Campbell Systematic Reviews
2006:11
First published: 18 September, 2006
Last updated: 18 September, 2006

The Effectiveness of Incarceration-Based Drug Treatment on Criminal Behavior

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

THE CAMPBELL COLLABORATION

Colophon

Title	The effectiveness of incarceration-based drug treatment on criminal behavior
Institution	The Campbell Collaboration
Authors	Mitchell, Ojmarrh Wilson, David B. MacKenzie, Doris L.
DOI	10.4073/csr.2006.11
No. of pages	58
Last updated	18 September, 2006
Citation	Mitchell O, Wilson DB, MacKenzie DL. The effectiveness of incarceration-based drug treatment on criminal behavior. Campbell Systematic Reviews 2006:11 DOI: 10.4073/csr.2006.11

Keywords	
Contributions	None stated.
Support/Funding	SFI Campbell, The Danish National Centre of Social Research, Denmark Jerry Lee Foundation, USA
Potential Conflicts of Interest	None of the authors have any financial interests in any existing or planned incarceration-based drug treatment or any competing types of interventions for drug using offenders.

Corresponding author	Ojmarrh Mitchell Division of Criminal Justice University of Cincinnati P.O. Box 210389 600 Dyer Hall Cincinnati, OH 45221-0389 USA E-mail: ojmarrh.mitchell@uc.edu

Campbell Systematic Reviews

The Campbell Collaboration (C2) was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. C2 offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of C2's editors, librarians, methodologists and external peer-reviewers contribute.

The Campbell Collaboration
P.O. Box 7004 St. Olavs plass
0130 Oslo, Norway
www.campbellcollaboration.org

Title and Authors

The Effectiveness of Incarceration-Based Drug Treatment on Criminal Behavior

Ojmarrh Mitchell (Corresponding Author), David B. Wilson, and Doris L. MacKenzie

Objectives

- To synthesize the extant evidence regarding the effectiveness of incarceration-based drug treatment in reducing drug relapse and recidivism.
- Specifically, this systematic review addressed:
  - Are incarceration-based drug treatment programs effective in reducing recidivism and drug use?
  - Approximately how effective are these programs (i.e., what's the magnitude of the effect)?
  - Are there particular types of drug treatment programs that are especially effective or ineffective?

Findings

- Sixty-six independent evaluations met our eligibility criteria, the most important of which was the use of a comparison group.
- Approximately 83% of recidivism odds-ratios indicated that the drug treatment group had less recidivism than the comparison group. The mean odds-ratio was 1.37, indicating that on average participation in these drug treatment programs was associated with a modest reduction in post-treatment offending; assuming a 35% recidivism rate for the comparison group (which was the average recidivism rate for comparison groups), this overall mean odds-ratio translates into a recidivism rate of approximately 28% for the treatment group.
- The effectiveness of drug treatment varied by type of treatment:
  - Boot camps aimed at drug involved offenders were ineffective in reducing re-offending and drug relapse.
  - Narcotic maintenance programs did not exhibit reductions in re-offending or drug use, but the evidence in this area was scant.
  - Group counseling programs exhibited reductions in re-offending but not drug use.
  - Therapeutic communities (TCs) exhibited the strong and consistent reductions in drug relapse and recidivism.

Caveats and Qualifications

- The small number of independent evaluations of narcotic maintenance programs undermines our ability to draw firm conclusions about these programs.
- Many of the included evaluations were methodologically weak, which limits our ability to draw firm conclusions from these evaluations.
- Many of our analyses were limited by low statistical power.

Implications for Decision Makers

- Programs that intensively focus on the multiple problems of substance abusers, such as TCs, are most likely to reduce drug use and recidivism.
- Correctional boot camps for drug offenders cannot be justified in terms of reduced recidivism.

Implications for Research

- Beneficial future research will focus on determining which components of effective programs are important and <u>why</u> are such components important.
- Beneficial future research will employ methodologically rigorous techniques that reduce the likely of selection bias and other threats to internal validity.

The Effectiveness of Incarceration-Based Drug Treatment on Criminal Behavior[*]

Ojmarrh Mitchell[†]

David B. Wilson[‡]

Doris L. MacKenzie[§]

Submitted to the Campbell Collaboration, Criminal Justice Review Group
September 2006

[*] We would like to gratefully acknowledge the Nordic Campbell Center and the Jerry Lee Foundation for partial support of this project.

[†] Division of Criminal Justice, 600 Dyer Hall, P.O. Box 210389, University of Cincinnati, Cincinnati, OH 45221-0389, ojmarrh.mitchell@uc.edu.

[‡] Administration of Justice, 10900 University Blvd., MS 4F4, George Mason University, Manassas, VA 20110-2203.

[§] Department of Criminology and Criminal Justice, 2220 LeFrak Hall, University of Maryland, College Park, MD 20742-8235.

Table of Contents

List of Tables

List of Figures

1 Background for the Review

Research indicates that a substantial proportion of incarcerated offenders are drug dependent. Peters and colleagues (Peters, Greenbaum, Edens, Carter, and Ortiz, 1993) for example, reported that 56% of a sample of Texas inmates were diagnosed as having a substance abuse or dependence disorder during the 30 days prior to their incarceration. Similarly, a survey of jail inmates in Ohio found that 51% were currently drug dependent (Lo and Stephens, 2000). In fact, it is estimated that about 40% of all Americans who clearly need drug treatment are under the supervision of the criminal justice system (Gerstein and Harwood, 1990:7).

Drug dependence also appears to be common among incarcerated offenders in many other countries besides of the United States. While international research assessing drug dependence among incarcerated offenders is more limited, the existing evidence indicates that drug dependence is common among incarcerated offenders in many nations. For instance, Bennett (1998) found that 45% of a sample of incarcerated arrestees in five English cities reported being drug dependent at one point in their lives, and 33% reported being currently drug dependent. Likewise, 31% of inmates incarcerated in Canadian federal prisons and 43% of inmates incarcerated in provincial prisons were found to be drug dependent (Pernanen, Cousineau, Brochu, and Sun, 2002).

In the absence of effective substance abuse treatment, it is likely that a high proportion of these drug dependent offenders will persist in crime. In fact, statistics reported by the Bureau of Justice Statistics indicate that among probationers, frequent drug abusers were 53% more likely to be re-arrested than non-drug abusers (Bureau of Justice Statistics 1995: 26). As such, the period of time when an offender is incarcerated represents a crucial opportunity to prevent crime by intervening in the cycle of drug abuse and crime.

Several aspects of correctional facilities (i.e., prisons, jails) make incarceration-based substance abuse treatment attractive. First, the availability of drugs is more limited in correctional facilities than in the community, which facilitates detoxification and abstinence during treatment. Second, there is an abundance of time available to focus on treatment and introspection. Perhaps most importantly, correctional facilities have the capacity to mobilize considerable coercive force to encourage substance abusing offenders to engage in treatment, many of whom otherwise would not do so.

Incarceration-based drug treatment is diverse, encompassing a broad array of treatment programs, including group and individual psychotherapy, 12-step programs, methadone maintenance and punitive interventions, such as boot camps for drug abusing offenders. For our purposes, the defining features of these programs are that they target substance abusers, intend to reduce substance abuse and other criminal behaviors, and these interventions are based in a correctional facility. Evaluations of existing incarceration-based drug treatment programs predominantly focus on assessing the effectiveness of therapeutic communities (TCs) and group counseling programs (e.g., drug education, 12-step programs, such as AA/NA). A considerably smaller number of evaluations have assessed the effectiveness of boot camps targeted specifically at substance abusers or narcotic maintenance programs.

The individual components of TCs vary widely. Yet, several components appear to be common. First, in order to create an environment conducive to rehabilitation, residents in therapeutic communities are most commonly housed in a separate, distinct treatment unit away from non-participating inmates. Second, residents are instrumentally involved in running the therapeutic community including leading treatment sessions, monitoring other residents for rule compliance, maintaining the treatment unit, and resolving disputes. Third, staff and residents of TCs tend to be confrontational with rule

violators, but residents also are supportive of each other's struggles to make positive changes. Fourth, the guiding philosophy of TCs is that drug use is symptomatic of more general personal disorders, thus the focus of the treatment is on the underlying disorders and not drug abuse, per se.

Counseling programs are somewhat harder to characterize. Generally these programs incorporate elements of group counseling programs (e.g., 12-step programs such as AA/NA), life skills training, cognitive skills training, drug education, and adult basic (academic) education. A key commonality among these programs is their reliance on group based therapies, in which substance abuse and other common problems are discussed among peers in an effort to solve mutual issues. However, not all counseling programs rely on peer therapy; some counseling programs are individual-based where the client and a clinician work together to remedy drug problems. 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 most TC programs confrontations most often occur between correctional staff and inmates—with drill instructors disciplining any deviation from established codes of conduct. In theory, the harsh, rigorous nature of boot camp programs serve as a deterrent to future criminal conduct, and the content of these programs instill self-discipline within program participants, which also leads to reduced recidivism (Wilson and MacKenzie, 2006).

Narcotic maintenance programs (i.e., methadone and levo-alpha-acetyl-methadol maintenance [LAAM]) are very different than other types of incarceration-based drug treatment programs. These programs attempt to reduce the harms associated with heroin dependency (e.g., disease transmission, criminal activity) by prescribing synthetic opioid medication. Unlike heroin, 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. Some long term narcotic treatments gradually reduce the amount of medication administered to the client until the opiate dependence is relieved; other programs maintain clients indefinitely.

Each of the above types of drug interventions ostensibly has the potential to reduce drug use and other criminal behaviors. Existing systematic reviews of this body of literature, however, only found strong evidence supporting the effectiveness of TC programs (Pearson and Lipton, 1999). In particular, Pearson and Lipton (1999) systematically reviewed the research assessing the effectiveness of corrections-based drug abuse programs in reducing recidivism. Their systematic review conducted a comprehensive search for quasi-experimental and experimental evaluations of interventions carried out in correctional settings [i.e., "prison, jail, or a similar residential correctional facility" (p. 390)], conducted in any country, and completed between 1968 and 1996, inclusive. Their search revealed 30 studies meeting their eligibility criteria. Pearson and Lipton's synthesis of these 30 studies indicated that TCs were effective in reducing recidivism. Specifically, these authors' analyses found that six of the seven TC studies reviewed produced substantial reductions in recidivism; the overall mean weighted *r* effect size was 0.133 ($p = 0.025$) with positive effect sizes ranging from .13 to .28 [one effect size was negative (-0.16)]. In contrast, the mean effect size was not statistically significant for either boot camp or group counseling programs—indicating that these programs are no more effective than no treatment. Additionally, Pearson and Lipton found that too few studies evaluated other types of interventions to draw strong conclusions about their effectiveness. Overall, however, these authors characterized the

evidence assessing the effectiveness of methadone maintenance, drug education, cognitive-behavioral, and 12-step programs as being promising.

In many regards, this systematic review is an extension of the work by Pearson and Lipton. Like the work of Pearson and Lipton, 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 substantive 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. Given this difference in time frames, our results may differ somewhat from those of Pearson and Lipton's work.

2 Objectives of the Review

The objective of this review was to systematically synthesize the available evidence regarding the effectiveness of incarceration-based drug treatment interventions in reducing drug relapse and recidivism. More specifically, this systematic review focused on addressing the following research questions: Are incarceration-based drug treatment programs effective in reducing recidivism and drug use? Approximately how effective are these programs (i.e., what's the magnitude of the effect)? Are there particular types of drug treatment programs that are especially effective or ineffective? What characteristics differentiate effective programs from ineffective programs? These questions are addressed using quantitative meta-analytic synthesis techniques.

3 Methods

3.1 Criteria for inclusion and exclusion of studies in the review

The scope of this review was 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 study evaluated an intervention which was administered in a correctional facility (i.e., prison or jail); (2) the intervention specifically targeted substance users; (3) the evaluation used an experimental or two-group quasi-experimental research design which included a no-treatment or minimal treatment comparison group; (4) the study reported an outcome measure involving post-release criminal behavior (this concept includes drug use); (5) the intervention was conducted between 1980 and 2004, inclusive; and, (6) the study had to report enough information to calculate an effect size. Note that eligible studies could be published or unpublished.

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 noteworthy studies. Specifically, programs designed to be alternatives to incarceration such as those reported in Dynia and Sung (2001) and Knight and Hiller (1997) were excluded by this criterion.

The second criterion restricted the focus of this review to studies 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 (1990) 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. Similarly, Jones, Oslon, Karr, and Urbas (2003 and other years) annually report on the effectiveness of a correctional boot camp in Illinois that does not appear to specifically target drug users; thus, this study was also excluded. 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 interventions comprised of drug users (and given the large proportion of incarcerated offenders who are drug users, such a review runs the risk of becoming a review of nearly all incarceration-based interventions).

The third criterion specified that all included evaluations must have a comparison/control group that received no treatment or minimal treatment. Therefore, 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 Swartz, Lurigio, and Slomka (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 Field (1985, 1989) and Berggen and Svard (1990), all of which used program drop-outs as the comparison group, were excluded from this systematic review. Evaluations that utilized program drop-outs as the comparison group were excluded, because extant research clearly demonstrates that drug treatment drop-outs and program completers often differ on important observed variables (and most likely on important unobserved variables as well) prior to the intervention (see e.g., Hiller, Knight, and Simpson, 1999); and thus, selection bias is particularly problematic in this research design.

The fourth and fifth criteria are largely self-explanatory. It is important to emphasize, however, all studies needed to report a post-release measure of recidivism. This criterion excluded a few notable studies, such as Shewan, Macpherson, Reid, and Davies (1996) and Dolan, Shearer, MacDonald, Mattick, Hall, and Wodak (2003), which reported in-prison outcomes.

The last criterion excluded studies that did not report enough information to calculate to an effect size. This criterion was necessary for practical purposes. Unfortunately, several otherwise eligible studies (e.g., Schippers, Van Den Hurk, Breteler, and Meerkerk, 1998; Guerin, 2002) were ruled ineligible based on this last criterion.

3.2 Search strategy for identification of relevant studies

The goal of the search strategy was to identify all studies, published or unpublished, meeting the above eligibility criteria. In order to achieve this objective, a multi-pronged search strategy was utilized. The search began by 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), RSAT, drunk driver, drink driver, DUI, DWI, inmates, incarceration, incarcerated, prison, evaluation, outcome evaluation, and recidivism. These keywords were used in various combinations. See Appendix 1 for a more detailed discussion of the specific combinations.

We also searched for eligible studies by carefully reading existing studies and literature reviews for unfamiliar studies. In particular, we reviewed the reference lists of existing syntheses to identify eligible studies. 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, Arsovksa, Giaime, Goril, Kahl, Król, and Moore, 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 (http://www.csc-scc.gc.ca/text/research_e.shtml); the Home Office (http://www.homeoffice.gov.uk/); RAND Drug Policy Research Center (http://www.rand.org/multi/dprc/); The Urban Institute's crime /justice research page (http://www.urban.org/justice/index.cfm); and, Vera Institute of Justice publications page (http://www.vera.org/publications/publications.asp).

We also hand searched the titles/abstracts of articles published between 1999 to 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.

Finally, our search strategy and its results were reviewed by an information specialist. The information specialist supplied a list of additional studies that appeared relevant.

All studies that appeared to be eligible based on a preliminary review of the title and abstract were retrieved and closely scrutinized to determine final eligibility status. Specifically, we reviewed the title and abstract of each search result for strong evidence of ineligibility. That is, we reviewed each title/abstract looking for clear evidence that the study violated one or more of the eligibility criteria (see section 3.1). For example, the first eligibility criterion is that all studies evaluated an intervention administered in a correctional facility; thus, if a title/abstract clearly indicated that the evaluation assessed a program that was not administered in a correctional facility, then the study was not retrieved for further scrutiny. Those studies that could not be ruled as ineligible based on the title/abstract review were retrieved for further assessment of eligibility. Retrieved studies were read closely to determine final eligibility status.

3.3 Description of methods used in the component studies

The basic research design for eligible studies was a treatment and comparison group design with a post-release outcome measure of interest, such as post-release criminal offending or drug use. Studies varied with respect to the method of constructing the comparison group; common variations were historical comparisons, adjacent jurisdictions, offenders eligible for the treatment program who chose not to participate, eligible offenders who did not participate due to limited space in the drug treatment program, and random assignment. The studies also varied with respect to the degree to which they employed statistical controls (matching, covariate analysis, etc.) to reduce the threat of selection bias. Included studies exhibited variation in the type of recidivism

measure (e.g., arrests, convictions, re-incarceration) and type of drug use measure (e.g., self-report, urinalysis). Our coding forms were designed to capture these methodological variations.

3.4 Criteria for determination of independent studies

Several types of statistical dependencies were evident in evaluations of incarceration-based drug treatment programs. One common dependency was created by multiple measures of criminal behavior (e.g., re-arrest, re-conviction, drug use) or multiple follow-up periods for the same indicator of criminal behavior ups (e.g., 6 months, 12 months). Another common dependency was produced by multiple studies reporting findings from the same sample of research participants.

The statistical methods detailed below required statistical independence of study findings. We utilized several strategies to maintain the statistical independence of study findings. First, all evaluations (i.e., treatment/comparison contrasts) were cross-checked against one another to ensure that multiple studies reporting the results of the same evaluation do not contribute multiple estimates of program effects to any analysis. Second, in evaluations that report multiple measures of criminal behavior, rather than averaging these multiple outcomes, we applied a set of selection criteria that created five data sets of effect sizes, with a particular evaluation contributing only one effect size to each of the data sets. In the first data set preference was given to effect sizes that: (1) were general (i.e., covered all offense types as opposed to being offense specific), (2) were based on arrest, (3) were dichotomous, and (4) followed sample members for 12 months. We preferred effect sizes meeting these criteria, because arrest is more proximate to offending than the other outcome measures, and because such effect sizes were commonly reported outcome measures. And thus effect sizes meeting these criteria provided a comparable measure of program effectiveness across studies. If no such effect size was available we selected the effect size that most closely matched these criteria. For example, property offenses were more general than violent offenses, effect sizes based on re-convictions were preferred over re-incarcerations, and effect sizes following sample members closest to 12 months were preferred over other effect sizes. Each independent evaluation contributed one, and only one, effect size to this "general recidivism" data set. This general recidivism data set served as the main data set in the analyses that follow.

We also created four more specific data sets: one data set for re-arrest, re-conviction, re-incarceration, and drug relapse outcomes. In creating these data sets we had considerably fewer effect sizes to choose from. When multiple effect sizes were available for any of these data sets, preference was given to effect sizes that: (1) were general (i.e., covered all offense types as opposed to being offense specific), (2) were dichotomous, and (3) followed sample members for 12 months. If a study did not report one of these specific types of outcomes, then that study did not contribute to the particular data set. For example, if a study reported only an arrest outcome, then this study would contribute to the arrest data set but not to the re-conviction, re-incarceration, or drug relapse data sets.

3.5 Details of study coding categories

The coding forms employed in this review are provided in Appendix 2. These coding forms were structured hierarchically, in order to explicitly recognize the nested nature of effect sizes within studies. Any number of effect sizes could be coded from each evaluation using these forms [see Lipsey and Wilson (2001) for a discussion of this issue].

The coding forms captured key features of the nature of the treatment, research participants, research methodology, outcome measures, and direction and magnitude of observed effects. Two coders assessed each study. Discrepancies between coders were resolved by one of the principal investigators.

3.6 Statistical procedures and conventions

An effect size was calculated for each evaluation contrast. We utilized the odds-ratio effect size for dichotomous outcomes as this type of effect size is the most appropriate for dichotomous outcome measures (Lipsey and Wilson, 2001). Indicators of criminal behavior based on a continuous scale were coded using the standardized mean difference effect size. These effect sizes were coded in manner such that positive effect sizes indicate the treatment group had a more favorable outcome than the comparison group (i.e., less recidivism or drug use). The odds-ratio effect size (ES_{or}) is defined as:

$$ES_{or} = \frac{P_c/(1-P_c)}{P_t/(1-P_t)}$$

where P_c is the probability of the event (e.g., re-arrest) for the comparison group and P_t is the probability of the same event for the treatment group.[5] The standardized mean difference effect size (ES_d) is defined as:

$$ES_d = \frac{\bar{X}_c - \bar{X}_t}{S_{pooled}}$$

where $\bar{X}_c$ is the comparison group mean, $\bar{X}_t$ is the treatment group mean, and s_{pooled} is the pooled within groups standard deviation, defined as:

$$S_{pooled} = \sqrt{\frac{(n_t-1)s_t^2 + (n_c-1)s_c^2}{(n_t-1)+(n_c-1)}}$$

where s_t^2 is treatment group variance, s_c^2 is the comparison group variance, n_t is the treatment group sample size, and n_c is the comparison group sample size. Odds-ratio effect sizes 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.

Our analyses of these effect sizes utilized the statistical approach outlined by Lipsey and Wilson (2001) and Wang and Bushman (1999). 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 study features) and unmeasured differences between studies. In order to capture unmeasured differences between studies, a random effects component was added to the fixed effects weights calculated for each effect, as follows:

$$v^* = v + v_\theta$$

[5] Note that we used the inverse of the odds-ratio, as we were interested in obtaining values greater than 1 to reflect a lower probability of recidivism in the treatment group relative to the comparison group.

where v is the sampling error variance and v_θ is the random effects variance estimated from the distribution of effect sizes.

Our analyses employed Stata macro programs written by D. B. Wilson.[6] These macro programs calculated the random effects variance component discussed above and computed various statistics such as the overall mean effect and the homogeneity of effects statistic. Further, we also used these macro programs to determine which study features were associated with observed study effects via meta-analytic analogs to analysis of variance and regression, assuming a mixed-effects model estimated via maximum likelihood (Raudenbush, 1994; Overton, 1998). Our publication bias analyses utilized the "metabias" (which performs two tests for publication bias) and the "metatrim" (which conducts a statistical correction for publication bias) Stata macro programs written by Thomas J. Steichen (both of these macros are available via Stata's "net install" command). Finally, we conducted power analyses utilizing the methods described in Hedges and Pigott (2001).

3.7 Treatment of qualitative data

We did not include qualitative research in this systematic review. However, we are open to suggestions from and collaboration with researchers specializing in such techniques in future updates of this synthesis.

4 Findings

4.1 Description of Eligible Studies

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 one study may contain multiple evaluations. In particular, nine unique studies reported the results of multiple evaluations. Seven of these studies reported the results of two evaluations, one study reported the results of three evaluations, and one study (Tunis et al., 1995) reported results from five independent evaluations. The 66 evaluations coded from the 53 unique studies are the unit of analysis for this synthesis.

The overwhelming majority of the included evaluations were conducted in the United States. Fifty-eight of these evaluations were conducted in the United States, three evaluations were conducted in Australia, three other evaluations were conducted in Canada, one evaluation was conducted in the United Kingdom, and one evaluation was 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, a little over half of the evaluations (34) were coded from studies made available after 1999. Interestingly, 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.

[6] As of this writing, David Wilson has made these macro programs available to the public at: http://mason.gmu.edu/~dwilsonb/ma.html

4.2 Overall Mean Effects Across Studies

Sixty-five of the 66 evaluations reported at least one measure of post-release offending (one evaluation reported only drug use outcomes). After applying our effect size selection criteria (see section 3.4 for a description of these criteria), approximately 83% of the 65 general recidivism odds-ratios indicated that the treatment group had less recidivism than the comparison group. Examination of the distribution of general recidivism odds-ratios suggested that one evaluation was an outlier (0.016)[7]; instead of dropping this effect size from the analyses, we decided to Windorize this effect size to the 5th percentile.

Table 1 displays the random effects mean odds-ratio for the general recidivism measure. The mean odds-ratio for this outcome was 1.37 with the 95 percent confidence interval ranging from a lower bound of 1.24 to an upper bound of 1.51, indicating that, on average, participation in these drug treatment programs was associated with a reduction in post-treatment offending. A more intuitive sense of this effect size can be gained by transforming this effect size into a percentage. For heuristic purposes, we assumed a 35% recidivism rate for the comparison group (which was the unweighted average rate of recidivism for all comparison groups), given this assumption the overall mean odds-ratio translates into a recidivism rate of approximately 28% for the treatment group; thus, participation in treatment was associated with a 20% reduction in recidivism (i.e., $\frac{28\% - 35\%}{35\%} \approx -20\%$).[8]

The distribution of the general recidivism measure 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 may moderate the size of the observed treatment effect. Analyses in the subsequent sections tested this possibility.

Table 1 also displays random mean odds-ratios for the four outcome specific data sets (see section 3.4 for a description of these data sets). From this table it is apparent that evaluations utilizing re-convictions as the outcome measure exhibited the largest mean odds-ratio, whereas evaluations utilizing re-incarceration measures of recidivism had the smallest mean odds-ratio. We believe that our general recidivism outcome measure is the best available, as it is the indicator of criminal behavior least likely to be differentially affected by criminal justice system actors based on the study condition of the offender. Furthermore, it is the most comparable indicator of program effectiveness across studies, and, thus, we utilized this measure in the analyses reported in section 4.2 (below).

Interestingly, only 20 of the 66 independent evaluations assessed the effect of drug treatment on post-release drug use (see Figure 3). The random effects mean odds-ratio for these 20 independent evaluations was 1.28 (with a 95% confidence interval of 0.92 to 1.78). Thus, this mean odds-ratio was not statistically significant; this non-significant finding was not due to a lack of statistical power, as our post-hoc power analyses indicated that the power of this analysis to detect a small effect (i.e., a logged odds-ratio of 0.20) was 0.99.[9] Additionally, the distribution of effect sizes exhibited more

[7] This odds-ratio converts into a logged odds-ratio of -4.61; no other log odds-ratio had an absolute value of 2 or more. This odds-ratio came from an evaluation of a narcotics maintenance program (Magura et al., 1993—female sample).

[8] If we assumed a 50% rate of recidivism for the comparison group, then this effect size translates into a 42% recidivism rate for the treatment group, a 16% reduction in recidivism.

[9] All power analyses were conducted using a two-tailed significance level of 0.05. All power analyses utilized observed data (i.e., standard errors, number of studies, etc.); thus, these power analyses were post-hoc.

variability than expected by sampling error alone (Q = 197.97, df = 19, p < 0.001), which again suggested that moderator variables may explain some of the variability in the drug relapse odds-ratios.

4.3 Analysis of Moderator Effects

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 effect size variability. We tried to capture important differences between studies by coding information from each of the included studies; however, our ability to code many relevant study features was limited by the quality of the descriptions provided by the primary authors.

The first moderator variable examined was primary type of intervention. As noted above (see section 1) the coded evaluations involved four types of primary treatment interventions: TCs, counseling programs, boot camps, and narcotic maintenance programs. The majority of the evaluations concerned TCs (30). Another sizeable portion of evaluations (25) assessed counseling programs. Only a handful of evaluations assessed boot camps or narcotic maintenance programs, 2 and 6, respectively. And three evaluations were not described in enough detail to allow categorization.

Table 2 reveals the mean general recidivism odds-ratio varied considerably by type of primary intervention (Q = 6.36, df = 3, p = 0.09). On average, TC and counseling interventions exhibited statistically significant reductions in general recidivism. In particular, evaluations of TC programs had a mean odds-ratio of 1.38 (with a 95% confidence interval of 1.17 to 1.62), which translates into a 28% recidivism rate for participants in these programs, if we continue to assume a 35% recidivism rate for the comparison group. Evaluations of counseling programs had a mean odds-ratio of 1.50 (with a 95% confidence interval of 1.25 to 1.79). This means odds-ratio translates into a 26% recidivism rate for counseling participants, assuming a 35% recidivism rate for the comparison group.

On the other hand, the mean odds-ratios for both boot camps and narcotic maintenance programs were not statistically different from 1 indicating that participation in these programs generally was not associated with statistically significant reductions in recidivism. More specifically, two evaluations of boot camp programs for drug offenders were included in the present research. Both of these evaluations yielded small, positive logged odds-ratios (0.06 and 0.14); neither of these logged odds-ratios were statistically significant. The fixed effects mean odds-ratio for the two boot camp evaluations was 1.10 (with a 95% confidence interval of 0.61 to 1.97). The finding regarding boot camps' lack of effectiveness is not due to a lack of statistical power. In fact, post-hoc power analysis indicated that the power of this analysis to detect a small effect (i.e., a logged odds-ratio of 0.20) was 0.93, which exceeds the standard benchmark of 0.80. Somewhat similarly, five odds-ratios were extracted from evaluations of narcotic maintenance programs. Four of the five logged odds-ratios were negative. The random effects mean odds-ratio for narcotic maintenance evaluations was 0.84 (with a 95% confidence interval of 0.54 to 1.30). The statistical power of this analysis to detect a small effect was very low; thus, the small number of boot camp evaluations did not limit statistical power, but the small number of incarceration-based narcotic maintenance programs did limit statistical power. As a result, little can be said about the potential of incarceration-based narcotic maintenance programs beyond that the limited existing evidence is discouraging.

Preliminary moderator analyses indicated that the association between odds-ratio and several moderator variables depended on whether the odds-ratio came from an

evaluation of a TC or counseling program. Therefore, we conducted separate moderator analyses for TCs and counseling programs, in a series of parallel analyses. The odds-ratios concerning boot camp and narcotic maintenance programs were set aside for these analyses.

It is important to note that the moderator analyses presented in Tables 3 through 11 have limited statistical power. The statistical power to detect a small effect for these moderator analyses ranged from approximately 0.10 to 0.40.[10] The limited statistical power of these analyses means that only contrasts with large effects were likely to be statistically significant; stated differently, many substantively meaningful effects were not statistically significant at conventional levels of significance (i.e., $p < 0.05$), as a result of these analyses low statistical power. To combat the low statistical power of these analyses, we employ two strategies. First, we interpret as statistically significant any contrast that has a probability of occurring by chance alone of less than 10% (i.e., $p < 0.10$). Second, instead of relying solely on statistical significance, in the following moderator analyses we also discussed substantively significant effects; that is, effects that appear to be substantively large, even if not statistically significant, were treated as noteworthy. We defined "substantively significant" relationships based on the 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 less than five evaluations and the difference between categories was 0.40 or more, then we considered such differences as substantively significant.

Another limitation of the moderator analyses presented in Tables 3 through 11 is that all of these analyses were bivariate. Unfortunately, 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., deleting one observation). As a result, these bivariate findings are vulnerable to spuriousness; consequently, the results of the moderator analyses should be viewed as suggestive.

Tables 3, 4, and 5 display the results of series of bivariate moderator analyses using odds-ratios from TCs only. Table 3 examines variation in the general recidivism odds-ratios by coded methodological features. 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., 2002). The lowest level of method quality was weak quasi-experimental designs; these studies utilized a comparison that lacks 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 comparable on important observed variables (e.g., age, gender, prior criminal history, prior drug use), or evaluations that employed slightly different treatment and comparison groups but also used multivariate analyses that controlled for pre-existing differences on important variables. The highest level of

[10] If we increase the significance level to 0.10, the statistical power to detect a small effect for these moderator analyses ranges from approximately 0.20 to 0.50. This level of statistical power is still well below the standard of 0.80. And thus, statistical power is still limited even if the 0.10 level of significance is utilized.

method quality, "experimental designs," randomly assigned research participants to conditions and did not have attrition problems (see the coding forms in Appendix 2).[11]

Based on this variable, most of the included evaluations were methodologically weak. Of the 30 TC evaluations, 11 (37%) were rated as rigorous quasi-experiments or experimental designs. The modal method ranking was "standard quasi-experiment"; 13 of the 30 evaluations earned this ranking. It is important to note that the mean odds-ratios for the three highest levels of method quality were statistically significant—indicating that the effectiveness of TCs was not confined to only methodologically weak evaluations.

The mean odds-ratio for each level of method quality exhibited a weak positive trend. That is, evaluations with the lowest method quality rating had the smallest odds-ratios and evaluations with the highest method quality had higher odds-ratios. The meta-analytic analog to analysis of variance indicated that the variation between levels of quality method was statistically significant ($p = 0.099$). This finding suggests that more methodologically rigorous evaluations found stronger evidence of treatment effectiveness.

Table 3 also indicates that few of the coded methodological features were associated with treatment effectiveness among the evaluations of TC programs. Methodological factors such as random assignment, subject-level matching, and the use of multivariate analyses to control for pre-existing differences between treatment and comparison groups all were not associated with effect size. Additionally, Table 3 reveals that nearly all of the mean odds-ratios were statistically significant, which suggests that the finding of the effectiveness of TCs was robust to methodological variations.

The moderator analysis found that published studies exhibited statistically larger effect sizes than 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 statistical tests for publication bias. Specifically, we conducted both the Begg and Mazumdar (1994), and the Egger, Smith, Schneider, and Minder (1997) tests for publication bias. The more statistically powerful Egger et al. method found evidence of publication bias; that is, the test of the null hypothesis that the intercept for the regression of the standardized effect estimates against their precision equals zero was rejected ($p = 0.001$). Given this finding, we conducted Duval and Tweedie's (1997) "trim and fill" method for accounting for publication bias. This procedure added seven effect sizes to the distribution, which in turn lowered the mean random effects odds-ratio to 1.21 with a 95% confidence interval of 1.08 to 1.36 ($Q = 252.90$, $df = 36$, $p < 0.001$); this publication bias adjusted mean odds-ratio translates into a recidivism rate of approximately 31% for the treatment group, if we continue to assume a 35% recidivism rate for the comparison group.

Table 4 presents the results of a similar bivariate analysis between the general recidivism odds-ratios and sample characteristics. Four sample characteristics were consistently reported by evaluators: age group (juvenile or adult), gender composition of sample, racial composition of sample, and type of offenders (violent/non-violent offenders). Our analyses found that none of the sample characteristics displayed a statistically or substantively significant relationship with effect size. Once again, however, it is important to note that nearly all of the mean odds-ratios in Table 4 were statistically significant and at least modestly large, which suggests that TC programs were effective with many different types of samples.

---

[11] 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.

Characteristics of each intervention were also coded. Bivariate analyses analyzing these characteristics as moderator variables are shown in Table 5. Six treatment characteristics were coded: mandatory aftercare, location of intervention (i.e., jail vs. prison), length of treatment, program maturity, nature of participation (i.e., strictly voluntary vs. at least some non-voluntary participation), and program capacity/average number of participants (not shown in Table 5). Once again, none of the coded characteristics were statistically or substantively related with effect size, and evidence of the effectiveness of TCs was largely robust to coded variation in treatment features. Specifically, although evaluations of TC programs that mandated aftercare treatment, or programs that required all participants to volunteer for treatment produced somewhat larger effect sizes than other programs, these features were not related to magnitude of effect size. Similarly, programs with short treatment durations were somewhat less effective than longer programs, but this difference also was not significant. In fact, these evaluations found evidence of the effectiveness of TC programs regardless of these coded differences in treatment characteristics. Thus, once more the evidence suggests that participants in TC programs had lower recidivism rates than non-participants, regardless of several prominent evaluation characteristics.

A parallel set of analyses were conducted for counseling programs (see Tables 6, 7, and 8). Again, most evaluations were methodologically weak. Over two-thirds of evaluations (72%) were rated as either “weak” or “standard” quasi-experiments. Only two evaluations employed an experimental design that randomly assigned offenders to treatment conditions. This lack of methodological rigor is particularly problematic as evaluations rated higher on this scale exhibited smaller non-statistically significant mean odds-ratios than evaluations rated lower on the scale. In particular, evaluations rated as “rigorous” quasi-experiments or “experimental designs” exhibited mean effect sizes of 1.33 (with a 95% confidence interval of 0.86 to 2.06) and 1.09 (with a 95% confidence interval of 0.52 to 2.30), respectively; neither of which were statistically significant. While the statistical test comparing the mean odds-ratios for the various levels of methodological rigor was not statistically significant, this finding suggests that the strongest evidence of the effectiveness of counseling programs in reducing re-offending came from methodologically weak evaluations.

The only methodological variables that had a statistically significant association with magnitude of odds-ratio were differential attrition and multivariate data analysis (see Table 6). In particular, evaluations that employed multivariate data analysis yielded larger effect sizes than evaluations that did not utilize such techniques. And evaluations in which differential attrition was apparent had statistically smaller odds-ratios than evaluations without substantial differential attrition. Further, while not statistically significant, evaluations without considerable overall attrition had substantively larger mean effect sizes than other evaluations. It is also worth noting that the moderator distinguishing published from unpublished studies found no difference in the magnitude of the effect for counseling programs (see Table 6). This finding comports with other tests of publication bias (not reported in the tables); that is, both the Begg and Mazumdar, and Egger et al. tests for publication bias retained their null hypotheses (no publication bias).

Two of the four moderator variables capturing sample differences were statistically related to effect size (Table 7). Evaluations with adult samples had a statistically larger mean odds-ratio than evaluations using juvenile samples. Likewise, evaluations that employed female samples exhibited statistically larger mean odds-ratio than either male samples, or mixed gender samples; in fact, post hoc contrasts indicated that all three mean odds-ratios statistically differed from one another. Racial composition of sample had no substantive or statistical relationship to effect size. In fact, counseling programs were effective in reducing re-offending in all of the racial categories.

In regards to treatment characteristics, mature counseling programs and voluntary programs exhibited statistically larger effect sizes than other evaluations (see Table 8). None of the other coded treatment characteristics had a substantive or statistical relationship with effect size. In concordance to the analyses of TC evaluations, programs with a mandatory aftercare component had a larger mean effect size than programs without aftercare, but the difference was not significant. Once again, it is important to note that the moderator variable analysis had low statistical power, so this finding (and other non-significant findings) may be due to a lack of power.

Lastly, we examined the moderator variables' ability to predict variation in drug relapse odds-ratios (see Tables 9, 10, and 11). Because of the limited number of effect sizes involved in these analyses (19), we were unable to conduct separate analyses for the different types of primary treatment; therefore, in these analyses all types of primary treatment were analyzed together.

Perhaps the most striking revelation from the moderator analyses of drug relapse outcomes is that few of the mean effect sizes were statistically significant. This finding suggests that, with few exceptions, regardless of variations in methodology, sample, or treatment characteristics incarceration-based drug treatment does not generally reduce post-release drug use. Only evaluations with the following characteristics exhibited odds-ratios statistically greater than 1 (indicating a significant reduction in drug use): use of random assignment or lack of subject-level matching, samples comprised of adults or females, and programs that were voluntary, had mandatory aftercare, or were based on the TC model. Further, only the methodological quality, age group of sample, and presence of aftercare moderator variables were statistically significant; whereas, the subject-level matching, gender composition, type of treatment, and voluntary participation moderator variables were all substantively significant. Last, it is important to note that while published evaluations exhibited somewhat larger effect sizes than unpublished evaluations, statistical tests for publication bias did not indicate the presence of publication bias in these evaluations.

5 Conclusions

Overall, this meta-analytic synthesis of evaluations of incarceration-based drug treatment programs found that such programs are modestly effective in reducing recidivism. Eight-three percent of the general recidivism odds-ratios favored the treatment group over the comparison group. Moreover, the random effects mean odds-ratio was 1.37, which translates into a 28% recidivism rate for the treatment group, if we assume a 35% rate of recidivism for the comparison group. Yet, the effectiveness of treatments programs clearly varied by type of treatment.

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

Similarly, we found limited evidence of the effectiveness of incarceration-based narcotic maintenance programs in reducing either re-offending or drug use. The scant available evidence suggests that narcotic maintenance programs do not reduce re-offending; in fact, program participants had somewhat higher re-offending rates than non-participants in four out of the five available evaluations. By contrast, all existing evaluations found somewhat lower rates of post-release drug use among participants than non-participants. Thus, incarceration-based narcotic maintenance programs may reduce drug use, but not re-offending. The limited statistical power of the analyses,

however, undermines our ability to draw firm conclusions in this area of research. 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 use. 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 (e.g., female only samples, male only samples, and adult samples), which suggests that TCs can be applied to wide-range of offenders. While TCs in general were clearly effective, TCs that combined incarceration-based treatment with mandatory post-release aftercare exhibited enhanced effectiveness in reducing re-offending. However, the possibility of publication bias in the available body of TC evaluations tempers our findings. That is, there was evidence of publication bias in this area of research that apparently over-estimated the effectiveness of TC programs.

The evidence regarding counseling programs indicated that these programs were effective in reducing re-offending but not drug use. Counseling programs appeared to be most effective in reducing re-offending when targeted towards adult or female offenders. The evidence also indicated that counseling programs that were strictly voluntary appeared to be more effective in reducing re-offending than other counseling programs. However, the strongest evidence of the effectiveness of counseling programs came from evaluations that were methodological weak. Further, while only a few evaluations of counseling programs assessed their effects on drug use, these existing studies did not generally find that participation in counseling programs reduced drug use.

Interestingly, all of the moderator analyses indicated that treatment programs that mandated aftercare after release from incarceration produced larger effect sizes than programs that do not. In all but the analyses of drug use effect sizes, these differences were not significant. However, given these analyses lack of statistical power, sensitivity to the deletion or inclusion of a single evaluation, and the existing evidence that finds aftercare to strengthen the effectiveness of such interventions, we believe the inclusion of a mandatory aftercare component most likely does intensify the effectiveness of incarceration-based drug treatment programs.

The implications of this research for policy-makers are clear. Policymakers seeking effective interventions for incarcerated substance abusers are most likely to find success with programs that intensively focus on the multiple problems of substance abusers, such as TC programs. Policymakers should expect smaller treatment benefits from less intensive treatment programs. Further, based on the existing literature 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 such programs to reduce 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 leaves findings vulnerable to alternative explanations (i.e., reductions in recidivism could be due to factors other than the intervention). Beneficial future research should address these issues.

6 Plans for Updating the Review

We plan to update this systematic review every three years in accordance with Campbell Collaboration guidelines.

7 Acknowledgements

We would like to thank the Nordic Campbell Center and the Jerry Lee Foundation for partial support of this project. We would also like to thank Matthew Makarios for his assistance in coding many of the included evaluations.

8 Statement Concerning Conflict of Interest

None of the authors have any financial interests in any existing or planned incarceration-based drug treatment or any competing types of interventions for drug using offenders.

9 References

Begg, Colin B. and Madhuchhanda Mazumdar. 1994. "Operating Characteristics of a Rank Correlation Test for Publication Bias." *Biometrics* 50(4):1088-101.

Bennett, Trevor. 1998. *Drugs and Crime: The Results of Research on Drug Testing and Interviewing Arrestees*. London: Home Office.

Berggren, O. and H. Svard. 1990. *Osteraker Project: A Further Follow-Up of the Drug Misuser Treatment Programme at Osterake Prison*. Sweden: Swedish Prison and Probation Administration.

Bureau of Justice Statistics. 1995. *Drugs and Crime Facts, 1994: A Summary of Drug Data Published in 1994*. Rockville, MD: U.S. Department of Justice, Bureau of Justice Statistics.

Dolan, Kate A., James Shearer, Margaret MacDonald, Richard P. Mattick, Wayne Hall, and Alex D. Wodak. 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.

Duvall, Sue and Richard Tweedie. 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, Paul and Hung-En Sung. 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, Matthias, George D. Smith, Martin Schneider, and Christoph Minder. 1997. "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* 315:629-34.

Farrington, David P., Denise C. Gottfredson, Lawrence W. Sherman, and Brandon C. Welsh. 2002. "The Maryland Scientific Methods Scale." Pp. 13-21 in *Evidence-Based Crime Prevention*, Editors Lawrence W. Sherman, David P. Farrington, Brandon C. Welsh, and Doris L. MacKenzie. New York, NY: Routledge.

Field, Gary. 1985. "The Cornerstone Program: A Client Outcome Study." *Federal Probation* 49(2):50-55.

———. 1989. *A Study of the Effects of Intensive Treatment on Reducing the Criminal Recidivism of Addicted Offenders*. Unpublished Manuscript: Oregon Correctional Treatment Programs.

Flanagan, Clare, Jana Arsovska, Alison J. Giaime, Mara Goril, Christian Kahl, Artur Król, and Bianca Moore. 2004. *Digest of Research on Drug Use and HIV/AIDS in Prisons*. London: The European Network for Drug Services in Prison, Cranstoun Drug Services.

Gerstein, Dean R. and Hernick J. Harwood, Editors. 1990. *Treating Drug Problems: A Study of the Evolution, Effectiveness, and Financing of Public and Private Drug Treatment Systems.* Washington, D.C.: National Academy Press.

Guerin, Paul. 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.

Hasselblad, Vic and Larry V. Hedges. 1995. "Meta-Analysis of Screening and Diagnostic Tests." *Psychological Bulletin* 117(1):167-78.

Hedges, Larry V. and Therese D. Pigott. 2001. "The Power of Statistical Tests in Meta-Analysis." *Psychological Methods* 6(3):203-17.

Hiller, Matthew L., Kevin Knight, and D. D. Simpson. 1999. "Risk Factors That Predict Dropout from Corrections-Based Treatment for Drug Abuse." *The Prison Journal* 79(4):411-30.

Jones, Robert J., Bruce W. Olson, Steven P. Karr , and Sheila M. Urbas. 2003. *Impact Incarceration Program: 2003 Annual Report to the Governor and General Assembly*. Springfield, IL: Illinois Department of Corrections.

Knight, Kevin and Matthew L. Hiller. 1997. "Community-Based Substance Abuse Treatment: A 1-Year Outcome Evaluation of the Dallas County Judicial Treatment Center." *Federal Probation* 61(2):61-68.

Lipsey, Mark W. and David B. Wilson. 2001. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage Publications.

Lo, Celia C. and Richard C. Stephens. 2000. "Drugs and Prisoners: Treatment Needs on Entering Prison." *American Journal of Drug and Alcohol Abuse* 26(2):229-45.

Magura, Stephen, Andrew L. C. Rosenblum, and Herman Joseph. 1993. "The Effectiveness of in-Jail Methadone Maintenance." *Journal of Drug Issues* 23(1):75-99.

Overton, Randall C. 1998. "A Comparison of Fixed-Effects and Mixed (Random-Effects) Models for Meta-Analysis of Moderator Variable Effects." *Psychological Methods* 3(3):354-79.

Pearson, Frank S. and Douglas S. Lipton. 1999. "A Meta-Analytic Review of the Effectiveness of Corrections-Based Treatment for Drug Abuse." *The Prison Journal* 79(4):384-410.

Pernanen, Kai, Marie-Marthe Cousineau, Serge Brochu, and Fu Sun. 2002. *Proportions of Crimes Associated With Alcohol and Other Drugs in Canada*. Ottawa, ON: Canadian Centre on Substance Abuse.

Peters, Roger H., Paul E. Greenbaum, John F. Edens, Chris R. Carter, and Madeline M. Ortiz. 1998. "Prevalence of DSM-IV Substance Abuse and Dependence Disorders Among Prison Inmates." *American Journal of Drug and Alcohol Abuse* 24(4):573-87.

Raudenbush, Stephen W. 1994. "Random Effects Models." Pp. 302-21 in *The Handbook of Research Synthesis*, Editors Harris Cooper and Larry V. Hedges. New York, NY: Russell Sage Foundation.

Schippers, Gerard M., Arie A. van den Hurk, Marinus H. M. Breteler, and Gert-Jan Meerkerk. 1998. "Effectiveness of a Drug-Fee Detention Treatment Program in a Dutch Prison." *Substance Use & Misuse* 33(4):1027-46.

Shaw, James W. and Doris L. MacKenzie. 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):501-16.

Shewan, David, A. Macpherson, Margaret M. Reid, and John B. Davies. 1996. "The Impact of the Edinburgh Prison Drug Reduction Programme." *Legal and Criminological Psychology* 1(1):83-94.

Swartz, James A., Arthur J. Lurigio, and Scott Slomka. 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):553-73.

Tunis, Sandra, James Austin, Mark Morris, Patricia Hardyman, and Melissa Bolyard. 1995. *Evaluation of Drug Treatment in Local Corrections: Final Report*. San Francisco, CA: National Council on Crime and Delinquency.

Wang, Morgan C. and Brad J. Bushman. 1999. *Integrating Results Through Meta-Analytic Review Using SAS Software*. Cary, NC: SAS Institute, Inc.

Wilson, David B. and Doris Layton MacKenzie. 2006. "Boot Camps." Pp. 73-86 in *Preventing Crime: What Works for Children, Offenders, Victims, and Places*, Editors Brandon C. Welsh and David F. Farrington. New York: Springer.

Wilson, David B., Doris L. MacKenzie, and Fawn N. Mitchell. 2005. *Effects of Correctional Boot Camps on Offending*. A Campbell Collaboration systematic review, available at: http://www.aic.gov.au/campbellcj/reviews/published.html.

Zhang, Sheldon. 2000. *An Evaluation of the Los Angeles County Juvenile Drug Treatment Boot Camp: Final Report*. San Marcos, CA: California State University.

10 Studies Included in Systematic Review

Anglin, M. D., Michael L. Prendergast, David Farabee, and Jerome Cartier. 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, Steve. 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.

Crundall, Ian and Kaye Deacon. 1997. "A Prison-Based Alcohol Use Education Program: Evaluation of a Pilot Study." *Substance Use & Misuse* 32(6):767-77.

Daley, Marilyn, Craig T. Love, Donald S. Shepard , Cheryl B. Petersen, Karen L. White, and Frank B. Hall. 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.

Dowden, Craig and Kelley Blanchette. 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, John R. and Ronald S. Everett. 1998. "An Experimental Test of Chemical Dependency Therapy for Jail Inmates." *International Journal of Offender Therapy and Comparative Criminology* 42(4):360-368.

Eisenberg, Michael, Lisa Riechers, and Nancy Arrigona. 2001. *Evaluation of the Performance of the Texas Department of Criminal Justice Rehabilitation Tier Programs*. Austin, TX: Criminal Justice Policy Council.

Eisenberg, Michael and Tony Fabelo. 1996. "Evaluation of the Texas Correctional Substance Abuse Treatment Initiative: The Impact of Policy Research." *Crime & Delinquency* 42(2):296-308.

Finigan, Michael W., Nancy Barron, and Shannon Carey. 2003. "Effectively Assessing and Preparing Inmates for Community Substance Abuse Treatment: The Portland Target Cities Project in-Jail Intervention." Pp. 165-78 in *Clinical Assessment and Substance Abuse Treatment: The Target Cities Experience*, Editors Richard C. Stephens, Christy K. Scott, and Randolph D. Muck. Albany, NY: State University of New York Press.

Gordon, Jill A. 2002. *Barrett Juvenile Correctional Center: Is It Effective*. Richmond, VA: Virginia Commonwealth University.

Gransky, Laura A. and Robert J. Jones. 1995. *Evaluation of the Post-Release Status of Substance Abuse Program Participants*. Chicago, IL: Illinois Criminal Justice Information Authority.

Hanson, Gina. 2000. *Pine Lodge Intensive Inpatient Treatment Program*. Olympia, WA: Planning and Research Section, Washington State Department of Corrections.

Hartmann, David J., James L. Wolk, L. S. Johnston, and Corey J. Colyer. 1997. "Recidivism and Substance Abuse Outcomes in a Prison-Based Therapeutic Community." *Federal Probation* 61(4):18-25.

Hughey, Ray and LLoyd W. Klemke. 1996. "Evaluation of a Jail-Based Substance Abuse Treatment Program." *Federal Probation* 60(4):40-44.

Hume, Sarah and Angela Gorta. 1988. *Results of Community Urinalyses for Clients on the NSW Prison Methadone Program. Study 6*. New South Wales, Australia: new South Wales Department of Corrective Services.

Hume, Sarah and Angela Gorta. 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, James A., Steven S. Martin, Clifford A. Butzin, Robert M. Hooper, and Lana D. Harrison. 1997. "An Effective Model of Prison-Based Treatment for Drug-Involved Offenders." *Journal of Drug Issues* 27(2):261-79.

Johnson, Sara L., Jennifer T. C. van de Ven, and Brian A. Grant. 2001. *Institutional Methadone Maintenance Treatment: Impact on Release Outcome and Institutional Behaviour (Research Rep. No. 119)*. Ottawa: Research Branch, Correctional Service of Canada.

Kelly, William R. 2001. *An Outcome Evaluation of the Texas Youth Commission's Chemical Dependency Treatment Program*. Austin, TX: University of Texas.

Kinlock, Timothy W., Robert J. Battjes, and Robert P. t. M. P. T. Schwartz. 2005. "A Novel Opioid Maintenance Program for Prisoners: Report of Post-Release Outcomes." *American Journal of Drug and Alcohol Abuse* 31:433-54.

Klebe, Kelli J. and Maureen O'Keefe. 1998. *Outcome Evaluation of the Crossroads to Freedom House and Peer I Therapeutic Communities*. Colorado Springs, CO: University of Colorado.

Knight, Kevin, D. D. Simpson, and Matthew Hiller. 1999. "Three-Year Reincarceration Outcomes for in-Prison Therapeutic Community Treatment in Texas." *The Prison Journal* 79(3):337-51.

Kunitz, Stephen J., W. G. Woodall, Hongwei Zhao, Denise R. Wheeler, Robert Lillis, and Everett Rogers. 2002. "Rearrest Rates After Incarceration for DWI: A Comparative Study in a Southwestern US County." *American Journal of Public Health* 92(11):1826-31.

Little, Gregory L. and Kenneth D. Robinson. 1989. "Treating Drunk Drivers With Moral Reconation Therapy: A One-Year Recidivism Report." *Psychological Reports* 64:960-962.

Magura, Stephen, Andrew L. C. Rosenblum, and Herman Joseph. 1993. "The Effectiveness of in-Jail Methadone Maintenance." *Journal of Drug Issues* 23(1):75-99.

Martin Carol, Elaine Player, and Sarah Liriano. 2003. "Results of Evaluations of the RAPt Drug Treatment Programme." *Prisoners' Drug Use and Treatment: Sever Research Studies* , Editor Malcolm Ramsey. London: Home Office Research.

Miller, J. M. and Barbara Koons-Witt. 2003. *Outcome Evaluation of the South Carolina Residential Substance Abuse Treatment Program for State Prisoners*. Columbia, SC: University of South Carolina.

Mosher, Clayton and Dretha Phillips. 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, Jeffrey 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, Jennifer A., Edward J. Latessa, and Melissa Winesburg. 2002. *Final Report: Mohican Youth Center RSAT Outcome Evaluation*. Cincinnati, OH: Center for Criminal Justice Research, University of Cincinnati.

Pelissier, Bernadette, William Rhodes, William Saylor, Gerry Gaes, Scott D. Camp, Suzy D. Vanyur, and Sue Wallace. 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.

Peters, Roger H., William D. Kearns, Mary R. Murrin, Addis S. Dolente, and Robert L. May II. 1993. "Examining the Effectiveness of in-Jail Substance Abuse Treatment." *Journal of Offender Rehabilitation* 19(3/4):1-39.

Porporino, F. J., D. Robinson, B. Millson, and J. R. Weekes. 2002. "An Outcome Evaluation of Prison-Based Treatment Programming for Substance Users." *Substance Use & Misuse* 37(8-10):1047-77.

Porter, Rachel. 2002. *Breaking the Cycle: Technical Report*. New York: Vera Institute of Justice.

Prendergast, Michael, Elizabeth Hall, and Jean Wellisch. 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, Michael L., Jean Wellisch, and Mamie M. Wong. 1996. "Residential Treatment for Women Parolees Following Prison-Based Drug Treatment: Treatment Experiences, Needs and Services, Outcomes." *The Prison Journal* 76(3):253-74.

Sealock, Miriam D., Denise C. Gottfredson, and Catherine A. Gallagher. 1997. "Drug Treatment for Juvenile Offenders: Some Good and Bad News." *Journal of Research in Crime and Delinquency* 34(2):210-236.

Siegal, Harvey A., Jichuan Wang, Russel S. Falck , Ahmmed M. Rahman, and Robert G. Carlson. 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, Cindy J. 1996. "The California Civil Addict Program: An Evaluation of Implementation and Effectiveness." Dissertation, University of California, Irvine, Ann Arbor, MI.

Taxman, Faye S. and David L. Spinner. 1996. *The Jail Addiction Services (JAS) Project in Montgomery County, Maryland*. College Park, MD: University of Maryland.

Tunis, Sandra, James Austin, Mark Morris, Patricia Hardyman, and Melissa Bolyard. 1995. *Evaluation of Drug Treatment in Local Corrections: Final Report*. San Francisco, CA: National Council on Crime and Delinquency.

Turley, Alan, Tim Thornton, Craig Johnson, and Sue Azzolino. 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):721-28.

Van Stelle, Kit R. and D. P. Moberg. 2000. *Outcome Evaluation of the Wisconsin Residential Substance Abuse Treatment Program: The Mental Illness-Chemical Abuse (MICA) Program at Oshkosh Correctional Institution, 1998-2000*. Madison, WI: University of Wisconsin Medical School.

Vaughn, Michael S., Furjen Deng, and Lou-Jou Lee. 2003. "Evaluating a Prison-Based Drug Treatment Program in Taiwan." *Journal of Drug Issues* 33(2):357-84.

Voas, R. B. and A. S. Tippetts. 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.

Welsh, Wayne N. 2002. *Evaluation of Prison Based Drug Treatment in Pennsylvania*. Philadelphia, PA: Temple University.

Wexler, Harry K., Gregory P. Falkin, and Douglas S. Lipton. 1990. "Outcome Evaluation of a Prison Therapeutic Community for Substance Abuse Treatment." *Criminal Justice and Behavior* 17(1):71-92.

Wexler, Harry K., Gerald Melnick, Lois Lowe, and Jean Peters. 1999. "Three-Year Reincarceration Outcomes for Amity in-Prison Therapeutic Community and Aftercare in California." *Prison Journal* 79(3):321-36.

Winesburg, Melissa, Edward J. Latessa, and Jennifer A. Pealer. 2002. *Final Report: Noble Choices RSAT Outcome Evaluation*. Cincinnati, OH: Center for Criminal Justice Research, University of Cincinnati.

Zhang, Sheldon. 2000. *An Evaluation of the Los Angeles County Juvenile Drug Treatment Boot Camp: Final Report*. San Marcos, CA: California State University.

11 Tables

Table 1. Mean random effects odds-ratio by type of recidivism measure

		95% Confidence Interval			
Outcome	**Mean ES**	**Lower**	**Upper**	***Q***	***k***[a]
General recidivism	1.37*	1.24	1.51	551.84*	65
Re-arrests	1.40*	1.25	1.56	196.52*	35
Re-convictions	1.43*	1.27	1.61	23.40	17
Re-incarcerations	1.22*	1.07	1.39	307.59*	35
Drug relapse	1.28	0.92	1.78	197.97*	20

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

Table 2. General Recidivism Odds-Ratio by Treatment Characteristics

		95% Confidence Interval		
Type of Program	**Mean ES**	**Lower**	**Upper**	***k***[a]
Therapeutic Community	1.38*	1.17	1.62	30
Counseling	1.50*	1.25	1.79	25
Narcotic Maintenance	0.84	0.54	1.30	5
Boot Camp	1.10	0.61	1.97	2

[a] Number of odds-ratios
* $p < 0.05$
# $p < 0.10$
Between $Q = 6.36$, $df = 3$, $p = 0.09$

Table 3. General Recidivism Odds-Ratio by Method Variables: TCs Only

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***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
Randomly assigned to conditions				
No	1.32*	1.15	1.52	28
Yes	1.90*	1.16	3.11	2
Used subject-level matching				
No	1.43*	1.23	1.66	24
Yes	1.10	0.81	1.50	6
Used multivariate data analysis				
No	1.27*	1.04	1.55	14
Yes	1.46*	1.20	1.78	16
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 $p < 0.10$.

Table 4. General Recidivism Odds-Ratio by Sample Characteristics: TCs Only

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***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
Racial composition of sample				
50% or less non-white	1.72*	1.24	2.38	7
51%-70% non-white	1.35*	1.03	1.78	10
More than 70% non-white	1.25	0.93	1.68	7
Offender type				
Non-violent offenders	1.49*	1.24	1.79	15
Mixed (violent and non-violent)	1.28*	1.02	1.62	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 $p < 0.10$.

Table 5. General Recidivism Odds-Ratio by Treatment Characteristics: TCs Only

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***k***[a]
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
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
Short treatment (less than 90 days)				
No	1.45*	1.26	1.68	22
Yes	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

[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 $p < 0.10$.

Table 6. General Recidivism Odds-Ratio by Method Characteristics: Counseling

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***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
Randomly assigned to conditions				
No	1.55*	1.25	1.92	23
Yes	1.09	0.51	2.32	2
Used subject-level matching				
No	1.48*	1.15	1.89	19
Yes	1.59#	0.97	2.61	5
Used multivariate data analysis+				
No	1.18	0.85	1.63	9
Yes	1.73*	1.35	2.22	16
Overall attrition apparent				
No	1.63*	1.28	2.07	18
Yes	1.15	0.77	1.72	6
Differential attrition apparent†				
No	1.68*	1.37	2.05	20
Yes	0.71	0.43	1.16	3
Published				
No	1.53*	1.07	2.18	9
Yes	1.49*	1.15	1.94	16

[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 $p < 0.10$.

Table 7. General Recidivism Odds-Ratio by Sample Characteristics: Counseling

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***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
Racial composition of sample				
50% or less non-white	1.48*	1.19	1.85	5
51%-70% non-white	1.44*	1.21	1.71	6
More than 70% non-white	1.50*	1.34	1.68	2
Offender type				
Non-violent offenders	1.48*	1.18	1.86	11
Mixed (violent and non-violent)	1.26[#]	0.98	1.62	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 $p < 0.10$.

Table 8. General Recidivism Odds-Ratio by Treatment Characteristics: Counseling

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***k***[a]
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
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
Short treatment (less than 90 days)				
No	1.44#	0.98	2.12	10
Yes	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

[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 $p < 0.10$.

Table 9. Drug Relapse Odds-Ratio by Method Variables

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***k***[a]
Overall method quality[†]				
Weak quasi-experiment	0.70	0.40	1.23	5
Standard Quasi-experiment	1.46	0.81	2.64	5
Rigorous quasi-experiment	1.30	0.84	1.99	8
Experimental design	4.49*	1.37	14.79	1
Randomly assigned to conditions				
No	1.18	0.83	1.68	18
Yes	2.51[#]	0.91	6.93	2
Used subject-level matching				
No	1.40[#]	0.95	2.05	16
Yes	0.93	0.43	2.00	4
Used multivariate data analysis				
No	1.14	0.68	1.90	9
Yes	1.43	0.90	2.29	11
Overall attrition apparent				
No	1.31	0.78	2.20	9
Yes	1.18	0.72	1.93	10
Differential attrition apparent				
No	1.33	0.84	2.10	11
Yes	1.12	0.64	1.96	8
Published				
No	1.16	0.71	1.88	10
Yes	1.45	0.88	2.39	10

[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 $p < 0.10$.

Table 10. Drug Relapse Odds-Ratio by Sample Characteristics

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***k***[a]
Age group of sample[†]				
Adults	1.58*	1.20	2.08	15
Juveniles	0.79	0.45	1.38	3
Gender composition of sample				
All female	2.06[#]	0.97	4.38	5
Mixed (male and female)	0.81	0.29	2.26	2
All male	1.18	0.77	1.82	12
Racial composition of sample				
50% or less non-white	1.86	0.69	5.03	2
51%-70% non-white	1.37	0.77	2.44	6
More than 70% non-white	1.34	0.76	2.35	6
Offender type				
Non-violent offenders	1.41	0.80	2.49	8
Mixed (violent and non-violent)	0.94	0.57	1.55	8

[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 $p < 0.10$.

Table 11. Drug Relapse Odds-Ratio by Treatment Characteristics

		95% Confidence Interval		
Variable	**Mean ES**	**Lower**	**Upper**	***k***[a]
Type of program				
Therapeutic community	1.41[#]	0.95	2.09	12
Boot camp	0.56	0.15	2.06	1
Counseling	0.78	0.35	1.73	3
Narcotic maintenance	1.95	0.87	4.40	4
Mandatory aftercare[†]				
No	0.75	0.42	1.34	6
Yes	1.79*	1.14	2.81	11
Treatment location				
Prison	1.25	0.87	1.80	18
Jail	1.97	0.52	7.49	2
Program maturity				
New program (less than 1 year)	1.01	0.56	1.82	7
Developing program (1-3 years)	1.69	0.78	3.67	5
Established program (3+ years)	1.26	0.61	2.60	5
Short treatment (less than 90 days)				
No	1.10	0.71	1.69	13
Yes	2.04	0.85	4.93	4
Strictly voluntary treatment				
No	0.93	0.44	1.95	4
Yes	1.58*	1.01	2.48	13

[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 $p < 0.10$.

12 Figures

Figure 1. General Recidivism Odds-Ratio and 95% Confidence Interval: TCs

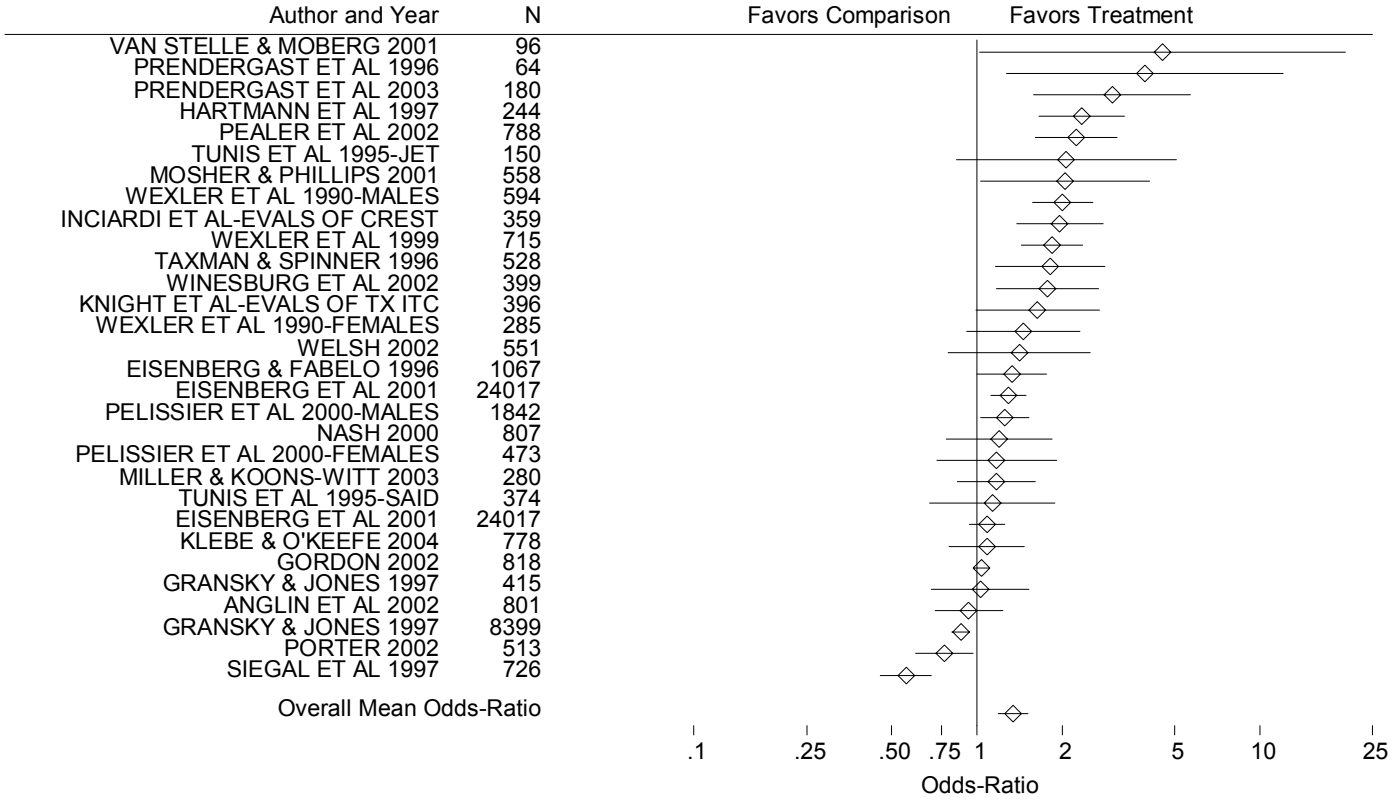


Figure 2. General Recidivism Odds-Ratio and 95% Confidence Interval: Counseling

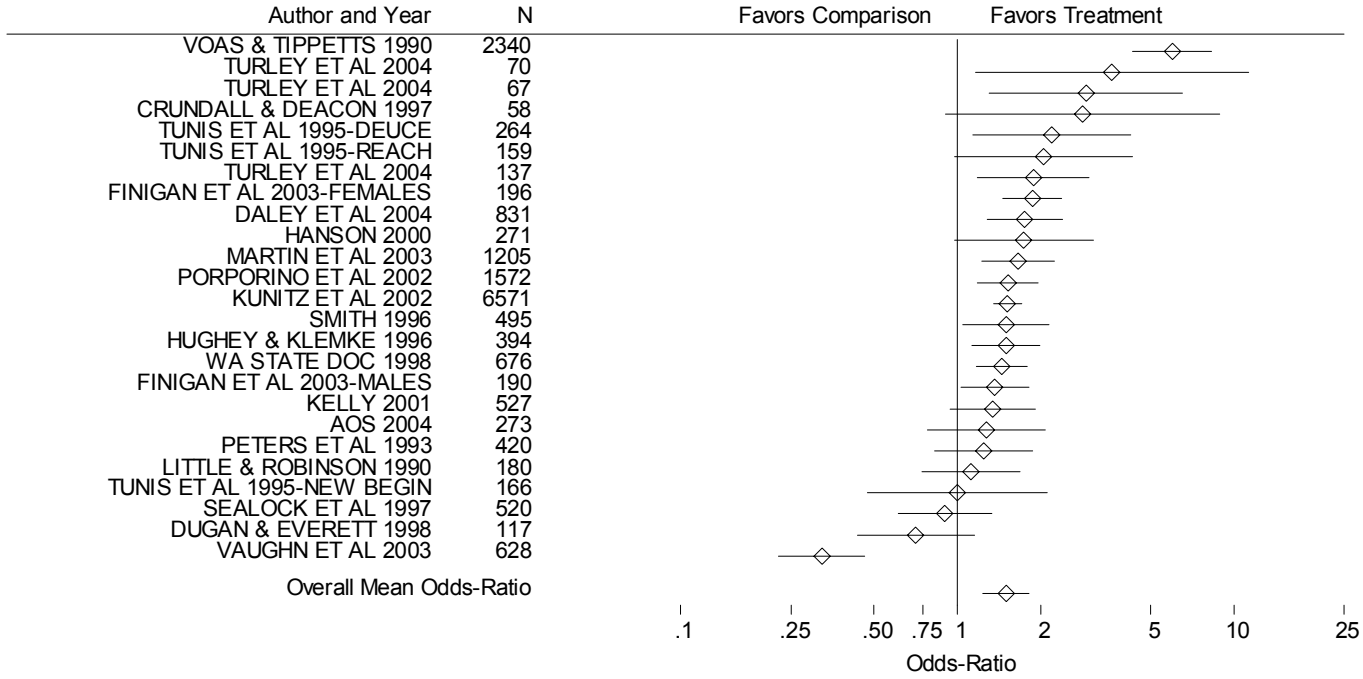
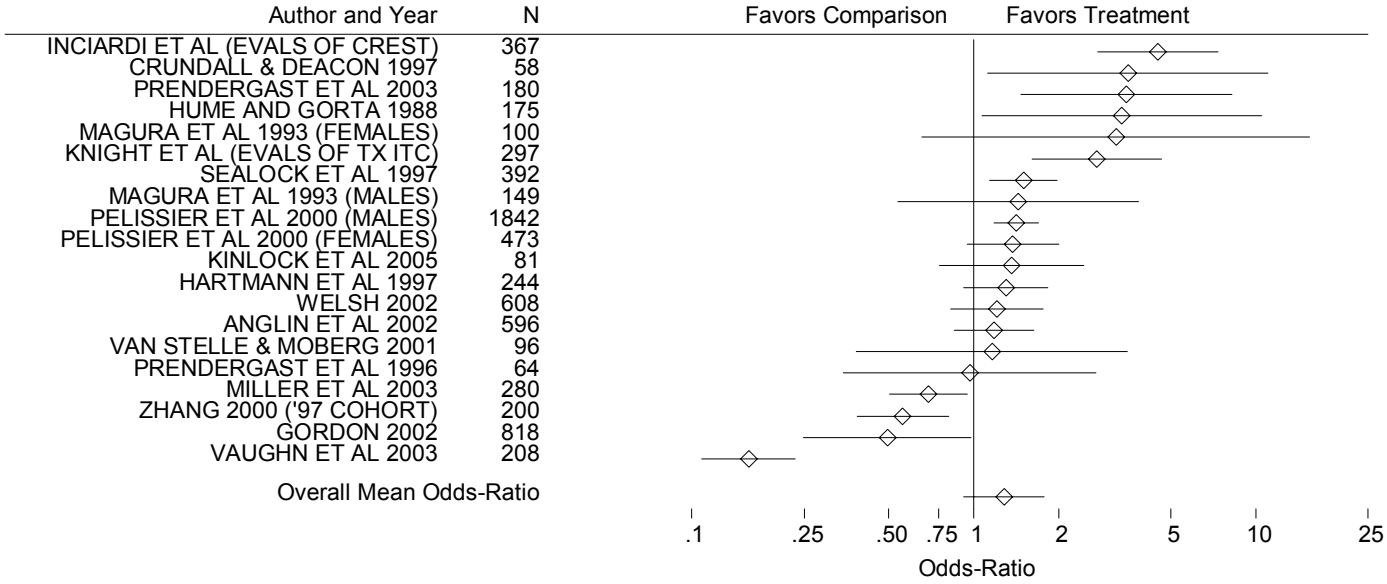


Figure 3. Drug Relapse Odds-Ratio and 95% Confidence Interval



13 Appendix 1: Search Terms

The search strategy for this review began by 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. We split our keywords into three groups: primary terms, secondary terms, and independent terms. The primary terms were too broad to be used independently; thus, we combined each primary term with each secondary term. We augmented this search strategy by utilizing a few independent search terms. The table below lists the primary, secondary, and independent search terms.

Terms used in computerized database search

Primary Terms	Secondary Terms	Independent Terms
Drug treatment	Offenders	Boot camps
Substance abuse treatment	Inmates	Residential substance abuse treatment
Drug counseling	Incarceration	RSAT
Substance abuse counseling	Incarcerated	
Methadone maintenance	Prison	
Therapeutic community(ities)	Evaluation	
Drunk driver	Outcome evaluation	
Drink driver	Recidivism	
DUI	Arrest	
DWI		

For example, the first primary term, “drug treatment,” was combined with each of the secondary terms to eliminate extraneous results. Then the second primary term, “substance term treatment,” was combined with each of the secondary terms. This process was repeated for each primary term. Lastly, we used the independent terms by themselves in each database.

14 Appendix 2: Coding Forms

Crime Prevention Meta-Analysis
Study Level Code Sheet

Identifying Information

Study (document) identifier [StudyID]

If multiple documents were used to code this study, indicate the supplemental study ID numbers
Cross references document identifier [CROSREF1]

Cross references document identifier [CROSREF2]

Cross references document identifier [CROSREF3]

Coder's initials [Coder]

Date coded [Date]

Author: [Author]

Publication type [PubType]

1 Book **4** Gov't Report, State/local

2 Book Chapter **5** Journal (peer reviewed)

3 Gov't Report, Federal **6** Unpublished (tech report, convention paper, dissertation

Year of publication:

Number of different "modules" included in report [MODS]

Is the same control/comparison group used in different modules? **(1 = Yes; 0 = No)** [SAME_CG]

Crime Prevention Meta-Analysis
Treatment-Comparison Contrast Level Code Sheet

A study may report on multiple independent evaluations, such as independent treatment and control group contrasts, or may have a design that includes multiple interventions of interest contrasted with a single control group. Each of these treatment/control contrasts of interest is treated as a separate "module" for coding purposes. Note that the treatment groups across modules must have independent (non-overlapping) subjects. A single control group may be used in more than one module.

Identifying Information

Study (document) identifier [StudyID]

Module identifier [ModID]

Coder's initials [CoderMod]

Program Description

Program description: [ProgDes 1]

(text)

Primary Treatment Type [PrimeTx]

1 Therapeutic Community (TC)
2 Individual Counseling
3 Group Counseling
4 Boot Camp/Shock Incarceration
5 Methadone Maintenance
6 Multiple modes of treatment (specific modality depends on client characteristics)
7 Other

Treatment Components (Check all that apply)

Life skills programs [TxComp1]
Cognitive behavioral programs [TxComp2]
12-step program [TxComp3]
Drug education [TxComp4]
Academic education [TxComp5]
Post treatment aftercare component [TxComp6]
Other [TxComp6]

In what format or social setting is the treatment delivered? [TxFormat]

1 One-on-one (e.g., therapist/client)
2 Group setting (e.g., classroom, group therapy)
3 Family setting (e.g., family therapy)
4 Mixed (i.e., any combination of the above)
9 Cannot tell

Where does the treatment group reside [TxLocale]

1 Jail
2 Prison
3 Halfway House
4 Other CJ institution
11 Mixed
99 Other

Who delivers or provides the treatment? [TxStaff]

1 Mental health professionals
2 CJ Professionals
3 Professional educator
4 Nonprofessional
5 Other
6 Cannot tell

Length of primary intervention in months (weeks/4.3)

a Minimum [TxMon1]
b Maximum [TxMon2]
c Mean [TxMon3]
d Fixed (same for all subjects) [TxMon4]

Length of aftercare or follow-up program component (weeks/4.3) [TxAfterM]

Describe the program for the comparison group if other than no treatment or treatment as usual. [ProgDes 2]

(text)

What happens to the comparison group? [CompGrp]

1 No treatment
2 Wait-list control
3 Placebo control or “strawman” alternative intervention
4 Treatment as usual; management as usual
5 Treatment drop-outs; unsuccessful participation
6 Nonparticipation in program
7 Mixed, any combination of above
8 Non-sex-offender specific mental health treatment (sex-offender studies only)
9 Cannot tell

Where does the comparison group reside [CgLocale]

1 Jail
2 Prison
3 Halfway House
4 Other CJ institution
11 Mixed
99 Other

Methodological Rigor

Use of control variables in statistical analyses to account for initial group differences **(1=Yes; 0 = No)** [CntrlVar]

Use of random assignment to conditions **(1=Yes; 0 = No)** [Random]

Use of subject level matching **(1=Yes; 0 = No)** [Matching]

Measurement of prior criminal involvement; not necessarily arrest **(1=Yes; 0 = No)** [PreTest]

Rating of initial group similarity **(7=highly similar; 1=highly dissimilar)** [SimRate]

Anchors:
- **7** Randomized design large N or small N with matching
- **5** Nonrandomized design with strong evidence of initial equivalence
- **1** Nonrandomized design, comparison group highly likely to be different or known different that are related to future recidivism

Was attrition discussed in the study reported? **(1=Yes; 0 = No)** [Attrit1]

Is there a potential generalizability threat from overall attrition? [Attrit2]

- **0** No
- **1** yes
- **8** N/A, no attrition problem
- **9** cannot tell

Is there a potential threat from differential attrition? [Attrit3]

(same as above)

Did the statistical analysis of outcome effects attempt to control for differential attrition effects? [Attrit4]
(1=Yes; 0=No; 8=NA)

Use of statistical significance testing **(1=Yes; 0 = No)** [SigTest]

Maryland methodology rating **(see Maryland scale)** [MethScor]

2 A comparison group is present but lacks comparability to the treatment group

3 A comparison group is present but differs slightly from the program group

4 A comparison group is present and it is very similar to program group, or a comparison group is present but it differs slightly from the program group, however, the data analysis controls for observed differences, or random assignment with large attrition

5 Random assignment and analysis of comparable program and comparison groups, including controls for attrition

Notes on Methodology

(text)

Crime Prevention Meta Analysis
Sample Level Code Sheet

Since a study may report results separately for distinct samples, a sample is a separate "level" in the coding scheme. For example if a study reports the results separately for

Identifying Information

Study (document) identifier	[StudyID]
Module identifier	[ModID]
Sample identifier (Note: each sample within a study gets a unique number)	[SampID]
Coder Initials	[CoderSmp]

Sample Description

Sample description treatment group (location, level of security, prior history, etc.) [SampDes1]

(Text)

Sample description comparison group (location, level of security, prior history, etc.) [SampDes2]

(Text)

Total number of individuals in treatment group at beginning of study	[TxN]
Total number of individuals in comparison group at beginning of study	[CgN]

Note: Above must equal the total sample size prior to any attrition. If multiple samples per module are being coded, the sum across samples must equal the total sample size prior to any attrition.

Approximate age range of study participants [Age]

1	Adolescent (12 to 18)	**4**	Adolescent and young adult
2	Young Adult (19 to 25)	**5**	Adolescent and adult
3	Adult (18+)	**9**	Unspecified or cannot tell

Young age included in sample (**99 if unknown)**	[YngAge]
Oldest age included in sample (**99 if unknown)**	[OldAge]
Exact proportion of males in sample if reported	[Males]

Approximate gender description of sample [Sex]

1 All males (>90%)

2 More males than females (60% to 90% males)

3 Roughly half males and half females

4 More females than males (60% to 90% females)

5 All females (>90%)

9 Cannot tell

Offender type general categories [SampType]

1 Violent, person crimes

2 Nonviolent, nonperson crimes

3 Mixed

Crime Prevention Meta-Analysis
Outcome (DV) Level Code Sheet

Identifying Information

Study (document) identifier	[StudyID]	
Outcome identifier (each coded outcome within a study gets a unique number)	[OutID]	
Coder Initials	[CoderDV]	

Outcome Information

Outcome label (label used in report) **(text)** [label]

Recidivism construct represented by this measure **(1=Yes; 0 = No)**

a	Arrest	[DV1]	
b	Conviction	[DV2]	
c	Reinstitutionalization / reincarceration	[DV3]	
d	Revocation	[DV4]	
e	Technical supervision violation	[DV5]	
f	Drug use	[DV6]	
g	Other indicator of criminal involvement	[DV7]	

Specific types of offenses included in recidivism measure **(1=Yes; 0 = No)**

a	All offenses	[DVType1]	
b	Drug offenses (including measures of drug use)	[DVType2]	
c	Person offenses, sexual	[DVType3]	
d	Person offenses, nonsexual	[DVType4]	
e	Person offenses, unspecified	[DVType5]	
f	Property offenses	[DVType6]	
g	Technical supervision or status offense	[DVType7]	
h	Other:	[DVType8]	

Type of measurement scale [Scale]

1 Dichotomy
2 Tricotomy
3 4-9 discrete ordinal categories
4 >9 discrete ordinal categories or continuous

Source of data [Source]

1 Self-report
2 Other report (e.g., teacher, parent)
3 Official record (e.g., school, police, probation, court, institution)
4 Other (e.g., urinalysis)
9 Cannot tell

Is this a valid or reasonable measure of recidivism? [Valid]
(1 = questionable; 2 = acceptable)

Crime Prevention Meta-Analysis
Effect Size Level Code Sheet

Identifying Information

Study identifier [StudyID]

Module identifier [ModID]

Sample identifier [SampID]

Outcome identifier [OutID]

Effect size identifier (number each effect size within a study sequentially) [ESID]

Coder's Initials [CoderES]

Effect Size Information

Effect size type [ES_Type]

1 Baseline (pretest; prior to start of intervention)

2 Post-test (first measurement point, post intervention)

3 Follow-up (all subsequent measurement points, post intervention)

Which group does the raw effect (difference) favor (ignoring statistical significance)? [ES_Direc]

1 Treatment group

2 Comparison group

3 Neither (ES equal zero)

9 Cannot tell (ES cannot be used if this option is selected)

Is this difference reported as statistically significant by the investigator? ES_Sig]

0 No **8** Not tested

1 yes **9** Cannot tell

Time frame in months captured by measure (weeks/4.3)

a Minimum [ES_Time1]

b Maximum [ES_Time2]

c Mean [ES_Time3]

d Fixed (same for all subjects) [ES_Time4]

Effect Size Data

Treatment group sample size for this effect size	[ES_TxN]	
Comparison group sample size for this effect size	[ES_CgN]	
Treatment group mean (clearly indicate decimal point)	[ES_TxM]	
Comparison group mean (clearly indicate decimal point)	[ES_CgM]	
Are the above mean adjusted? **(1=Yes; 0 = No)**	[ES_MAdj]	
Treatment group standard deviation (clearly indicate decimal point)	[ES_TxSD]	
Comparison group standard deviation (clearly indicate decimal point)	[ES_CgSD]	
Treatment group standard error (clearly indicate decimal point)	[ES_TxSE]	
Comparison group standard error (clearly indicate decimal point)	[ES_CgSE]	
Treatment group; number successful	[ES_TxNS]	
Comparison group; number successful	[ES_CgNS]	
Treatment group; proportion successful	[ES_TxPS]	
Comparison group; proportion successful	[ES_CgPS]	
Are the above proportion adjusted for initial group nonequivalence? **(1=Yes; 0 = No)**	[ES_PAdj]	
t-value from an independent t-test or square root of F-value from a one-way analysis of variance with one *df* in the numerator (only two groups)	[ES_T]	
Exact probability for a t-value from an independent *t*-test or square root of F-value from a one-way analysis of variance with one *df* in the numerator (only two groups)	[ES_T_P]	

Chi-square value with *df* = 1 (2 by 2 contingency table)	[ES_ChiSQ]
Correlation coefficient (point biserial)	[ES_RPB]
Correlation coefficient (phi)	[ES_RPHI]
Computer Calculated ES	[ES]
Hand Calculated ES	[HAND_ES]
Hand Calculated SE of ES	[HAND_SE]