaluations of multifaceted programs yielded the smallest mean odds ratio (1.21, with 95% CI of 0.85–1.72). Post-hoc contrasts indicated that the evaluations in the unknown emphasis category had statistically larger mean odds ratio than did either evaluations of cognitive behavioral therapy or multifaceted counseling; none of the other categories differed statistically from one another. Thus, the strongest evidence of the effectiveness of counseling programs comes from evaluations without a clear depiction of the treatment provided to program participants. Interestingly, the evidence suggests that the most popular incarceration-based drug treatment modality, 12-step/drug education, exhibited a substantively and statistically significant mean odds ratio of 1.54 (95% CI 1.00–2.36).

One other treatment characteristic was related to odds ratio. Evaluations of programs that were strictly voluntary had a statistically larger mean odds ratio than did other evaluations of counseling programs. In fact, evaluations in which at least some program participants did not volunteer for treatment produced a mean odds ratio (1.07) that did not differ statistically or substantively from 1.00, which suggests that in such evaluations program participants and non-participants had similar rates of re-offending.

Discussion

In concordance with the authors of existing reviews (e.g., Wilson et al. 2005; Pearson and Lipton 1999), we found no evidence that participation in boot camp programs reduced post-release offending or drug use. While the number of existing evaluations of boot camp programs for drug offenders was very small, given the consistency of our findings to other research on boot camps, it appears unlikely that boot camp programs generally reduce recidivism.

The evidence concerning the effectiveness of narcotic maintenance programs was mixed. We found no evidence of the effectiveness of such programs in reducing re-offending; in fact, program participants had somewhat higher re-offending rates than non-participants in four out of the five evaluations. By contrast, all existing evaluations found somewhat lower rates of post-release drug use among participants than non-participants, and the random-effects mean odds ratio for these four evaluations was substantively large (1.95). Thus, incarceration-based narcotic maintenance programs may reduce drug use, but the scant existing evidence does not indicate that such program reduce re-offending. Our ability to draw firm conclusions in this area of research, however, is undermined by the limited statistical power of our analyses. Continued research investigating the effectiveness of these programs would be a significant contribution to the knowledge base.

The most consistent evidence of treatment effectiveness came from evaluations of TC programs. These programs consistently showed post-release reductions in re-offending and drug relapse. This finding was robust to methodological variation. In fact, even among the most rigorous evaluations, participation in TC programs was

consistently related to reductions in re-offending. We also found that TCs were effective in several different types of samples, which suggests that TCs can be applied to a wide-range of offenders. However, there was evidence of publication bias in this area of research, and this bias appeared to cause the effectiveness of TC programs to be over-estimated. This possibility tempers our findings.

Ten of the 30 evaluations of TC programs were evaluations of RSAT-funded programs. These ten evaluations found evidence of effectiveness of RSAT programs in reducing re-offending. Four of these evaluations assessed the impact of the RSAT programs' effectiveness on drug use. This limited evidence produced mixed results: two evaluations found that participants had lower post-release rates of drug use, and two evaluations found that participants had higher post-release rates of drug use, than did non-participants. Thus, such evaluations are as likely to find reductions in drug use as they are to find increases.

Similarly, the evidence regarding counseling programs indicated that these programs were effective in reducing re-offending, but, once again, the limited evidence available on these programs' effects on drug use was unclear. The analyses of odds ratios from evaluations of counseling programs revealed that, while counseling programs had a larger mean odds ratio than evaluations of TC programs on offending outcomes, the positive effects of counseling programs were more contingent. Evidence of the effectiveness of counseling programs was strongest in evaluations that were rated as methodologically weak, composed of adult offenders, and strictly voluntary. Regrettably, counseling interventions generally were poorly described by the primary authors, and, as a result, we were unable to code precise program features that may have distinguished the most effective programs.

Interestingly, evaluations of both TC and counseling programs found that treatment was somewhat more effective in samples of women than in samples of either men or mixed genders. This finding is particularly interesting, as incarcerated female offenders are disproportionately drug involved, and female offenders traditionally have had limited access to high-quality treatment programs (see e.g., Pollock 2004).

A surprising finding was the smaller effects of both TC and counseling programs on drug-relapse measures than on measures of re-offending. A presumption of these programs is that their effectiveness in reducing re-offending is mediating by reductions in drug use. As such, we would expect drug-relapse effects to be as large as, or larger than, re-offending effects. An analysis of the relationship between drug relapse odds ratios and re-offending odds ratios showed a strong relationship between these two (unweighted correlation of 0.56 and a weighted correlation of 0.64), indicating that the larger a program's effect on drug-relapse the larger the effect on recidivism (and vice versa), consistent with the notion that drug use mediates criminality. This does not resolve, however, the counterintuitive finding of lower drug-relapse effects. Two possibilities come to mind. First, although official measures of recidivism are known to be only modestly reliable indicators of criminal behavior, they may be more reliable than measures of drug-relapse. If this is true, then the latter are attenuated to a greater degree than the former. Second, drug use may not mediate the relationship between these programs and criminal behavior. These programs may successfully change criminogenic attitudes, attachment of antisocial peers, maladaptive cognitive processes, or other psychosocial character-

istics of the offenders that maintain or foster their criminality, and, to a lesser degree, their drug use. Additional research is needed that examines the role of reducing an offender's drug use in reducing future non-drug-use criminal behavior.

We believe that the implications of this research for policy makers are clear. Policy makers seeking effective interventions for incarcerated substance abusers are most likely to find success with programs that focus on the multiple problems of substance abusers in an intensive manner, such as TC programs. Policy makers should expect smaller treatment benefits from less intensive treatment programs. Further, there is no evidence that correctional boot camps targeted at substance abusers reduce either post-release offending or drug use, and, thus, policy makers should not expect these programs to produce reductions in recidivism.

We believe that this research also has implications for researchers. Specifically, we believe that while the extant research clearly supports the effectiveness of certain programs, there is a lack of understanding concerning which particular components of treatment programs are most important and which combination of components are most effective. Further, the general methodological weakness of this area of research makes their findings vulnerable to alternative explanations (i.e., reductions in recidivism could be due to factors other than the intervention). Beneficial future research will address these issues.

Acknowledgments We would like to gratefully acknowledge the Nordic Campbell Center and the Jerry Lee Foundation for partial support of this project. We would also like to thank Matthew Makarios, who helped code numerous evaluations.

References

*Anglin, M. D., Prendergast, M. L., Farabee, D., & Cartier, J. (2002). *Final report on the substance abuse program at the California Substance Abuse Treatment Facility and state prison at Corcoran*. Los Angeles, CA: The UCLA Integrated Substance Abuse Program.

*Aos, S. (2004). *Washington State's Family integrated transitions program for juvenile offenders: Outcome evaluation and benefit-cost analysis*. Olympia, WA: Washington State Institute for Public Policy.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*(4), 1088–1101.

Bennett, T. (1998). *Drugs and crime: The results of research on drug testing and interviewing arrestees*. London: Home Office.

Berggren, O., & Svard, H. (1990). *Osteraker Project: A further follow-up of the drug misuser treatment programme at Osteraker Prison*. Sweden: Swedish Prison and Probation Administration.

Bureau of Justice Assistance. (2005). *Residential substance abuse treatment for state prisoners (RSAT) program*. Washington, DC: Bureau of Justice Assistance.

Bureau of Justice Statistics. (1995). *Drugs and crime facts, 1994: A summary of drug data published in 1994*. Washington, DC: Bureau of Justice Statistics.

Cooper, H., & Hedges, L. V. (Eds). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.

*Crundall, I., & Deacon, K. (1997). A prison-based alcohol use education program: Evaluation of a pilot study. *Substance Use & Misuse, 32*(6), 767–777.

*Daley, M., Love, C. T., Shepard, D. S., Petersen, C. B., White, K. L., & Hall, F. B. (2004). Cost effectiveness of Connecticut's in-prison substance abuse treatment. *Journal of Offender Rehabilitation, 39*(3), 69–92.

*Darabi, G. A. (1992). *Psychological impact of tier programs: An outcome evaluation*. Tallahassee, FL: Bureau of Planning, Research, and Statistics.

Dolan, K. A., Shearer, J., MacDonal, M., Mattick, R. P., Hall, W., & Wodak, A. D. (2003). A randomised controlled trial of methadone maintenance treatment versus wait list control in an Australian prison system. *Drug and Alcohol Dependence, 72*(1), 59–65.

*Dowden, C., & Blanchette, K. (2002). An evaluation of the effectiveness of substance abuse programming for female offenders. *International Journal of Offender Therapy and Comparative Criminology, 46*(2), 220–230.

*Dugan, J. R., & Everett, R. S. (1998). An experimental test of chemical dependency therapy for jail inmates. *International Journal of Offender Therapy and Comparative Criminology, 42*(4), 360–368.

Duvall, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98.

Dynia, P., & Sung, H.-E. (2000). The safety and effectiveness of diverting felony drug offenders to residential treatment as measured by recidivism. *Criminal Justice Policy Review, 11*(4), 299–311.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

*Eisenberg, M., & Fabelo, T. (1996). Evaluation of the Texas correctional substance abuse treatment initiative: The impact of policy research. *Crime & Delinquency*, *42*(2), 296–308.

*Eisenberg, M., Riechers, L., & Arrigona, N. (2001). *Evaluation of the performance of the Texas Department of Criminal Justice Rehabilitation Tier programs*. Austin, TX: Criminal Justice Policy Council.

Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2006). The Maryland Scientific Methods Scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds), *Evidence-based crime prevention* (pp. 13–21). New York, NY: Routledge.

Field, G. (1985). The Cornerstone program: A client outcome study. *Federal Probation, 49*(2), 50–55.

Field, G. (1989). *A study of the effects of intensive treatment on reducing the criminal recidivism of addicted offenders*. Unpublished Manuscript: Oregon Correctional Treatment Programs.

*Finigan, M. W., Barron, N., & Carey, S. (2003). Effectively assessing and preparing inmates for community substance abuse treatment: The Portland Target Cities Project in-jail intervention. In R. C. Stephens, C. K. Scott, & R. D. Muck (Eds), *Clinical assessment and substance abuse treatment: The Target Cities experience* (pp. 165–178). Albany, NY: State University of New York Press.

Flanagan, C., Arsovska, J., Giaime, A. J., Goril, M., Kahl, C., Król, A., & Moore, B. (2004). *Digest of research on drug use and HIV/AIDS in prisons*. London: The European Network for Drug Services in Prison, Cranstoun Drug Services.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*(4), 575–607.

*Gordon, J. A. (2002). *Barrett Juvenile Correctional Center: Is it effective*. Richmond, VA: Virginia Commonwealth University.

*Gransky, L. A., & Jones, R. J. (1995). *Evaluation of the post-release status of substance abuse program participants*. Chicago, IL: Illinois Criminal Justice Information Authority.

Guerin, P. (2002). *Outcome evaluation of the New Mexico Corrections Department Genesis Residential Substance Abuse Treatment program for state prisoners*. Albuquerque, NM: Institute for Social Research, University of New Mexico.

*Hanson, G. (2000). *Pine Lodge intensive inpatient treatment program*. Olympia, WA: Planning and Research Section, Washington State Department of Corrections.

*Hartmann, D. J., Wolk, J. L., Johnston, L. S., & Colyer, C. J. (1997). Recidivism and substance abuse outcomes in a prison-based therapeutic community. *Federal Probation, 61*(4), 18–25.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychological Bulletin, 117*(1), 167–178.

Horney, J., Osgood, D. W., & Marshall, I. H. (1995). Criminal careers in the short-term: Intra-individual variability in crime and its relation to local life circumstances. *American Sociological Review, 60*, 655–673.

*Hughey, R., & Klemke, L. W. (1996). Evaluation of a jail-based substance abuse treatment program. *Federal Probation, 60*(4), 40–44.

*Hume, S., & Gorta, A. (1988). *Results of community urinalyses for clients on the NSW Prison Methadone Program*. Sydney, NSW: Research and Statistics Division, NSW Department of Corrective Services.

*Hume, S., & Gorta, A. (1989). *Effects of the N.S.W. prison methadone program on criminal recidivism and retention in methadone treatment*. New South Wales, Australia: New South Wales Department of Corrective Services.

*Inciardi, J. A., Martin, S. S., Butzin, C. A., Hooper, R. M., & Harrison, L. D. (1997). An effective model of prison-based treatment for drug-involved offenders. *Journal of Drug Issues, 27*(2), 261–279.

*Johnson, S. L., van de Ven, J. T. C., & Grant, B. A. (2001). *Institutional methadone maintenance treatment: Impact on release outcome and institutional behaviour (Research Rep. No. 119)*. Ottawa: Research Branch, Correctional Service of Canada.

Karberg, J. C., & James, D. J. (2005). *Substance dependence, abuse, and treatment of jail inmates, 2002*. Washington, DC: Bureau of Justice Statistics.

*Kelly, W. R. (2001). *An outcome evaluation of the Texas Youth Commission's Chemical Dependency Treatment Program*. Austin, TX: University of Texas.

*Kinlock, T. W., Battjes, R. J., & Schwartz, R, The MTC Project Team. (2005). A novel opioid maintenance program for prisoners: Report of post-release outcomes. *American Journal of Drug and Alcohol Abuse, 31*, 433–454.

*Klebe, K. J., & O'Keefe, M. (1998). *Outcome evaluation of the Crossroads to Freedom House and Peer I therapeutic communities*. Colorado Springs, CO: University of Colorado.

Knight, K., & Hiller, M. L. (1997). Community-based substance abuse treatment: A 1-year outcome evaluation of the Dallas County Judicial Treatment Center. *Federal Probation, 61*(2), 61–68.

*Knight, K., Simpson, D. D., & Hiller, M. (1999). Three-year reincarceration outcomes for in-prison therapeutic community treatment in Texas. *The Prison Journal, 79*(3), 337–351.

Kooyman, M. (1993). *The therapeutic community for addicts: intimacy, parent involvement, and treatment success.* Berwyn, PA: Swets & Zeitlinger.

*Kunitz, S. J., Woodall, W. G., Zhao, H., Wheeler, D. R., Lillis, R., & Rogers, E. (2002). Rearrest rates after incarceration for DWI: A comparative study in a Southwestern US county. *American Journal of Public Health, 92*(11), 1826–1831.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Lipton, D. S., Falkin, G. P., & Wexler, H. K. (1992). Correctional drug abuse treatment in the United States; An overview. In C. G. Leukefeld, and F. M. Tims (Eds), *Drug abuse treatment in prisons and jails*. Rockville, MD: National Institute on Drug Abuse.

*Little, G. L., Robinson, K. D., & Burnette, K. D. (1991). Treating drug offenders with moral reconation therapy: A three-year recidivism report. *Psychological Reports, 69*(3), 1151–1154.

MacKenzie, D. L., Browning, K., Skroban, S. S., & Smith, D. A. (1999). The impact of probation on the criminal activities of offenders. *Journal of Research in Crime and Delinquency, 36*(4), 423–453.

Magura, S., Rosenblum, A. L. C., & Joseph, H. (1993). The effectiveness of in-jail methadone maintenance. *Journal of Drug Issues, 23*(1), 75–99.

*Martin, C., Player, E., & Liriano, S. (2003). Results of evaluations of the RAPt drug treatment programme. In M. Ramsey (Ed), *Prisoners' drug use and treatment: Sever research studies*. London: Home Office Research.

*Miller, J. M., & Koons-Witt, B. (2003). *Outcome evaluation of the South Carolina residential substance abuse treatment program for state prisoners*. Columbia, SC: University of South Carolina.

*Mosher, C., & Phillips, D. (2002). *Final report on the program evaluation of the Pine Lodge Pre-Release Residential Therapeutic Community for women offenders in Washington State*. Pullman, WA: Washington State University.

*Nash, J. E. (2000). *Final report of outcomes for Ozark Correctional Center drug treatment program*. Springfield, MO: Center for Social Sciences and Public Policy Research.

*Oregon Department of Corrections. (1994). *Comparison of outcomes and costs residential and outpatient treatment programs for inmates alcohol and drug, mental health, sex offender, and social skills treatment*. Oregon: Oregon Department of Corrections.

Pealer, J. A., Latessa, E. J., & Winesburg, M. (2002). *Final report: Mohican Youth Center RSAT outcome evaluation*. Cincinnati, OH: Center for Criminal Justice Research, University of Cincinnati.

Pearson, F. S., & Lipton, D. S. (1999). A meta-analytic review of the effectiveness of corrections-based treatment for drug abuse. *The Prison Journal, 79*(4), 384–410.

*Pelissier, B., Rhodes, W., Saylor, W., Gaes, G., Camp, S. D., Vanyur, S. D., & Wallace, S. (2000). *TRIAD drug treatment evaluation project final report of three-year outcomes: Part I*. Washington, D.C.: Federal Bureau of Prisons, Office of Research and Evaluation.

Pernaen, K., Cousineau, M.-M., Brochu, S., & Sun, F. (2002). *Proportions of crimes associated with alcohol and other drugs in Canada*. Ottawa: Canadian Centre on Substance Abuse.

*Peters, R. H., Kearns, W. D., Murrin, M. R., Dolente, A. S., & May, R. L. II. (1993). Examining the effectiveness of in-jail substance abuse treatment. *Journal of Offender Rehabilitation, 19*(3/4), 1–39.

Pollock, J. M. (2004). *Prisons and prison life: Costs and consequences*. Los Angeles: Roxbury.

*Porporino, F. J., Robinson, D., Millson, B., & Weekes, J. R. (2002). An outcome evaluation of prison-based treatment programming for substance users. *Substance Use & Misuse, 37*(8–10), 1047–1077.

*Porter, R. (2002). *Breaking the cycle: technical report*. New York: Vera Institute of Justice.

*Prendergast, M., Hall, E., & Wellisch, J. (2003). *An outcome evaluation of the Forever Free substance abuse treatment program: One-year post-release outcomes*. Santa Monica, CA: Drug Abuse Research Center.

Prendergast, M. L., Wellisch, J., & Wong, M. M. (1996). Residential treatment for women parolees following prison-based drug treatment: Treatment experiences, needs and services, outcomes. *The Prison Journal, 76*(3), 253–274.

Sánchez-Meca, J., Chacón-Moscoso, S., & Marín-Martínez, F. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8*, 448–467.

Schippers, G. M., van den Hurk, A. A., Breteler, M. H. M., & Meerkerk, G.-J. (1998). Effectiveness of a drug-fee detention treatment program in a Dutch prison. *Substance Use & Misuse, 33*(4), 1027–1046.

*Sealock, M. D., Gottfredson, D. C., & Gallagher, C. A. (1997). Drug treatment for juvenile offenders: Some good and bad news. *Journal of Research in Crime and Delinquency, 34*(2), 210–236.

Shaw, J. W., & MacKenzie, D. L. (1992). The one-year community supervision performance of drug offenders and Louisiana DOC-identified substance abusers graduating from shock incarceration. *Journal of Criminal Justice, 20*(6), 501–516.

Shewan, D., Macpherson, A., Reid, M. M., & Davies, J. B. (1996). The impact of the Edinburgh prison drug reduction programme. *Legal and Criminological Psychology, 1*(1), 83–94.

*Siegal, H. A., Wang, J., Falck, R. S., Rahman, A. M., & Carlson, R. G. (1997). *An evaluation of Ohio's prison-based therapeutic community treatment programs for substance abusers: Final report*. Dayton, OH: Wright State University, School of Medicine.

*Smith, C. J. (1996). *The California Civil Addict program: An evaluation of implementation and effectiveness.* Unpublished doctoral dissertation, University of California, Irvine, Ann Arbor, MI.

Swartz, J. A., Lurigio, A. J., & Slomka, S. (1996). The impact of IMPACT: An assessment of the effectiveness of the Cook County Jail's substance abuse treatment program. *Crime and Delinquency, 42*(4), 553–573.

*Taxman, F. S., & Spinner, D. L. (1996). *The Jail Addiction Services (JAS) project in Montgomery County, Maryland*. College Park, MD: University of Maryland.

*Tunis, S., Austin, J., Morris, M., Hardyman, P., & Bolyard, M. (1995). *Evaluation of drug treatment in local corrections: Final report*. San Francisco, CA: National Council on Crime and Delinquency.

*Turley, A., Thornton, T., Johnson, C., & Azzolino, S. (2004). Jail drug and alcohol treatment program reduces recidivism in nonviolent offenders: A longitudinal study of Monroe County, New York's, jail treatment drug and alcohol program. *International Journal of Offender Therapy and Comparative Criminology, 48*(6), 721–728.

*Van Stelle, K. R., & Moberg, D. P. (2000). *Outcome evaluation of the Wisconsin Residential Substance Abuse Treatmen program: The mental illness-chemical abuse (MICA) program at Oshkosh Correctional Institution, 1998–2000*. Madison, WI: University of Wisconsin Medical School.

*Vaughn, M. S., Deng, F., & Lee, L.-J. (2003). Evaluating a prison-based drug treatment program in Taiwan. *Journal of Drug Issues, 33*(2), 357–384.

*Voas, R. B., & Tippetts, A. S. (1990). Evaluation of treatment and monitoring programs for drunken drivers. *Journal of Traffic Medicine, 18*, 15–26.

*Washington State Department of Corrections. (1988). *Substance abuse treatment program evaluation of outcomes and management report*. Olympia, WA: Washington State Department of Corrections.

Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice. *Annals of the American Academy of Political and Social Sciences, 578*, 50–70.

*Welsh, W. N. (2002). *Evaluation of prison based drug treatment in Pennsylvania*. Philadelphia, PA: Temple University.

*Wexler, H. K., Falkin, G. P., & Lipton, D. S. (1990). Outcome evaluation of a prison therapeutic community for substance abuse treatment. *Criminal Justice and Behaviors, 17*(1), 71–92.

*Wexler, H. K., Melnick, G., Lowe, L., & Peters, J. (1999). Three-year reincarceration outcomes for Amity in-prison therapeutic community and aftercare in California. *Prison Journal, 79*(3), 321–336.

Wilson, D. B., & MacKenzie, D. L. (2006). Boot camps. In B. C. Welsh & D. F. Farrington (Eds), *Preventing crime: What works for children, offenders, victims, and places* (pp. 73–86). New York: Springer.

Wilson, D. B., MacKenzie, D. L., & Mitchell, F. N. (2005). *Effects of correctional boot camps on offending*. A Campbell Collaboration systematic review: Available at: http://www.aic.gov.au/campbellcj/reviews/published.html.

*Winesburg, M., Latessa, E. J., & Pealer, J. A. (2002). *Final report: Noble Choices RSAT outcome evaluation*. Cincinnati, OH: Center for Criminal Justice Research, University of Cincinnati.

*Zhang, S. (2000). *An evaluation of the Los Angeles County juvenile drug treatment boot camp: Final report*. San Marcos, CA: California State University.

Zhang, Z. (2003). *Drug and alcohol use and related matters among arrestees, 2003*. Washington, DC: Arrestee Drug Abuse Monitoring Program.

Ojmarrh Mitchell is an assistant professor of Criminal Justice at the University of Cincinnati. Professor Mitchell earned his Ph.D. in Criminal Justice and Criminology from the University of Maryland with a doctoral minor in Measurement, Statistics, and Evaluation. His research interests include deviant self-identity, evaluation of correctional programs, race and sentencing, drugs and crime, measurement theory, and meta-analysis.

David B. Wilson, Ph.D. is an associate professor in the Administration of Justice Program at George Mason University. His research interests are the effectiveness of offender rehabilitation and crime prevention efforts, program evaluation methodology, quantitative methods, and meta-analysis.

Doris Layton MacKenzie is a professor of Criminology and Criminal Justice at the University of Maryland. Her most recent book is titled "What Works In Corrections: Examining the Criminal Activities of Offenders and Delinquents" (2006, Cambridge University Press). The work summarizes systematic reviews and meta-analyses of evaluations of correctional management strategies and rehabilitation programs. She has co-edited two recent books: "Correctional Boot Camps: Military Basic Training or a Model for corrections?" (2004, Sage Publications) and "Different Crimes, Different Criminals" (2006, Lexis/Nexis/Anderson). She has testified before US Senate and House committees and was invited to present her work at the United Nations, Vienna. She is a fellow and currently president-elect of the Academy of Experimental Criminology, vice-president of the American Society of Criminology (ASC) and past chair of the Division on Corrections and Sentencing, ASC. She has received the ASC Bloch and the Most Distinguished Scientist award from the Division on Corrections and Sentencing, ASC. In 2007, as a Fulbright Research Scholar, she is examining China's new community corrections programs.