ELSEVIER

Drug and Alcohol Dependence 67 (2002) 53–72

DRUG and ALCOHOL DEPENDENCE

www.elsevier.com/locate/drugalcdep

The effectiveness of drug abuse treatment: a meta-analysis of comparison group studies☆

Michael L. Prendergast [a,*], Deborah Podus [a], Eunice Chang [b], Darren Urada [a]

[a] *Integrated Substance Abuse Programs, UCLA Drug Abuse Research Center, Neuropsychiatric Institute, University of California, 11050 Santa Monica Blvd., Suite 150, Los Angeles, CA 90025, USA*

[b] *Prescription Solutions, Mail Stop LC07-286, 3515 Harbor Blvd., Costa Mesa, CA 92626, USA*

Received 9 July 2001; received in revised form 15 January 2002; accepted 21 January 2002

Abstract

A meta-analysis was conducted on 78 studies of drug treatment conducted between 1965 and 1996. Each study compared outcomes among clients who received drug treatment with outcomes among clients who received either minimal treatment or no treatment. Five methodological variables were significant predictors of effect size. Larger effect sizes were associated with studies with the following characteristics: smaller numbers of dependent variables, significant differences between groups at admission, low levels of attrition in the treatment group, a passive comparison group (no treatment, minimal treatment) as opposed to an active comparison group (standard treatment), and drug use determined by a drug test. Controlling for these methodological variables, further analyses indicated that drug abuse treatment has both a statistically significant and a clinically meaningful effect in reducing drug use and crime, and that these effects are unlikely to be due to publication bias. For substance abuse outcomes, larger effect sizes tended to be found in studies in which treatment implementation was rated high, the degree of theoretical development of the treatment was rated low, or researcher allegiance to the treatment was rated as favorable. For crime outcomes, only the average age of study participants was a significant predictor of effect size, with treatment reducing crime to a greater degree among studies with samples consisting of younger adults as opposed to older adults. Treatment modality and other variables were not related to effect sizes for either drug use or crime outcomes

Keywords: Drug abuse treatment; Treatment outcomes; Treatment effectiveness; Meta-analysis

1. Introduction

The extent to which treatment for drug abuse is effective has been of concern to policymakers, researchers, practitioners, and the general public, particularly in recent years. Numerous studies conducted over the past three decades have concluded that drug abuse treatment is effective, that is, that treatment produces measurable and significant expected changes in drug use and other behaviors compared with no treatment (or minimal treatment) or compared with pre-treatment status (e.g. Anglin and Hser, 1990; Apsler and Harding, 1991; Berg, 1992; Brown, 1984; Cooper et al., 1983; Crits-Christoph and Siqueland, 1996; Gerstein and Harwood, 1990; Hubbard, 1992; Kleber, 1989; McLellan et al., 1992, 1996; Sisk et al., 1990; Sorensen and Copeland, 2000; see Prendergast et al., 1998 for a bibliography of reviews of drug treatment outcome studies). In addition, recent meta-analyses have provided quantitative support for the effectiveness of specific types of treatment (methadone maintenance, Brewer et al., 1998; Marsch, 1998; contingency management, Griffith et al., 2000; family-couples treatment for drug abuse, Stanton and Shadish, 1997). On the basis of the body of research included in these narrative reviews and meta-analyses, federal agencies and treatment organizations have widely publicized the motto: 'Treatment Works' (e.g. Landry, 1995; National Association of State Alcohol and Drug Abuse

☆ Readers interested in the codebook used for studies included in this meta-analysis can view it on the journal internet home page at http://www.elsevier.com/locate/drugalcdegsuppmat/.

* Corresponding author. Tel.: +1-310-445-0874; fax: +1-310-312-0559.

E-mail address: mlp@ucla.edu (M.L. Prendergast).

03765-8716/02/$ - see front matter
PII: S0376-8716(02)00014-5

Directors, 1990; the motto also appears in materials from the Center for Substance Abuse Treatment).

Despite the conclusions of research reviews, varying degrees of skepticism still exist regarding the effectiveness of drug treatment among the general public, policymakers, third-party payers, researchers, and journalists (e.g. Apsler, 1994; Bennett et al., 1996; Krauthammer, 1997). Such skepticism has assumed added urgency in recent years due to health care reform and managed care, and the resultant emphasis on tightened federal, state, and municipal budgets and cost containment. Within this new climate, the drug treatment field is being asked to document the effectiveness of treatment generally and of specific treatment types or approaches. At the same time, government agencies and providers seek answers from treatment research on how to improve the delivery of services to drug abusing clients. The present study addresses the question of drug treatment effectiveness by using meta-analysis techniques to combine quantitative outcomes across a large set of studies and to examine methodological, client, and program factors that may influence the magnitude of treatment effects using regression analysis. In comparison with previous meta-analyses of drug abuse treatment, this one covers all of the major treatment modalities and includes a larger set of studies. While the sample of studies analyzed is limited to the United States, it is expected that the findings will be of interest to researchers and practitioners in other countries.

Meta-analysis is the term that describes various quantitative and analytic methods used to combine results across research studies (Cooper, 1984, 1989; Cooper and Hedges, 1994; Glass et al., 1981; Hedges and Olkin, 1985; Hunter and Schmidt, 1990; Rosenthal, 1991). The statistical techniques used in meta-analysis allow the combination of effect sizes across studies that permit inferences about the magnitude, direction, and consistency of treatment effects. According to Rosenthal (1995), meta-analysis provides a quantitative summary of a research domain that describes 'the typical strength of the effect of phenomenon, its variability, its statistical significance, and the nature of the moderator variables from which one can predict the relative strength of the effect or phenomenon'. Although most of the objectives of meta-analysis are similar to those of narrative literature reviews, meta-analysis differs from such reviews in its focus on quantification and statistical analysis. In comparison with findings from individual studies, meta-analysis results have two benefits: (1) statistical power is increased due to the combination of data from multiple studies bearing on a particular research question, and (2) generalizability is increased since the findings of the synthesis are based on a diverse set of study samples rather than on a single study sample (Nurius and Yeaton, 1987).

Following an overview of the drug abuse treatment system in the United States, this paper describes the criteria used for the inclusion of studies in this meta-analysis, the procedures used for finding and coding the studies, and the methods employed in calculating and combining the studies' effect sizes. It then reports on the characteristics of the studies included in the meta-analysis, the effect sizes of the studies, and the moderators that may account for the variation in effect sizes among the studies. Finally, the studies' overall treatment effects and the moderators of those effects are discussed and the limitations of this meta-analysis presented.

2. The drug abuse treatment system in the United States

Integration of drug treatment effectiveness studies begins with an understanding of the drug treatment system in the United States, which can be conceptualized in terms of a hierarchical classification consisting of services, programs, and modalities. At the 'micro' level, drug treatment achieves its goals by having clients participate in specific services or behavioral change techniques. Some of these services directly address drug use, such as aversion therapy, drug testing, drug counseling, 12-step groups, and relapse-prevention training, while other services are intended to ameliorate problems associated with drug use or to improve the life situation of the drug user, such as job skills training, social skills training, women's groups, family therapy, and primary medical care. In typical program settings, specific services or techniques are seldom provided alone but are included along with other services or techniques as part of the overall treatment protocol of the program. Some of these services have been extensively studied for their effectiveness, whereas others have received only limited attention.

At the next level are drug treatment programs. A drug treatment program delivers a combination of services and techniques designed to achieve specified treatment goals. It may be more formally defined as an identifiable operational entity that has a designated staff, specific policies and procedures that govern its operations, an allocated budget, and eligibility criteria that are applied to people who request services (National Institute on Drug Abuse, NIDA, 1991). Publicly funded programs may be operated directly by government agencies, but more commonly, they are operated under contract to nonprofit or for-profit providers. In addition to public funding, programs are supported by client fees or by third-party payers.

Treatment programs have been commonly classified into modalities. Treatment modalities are conceptual categories whose defining characteristics have been abstracted from the salient features of individual treatment programs. The four traditional modalities have

been methadone maintenance programs, therapeutic communities, outpatient drug-free (or nonmethadone) programs, and detoxification programs. Although programs can be placed within one of these modalities, even within the same modality there remains a wide range of variation in program characteristics, structure, philosophy, funding source, services offered, and other features.

This four-fold classification was originally based on the characteristics of publicly funded treatment programs in the early 1970s (Cole and James, 1975; Cole and Waterson, 1976). Beginning in the 1980s, however, privately funded inpatient treatment became an increasingly prominent component of the drug treatment system in the United States. These chemical dependence programs were based in hospitals or other residential facilities, had a 28-day inpatient phase, often included some type of outpatient aftercare, and followed the 12-step model of recovery. These largely private programs tended to be expensive, and under managed care have become less common in recent years. The addition of chemical dependence programs to the typology brings the number of modalities to five. Mixed modality programs offering a combination of the above modalities exist, but in small numbers (e.g. De Leon et al., 1997; Sorensen et al., 1987).

3. Procedures

The procedures for conducting this meta-analysis can be broken down into several steps: (1) determination of the eligibility criteria for studies; (2) identification and selection of studies based on the eligibility criteria; (3) systematic coding of the substantive and methodological characteristics of each study; (4) calculation of the effect size and direction of the treatment effect for the dependent variables in each study; and (5) calculation of average effect size across studies and statistical analysis of the relationships between study characteristics and effect size. We discuss each in turn.

3.1. Eligibility criteria

At the start of the project, a set of inclusion and exclusion criteria was developed to determine the eligibility of a study for the meta-analysis. The criteria were modified and refined during the early stage of the project; the final set of criteria is shown in Appendix A. Briefly, eligible studies were outcome evaluations of drug abuse treatment programs or specific techniques for adults in the United States and Canada published or issued from 1965 through 1996, using either single-group pre-posttest designs or treatment-comparison group designs (although the present analysis includes only studies using treatment-comparison group designs)[1]. The rationale for the focus on studies from the United States and Canada was both conceptual and practical. We believed that including studies from countries with diverse political, legal, and cultural traditions would introduce additional diversity into a dataset that would be heterogeneous already. The practical issue was the lack of resources to retrieve and translate studies in languages other than English.

The eligibility criteria were applied at three stages. First, during the literature search, the criteria shaped the selection of terms used to search electronic databases for potential references. Downloaded search results were screened against the criteria, and those that appeared to be eligible on the basis of titles or abstracts were retrieved. The selection criteria also guided the inspection of bibliographies and reference lists. Second, project staff conducted a second screen of the printed copies of all retrieved documents using the eligibility criteria. Third, the coders determined, after reading the study document(s), whether the study met the criteria for eligibility before beginning to code the study. For those studies that coders deemed ineligible, the project director reviewed the documents to confirm (or overrule) the coders' judgment.

3.2. Literature search

We used three main strategies to identify relevant documents that met the eligibility criteria: searches of online bibliographic databases, checking printed sources, and requests to colleagues and organizations. The following bibliographic databases were searched: Current Contents (Social and Behavioral Sciences), Dissertation Abstracts, ETOH (Alcohol and Alcohol Problems Science Database), GPO Monthly Catalog, Magazine and Newspaper Index, MEDLINE, NTIS (National Technical Information Service), PsychINFO, PAIS (Public Affairs Information Service), Sociological Abstracts, and Social Work Abstracts. Three searches were conducted: an initial search and two update searches 12 and 18 months later.

Secondly, we scanned both printed bibliographies that did not have an online counterpart and specialized bibliographies in substance abuse. In addition, as documents were retrieved and catalogued, we examined their reference lists for potentially relevant documents. Other sources of outcome studies were the printed proceedings of conferences and professional meetings.

[1] The term 'comparison group' is used throughout this paper rather than 'control group' since it is more inclusive of the types of groups that were compared with the treatment group in the studies coded. Although some studies did use a 'control group' typical of randomised control trails, most others used matched comparison groups convenience samples, or intact groups.

Finally, we sent letters to researchers, organizations, and agencies in the drug abuse field, asking them to send us references (and copies if available) of documents that were not likely to be identified through printed or online sources. We sent out 138 letters and received 59 documents in response. Requests for documents were also circulated at professional meetings and conferences and through drug-related newsletters.

Seven documents requested through interlibrary loan were not received by the cutoff date for document retrieval. Had these documents been retrieved, they may or may not have been eligible for coding.

3.3. Coding procedures

The codebook developed for this meta-analysis consisted of questions from previous meta-analyses on related subjects and questions developed specifically for drug treatment evaluations. The core set of questions came from the codebook used in Mark Lipsey's meta-analysis of juvenile delinquency interventions (Lipsey, 1992). Other questions were drawn from meta-analyses conducted by Andrews et al. (1990), Pearson and Lipton (1999) and Tobler (1997). Project staff pilot tested the initial version of the codebook. Further revisions occurred during the early stage of coding. Questions were added or deleted, and decisions and clarifications regarding specific questions were recorded in a coding policy manual. The codebook consisted of 266 questions organized into five categories: study context, methodology, participant characteristics, treatment characteristics, and dependent variable characteristics and effect size calculation. A copy of the final version of the codebook is available on the journal home page at http://www.elseviar.com/locate/drugalcdepsuppmat/.

Each study was coded by one of seven coders, all masters- or doctoral-level students. Coders attended a 2-day training session on use of the codebook and effect size calculation; they then coded and discussed three practice studies before beginning work on the project. Coders met with senior project staff to discuss coding questions every 2 weeks for the first several months of the study, and less often thereafter. In addition, all coders and senior staff periodically coded a selected study and reviewed the results. All coded studies were checked for discrepancies by one of two senior investigators before data entry.

It should be noted that we coded studies, not documents. That is, all retrieved documents associated with a given study were grouped under a single study number before being assigned for coding. Coders drew information to complete coding of that study from any of the documents, but one of the documents was selected as the 'key' document in order to complete certain of the questions for that study (e.g. first author, document date) and to be the standard source of data should there be discrepancies among documents (e.g. varying sample sizes in different study documents). The key document for a study was the one that was more recent, more complete, and/or published. Most studies were reported in only one document. In other cases, a single document reported on more than one study.

3.4. Calculating and combining effect sizes

The quantitative findings from separate treatment outcome studies can be expressed in a common metric in order to calculate estimates of the magnitude of treatment effects across studies. This metric, commonly called an effect size, may be taken to mean 'the degree to which the phenomenon is present in the population' or 'the degree to which the null hypothesis is false' (Cohen, 1988). By convention, an outcome for which the treatment group shows more success than the comparison group is indicated by a positive sign, whereas an outcome that favors the comparison group is indicated by a negative sign[2].

The statistical methods used to calculate, combine, and analyze effects sizes were those of Hedges and Olkin (1985), supplemented by procedures described in Cooper and Hedges (1994). The most common type of effect size for treatment evaluation studies in the social sciences is the standardized mean difference, which is computed by subtracting the mean outcome score of the comparison group from that of the treatment group and dividing the difference by the pooled standard deviation:

$$g = \frac{(M_t - M_c)}{\text{S.D.}_{\text{pooled}}}$$

where g is the effect size estimate, M_t is the mean of the treatment group, M_c is the mean of the comparison group, and $\text{S.D.}_{\text{pooled}}$ is the pooled standard deviation of the two groups. While the standard deviation of the comparison group is sometimes used as the denominator in this type of research, where the assumption of equal population variances is reasonable, the most precise estimate of the population variance is obtained by pooling the standard deviations of the treatment and the comparison groups (Hedges and Olkin, 1985). If means and standard deviations are not available, effect sizes may be estimated from the reported value of the t, F, or χ^2 statistic using formulas found in standard meta-

[2] The sign of the effect size for each dependent variable was verified during editing. In addition, data cleaning included comparing the sign of the effect size against a question in the codebook asking whether the raw scores favoured the treatment group or the comparison group. The reliability of this check, however, assumes that the question about raw scores was answered correctly. Errors in the sign of effect sizes would likely favour incorrect negative over incorrect positive effect sizes, leading to an underestimation of the overall average effect size.

analysis texts (see Cooper and Hedges, 1994; Hedges and Olkin, 1985).

Since effect sizes calculated from means and standard deviations in studies with small samples provide overestimates of the population effect size, it has become standard practice, following Hedges and Olkin (1985), to apply a correction to all such effect sizes, regardless of sample size, in order to provide an (approximately) unbiased population effect size estimate *d*:

$$d \cong \left[1 - \left(\frac{3}{(4N_1 + 4N_c - 9)}\right)\right] g$$

where N_t is the sample size of the treatment group and N_c is the sample size of the comparison group.

Effect sizes for results reported as proportions or percentages may be calculated using the arcsin transformation suggested by Cohen (1988):where $p1$ is the proportion of 'success' for the treatment group and $p2$ is the proportion of 'success' for the comparison group. Unlike the correction factor described above for means and standard deviations, there is no correction available in the literature to reduce small-sample bias for effect sizes calculated from proportions or percentages (g is assumed to equal d).

Since some studies have many outcome variables and others have only a few, calculating an overall average effect size from individual variable-level effect sizes would give greater weight to studies with many outcomes variables. Further, effect sizes within a given study are correlated and thus violate the assumption of independence required for multivariate analysis techniques. To address this problem, we calculated separate effect sizes and conducted separate analyses for distinct outcome variables (i.e. drug use and crime). In those studies where a particular outcome was represented by more than one measure (e.g. drug use measured by self-report and by urinalysis), the effect sizes for each of the measures were averaged.

Since many studies report outcome results for multiple measurement points (e.g. during treatment, end of treatment, post-treatment), we often had more than one time-related measure of the same dependent variable in a particular study. In such cases (again to avoid creating dependencies among dependent variables within studies), we selected the effect size calculated at the first post-treatment assessment point or, if all measures were taken during treatment, at the assessment point nearest to the end of the treatment.

Since studies with large samples provide more precise and stable estimates of the population effect size than do studies with small samples, standard meta-analytic practice is to weight each effect size estimate by the inverse of its variance. The variance of d is estimated by

$$V = N_t + N_c/N_t \times N_c + d^2/2(N_t + N_c)$$

$g = \mathrm{arc}\sin(p1)\text{-}\mathrm{arc}\sin(p2)$for effect sizes from means and standard deviations,

$$V = N_t + N_c/N_t \times N_c$$

for effect sizes from proportions, where d is the effect size, N_t is the sample size of the treatment group, and N_c is the sample size of the comparison group. To prevent studies with very large sample sizes from dominating the effect size averages, the sample sizes of large studies were Windsorized at 160 (80 for the treatment and 80 for the comparison group), which falls at approximately the 75% percentile of the sample size distribution for the studies included in the analysis. Each effect size was weighted using the formula:

$$\frac{\sum W_i d_i}{\sum W_i}, \text{ where } W_i = \frac{1}{V_i}$$

Weighting each effect size by the inverse of its variance assumes a fixed effects model in which the combined individual effect sizes estimate a single, or common, population effect size (i.e. the population variance is zero). That is, any differences in the estimated effect sizes among studies are assumed to be due to sampling error alone. For this set of drug treatment outcome studies, however, it is unlikely that this assumption is correct, since variation observed between studies is as likely to influence the mean effect size estimate as variation within studies. A more plausible assumption is that the study-level effect sizes are a sample drawn from a random distribution of population effect sizes, leading to the use of a random effects model to calculate average effect sizes. A random effects variance component, based on an estimate of variability among the population effect sizes, is added to the individual effect size variance. Generally, random effects estimates are more conservative than fixed effects estimates. Both fixed-effects and random-effects weighted means are reported.

4. Results

Examination of effect sizes from the drug treatment outcome studies involved two stages of analysis. The overall results are summarized in terms of descriptive statistics using inverse-weighted techniques for combining effect sizes. The second stage of analysis focuses on examining moderators of effect size using multivariate modeling of client characteristics and program characteristics, with effect sizes adjusted for differences in methodology across studies. Before reporting results from these analyses, we describe the outcome variables and summarize characteristics of the studies examined.

4.1. Outcome variables

A treatment outcome, for purposes of this study, is a behavior that the treatment is expected to change and for which measurement data are reported. Studies may include multiple types of outcomes (e.g. drug use, crime, employment, psychological status), multiple measures of a particular type of outcome (e.g. self-report and urinalysis for drug use), and multiple assessment points (e.g. during treatment, treatment termination, 6-month follow-up; Wells et al., 1988a,b). With a few exceptions (e.g. client satisfaction), the coding protocol for this meta-analysis specified that an effect size was to be calculated for all outcomes in a study for which sufficient quantitative information was available[3]. For the present paper, however, we limit the outcome variables to drug use and criminal behavior, the two outcomes that are most commonly reported in the studies and that historically have been most frequently referenced in policy discussions of drug abuse treatment.

Studies measured drug use either by drug testing or self-report, or both. Both ways of measuring drug use have their limitations. Urinalysis, the most commonly used biological assay method for illicit drugs, is able to determine drug use over the past 2–3 days for most drugs. Urinalysis results allow one to report that a certain percentage of participants had not been using drugs in the immediate past; it is likely, however, that at least some of these 'abstinent' participants did use drugs over the total assessment period. The reliability of self-reports of the use of drugs over a defined time period since leaving treatment (for example, the past 6 months) is open to question. While underreporting is common (Harrison, 1995; Messina et al., 2000), there is some evidence (Farabee and Fredlund, 1996) that people who have participated in treatment may be more likely to report that they have used drugs than are those who have not been in treatment, which, if true generally, would tend to reduce differences in drug use outcomes between the treatment and the comparison group. Criminal behavior was assessed from self-report or, less often, from official records. Self-reported criminal behavior includes crimes on the street, many of which are not detected by the criminal justice system, while official records include data on criminal justice processing only (arrest, conviction, incarceration).

Table 1
Characteristics of outcome studies of drug treatment programs ($N = 78$)

	N	%
Type of comparison conditions		
Passive comparison		
No treatment	6	7.7
Delayed treatment/wait list	3	3.8
Minimal contact	7	9.0
Active comparison		
Routine treatment	40	51.3
Placebo treatment	11	14.1
Alternative treatment	11	14.1
Assignment procedure		
Random or quasi-random	46	59.0
Nonrandom; matching	5	6.4
Nonrandom; no matching	27	34.6
Publication type		
Journal article	52	66.7
Book or book chapter	1	1.3
Technical report	7	9.0
Dissertation	3	3.8
Unpublished paper	8	10.3
Abstract	3	3.8
Other	3	3.8
Missing	1	1.3
Program modality		
Specific technique	50	64.1
Methadone maintenance	8	10.3
Therapeutic community	8	10.3
Outpatient drug free	8	10.3
Detoxification	2	2.6
Other	2	2.6
When conducted		
1960s	3	3.8
1970s	14	17.9
1980s	14	17.9
1990s	20	25.6
Missing	27	34.6
Primary funding source		
Federal	53	67.9
Other	9	11.5
Missing	16	20.5
Number of subjects[a]		
Mean (S.D.)	156.3 (300.1)	
Min/max	6/2,544	
Median	81	

[a] *N* for treatment and comparison groups combined.

4.2. Study characteristics

The final set of studies included in this paper consists of 78 treatment-comparison group studies of treatment programs or techniques that included at least one drug-use or crime variable and that contained sufficient data to calculate effect sizes and weights[4]. (As noted below, the imputation of missing sample sizes was needed to compute variances in two studies used in the regression analysis.) All of these contained at least one drug-use

[3] Effect sizes were only calculated for variables in which the unit of measure was based on the person. A measure like '% positive urine screens' was not coded if it was clear from the report that different people contributed different numbers of urine specimens.

[4] Because methadone treatment is directed specifically against heroin and other opiates, the drug outcome variable for methadone programs was confined to measures of opiate use. Thus, six studies of methadone maintenance programs that did not have an opiate-related outcome variable were excluded.

outcome variable; 25 of them also contained at least one crime outcome variable.

Table 1 displays the characteristics of the studies included in the current analysis. Nearly 80% of the studies used an active comparison group (routine treatment, placebo treatment, or alternative treatment), with the remaining using a passive comparison group (no treatment, delayed treatment, minimal contact). A majority (59%) of the studies assigned participants randomly or quasi-randomly. Two thirds of the studies were reported in journal articles. With regard to program type, nearly two thirds (64%) of the studies assessed some type of treatment technique, 10% assessed methadone maintenance treatment, therapeutic community treatment, or outpatient drug free programs, and about 3% assessed detoxification programs or other types of programs. It is noteworthy that the traditional treatment modalities are underrepresented in this set of treatment-comparison group studies; evaluations of these types of community-based programs tend to use single-group designs (see Prendergast et al., 2000), whereas studies of techniques are more likely to occur in more controlled settings (e.g. university-based research centers) and to use comparison group designs, although not necessarily with random assignment. The time period of the studies was about equally divided between the 1970s, 1980s, and 1990s. Federal agencies were the primary funding source for over two thirds of the studies. All studies were conducted in the United States. A few studies had very large samples, which resulted in a highly skewed distribution of sample sizes. The mean number of participants was 156.3, the median, 81.

4.3. Descriptive findings

Figure 1 displays a stem-and-leaf plot of the study-level effect sizes for drug use and crime outcomes for each of the studies included within the analysis. The stem identifies the first digit(s) of an effect size and the leaf identifies the final digit of an effect size. Each digit in the leaf represents a single effect size. While the stem-and-leaf plot shows that the effect sizes for drug treatment studies are widely distributed and that some of them are negative (favor the comparison group), the majority of effect sizes are positive (favor the treatment group) for both drug use and crime outcomes.

The summary statistics shown in Table 2 provide a quantitative view of the data. Under a fixed-effects model, the average weighted effect size for drug abuse outcomes is 0.30; for crime outcomes, 0.13. Note that the weighted means are somewhat smaller than the unweighted means, indicating, as has been found previously (Lipsey and Wilson, 1993), that studies with larger samples tend to have smaller effect sizes. Since neither of the confidence intervals include zero, the null hypothesis that the population effect size is zero can be rejected. In short, for drug use and crime outcomes, the weighted average effect size for drug abuse treatment is positive and statistically significant. Not surprisingly, given its primary focus on drug use, treatment has a greater impact on drug outcomes than on crime outcomes.

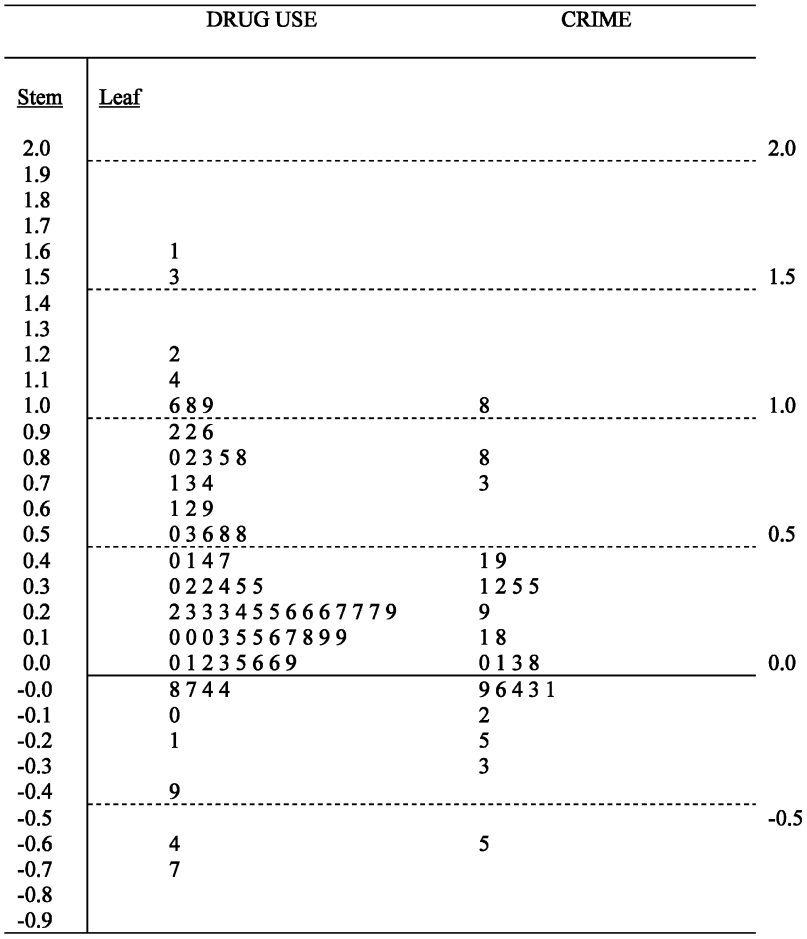
Table 2 also indicates that there is considerable variability in effect sizes. Table 2 shows the results of a homogeneity test (Q) of the effect-size estimates. This test, which has become fairly standard in meta-analysis, is intended to answer the question: Do all of the effect sizes from the sample of studies estimate, within sampling error, the same population effect size for the specified outcome variable (Cooper and Hedges, 1994; Hedges and Olkin, 1985)? As can be seen, the homogeneity test of the means for each of the outcomes is statistically significant. If the Q statistic is significant, as is the case here, then the effect sizes from the studies can be considered more heterogeneous than would be expected from sampling error alone. This suggests that the effect sizes are a sample from a population of randomly distributed effects, in which case a random-effects model for calculating the mean effect size is more appropriate. As seen in Table 2, the random-effects weighted means for the two outcome measures are similar to fixed-effects weighted means: 0.33 for drug use and 0.13 for crime. Also, a significant Q statistic indicates that the sample effect sizes do not share a common population effect size, suggesting that an examination of moderators may account for the variation among effect sizes.

4.4. Moderator analysis

In order to examine moderators that might account for the observed variation among effect sizes for drug use and crime, we used a weighted multiple regression approach appropriate for meta-analysis data (Hedges and Olkin, 1985). Before reporting the results of that analysis, two methodological issues need to be addressed: missing data and methodological differences among studies.

4.4.1. Missing data

Data for many of the variables of interest could not be determined for all of the studies. The most common reason for missing data is that the information for the particular variable is not present in the document(s) reporting on the study. For some variables, the cause of the missing data is the document type; that is, certain types of documents (e.g. conference abstracts) allow for only a limited amount of information to be reported. Other possible reasons for missing data include coding errors, failure to identify data that are reported, and unclear question wording.



Note: Each leaf represents an effect size from an individual study.

Fig. 1. Stem-and-leaf plot of study-level effect sizes for drug use and crime.

Table 2
Summary statistics of study-level effect sizes for drug use and crime outcomes

	Drug use	Crime
Number of studies	78	25
Unweighted mean	0.38	0.16
Standard deviation (S.D.)	0.43	0.38
Maximum	1.61	1.08
Median	0.27	0.08
Minimum	−0.77	−0.65
Fixed effects weighted mean (95% CI)	0.30 (0.25, 0.35)	0.13 (0.04, 0.21)
Homogeneity test (Q)	182.56*	48.68*
Random-effects weighted mean (95% CI)	0.33 (0.25, 0.42)	0.13 (−0.004, 0.27)
Homogeneity test (Q)	81.17	21.66

* $P < 0.05$.

Missing data create problems for analysis. In particular, when listwise deletion of studies with missing values in multiple regression analysis is performed, a considerable reduction in the number of observations included in the analysis will occur. Unless some strategy is applied to impute data for specific cases, the power to detect differences among variables is reduced. We used the EM algorithm technique (Dempster et al., 1997; Little and Rubin, 1987) to impute missing values of the independent variables involved in the regression models discussed below. NORM (Version 2.02) was used to create a complete set of data (Schafer, 1999). Some continuous variables in the imputation model were recoded as ordinal variables because of excessive skewness. In order for the imputation model to converge, variables with a large percentage ($\geq 36\%$) of missing values had to be dropped. Although missing effect sizes were not imputed, the imputation model was used to impute missing sample sizes, which are needed to compute the variances and weights. This resulted in two additional studies being added to the imputed dataset that was used in subsequent analyses.

4.4.2. *Adjusting for methodological differences*

The variation in effect sizes among studies is likely the result of several factors, one of which is a difference in

methodology across studies. The effect size observed in a given study may be partly accounted for by the methodological characteristics of the study rather than by the actual impact of treatment on the outcome measures. We used weighted multiple regression analysis to control for effect size variation due to differences in a variety of methodological variables (see Appendix B).

In the weighted regression analysis, five of the methodological variables were significant predictors of effect size, accounting for 36% of the variance.

- Number of dependent variables coded: studies with a larger number of dependent variables had smaller effect sizes.
- Comparability of treatment and comparison groups: studies in which there were no or negligible differences between the treatment and comparison groups at admission had smaller effect sizes than studies in which significant differences between groups existed at admission.
- Attrition rate of the treatment group: studies with high levels of attrition in the treatment group had smaller effect sizes.
- Type of comparison condition: studies that used a passive comparison group (no treatment, minimal treatment) had larger effect sizes than studies that used an active comparison group (standard treatment).
- Measure of drug use: studies in which drug use was determined by a drug test had larger effect sizes.

The adjusted mean effect sizes for drug use and crime (called method-adjusted effect size, a term adopted from Lipsey and Wilson (1998)) were the expected values from the final model with all significant independent variables centered around the mean of each variable. That is, if the regression model is as follows.

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_i$$

Then, the adjusted effect size is:

$$Y_{iAdj} = Y_i - \beta_1(X_{1i} - \bar{X}_1) - \beta_2(X_{2i} - \bar{X}_2) - \ldots - \beta_k(X_{ki} - \bar{X}_k)$$

For substance abuse, the method-adjusted average effect size was 0.34; for crime, it was 0.16. Note that these method-adjusted effect sizes are only slightly higher than the unadjusted effect sizes for both outcomes shown in Table 2.

4.4.3. Regression analysis

After imputing missing data and adjusting effect sizes by methodological characteristics, we conducted regression analyses in order to determine substantive characteristics of treatment that might account for variation among the method-adjusted effect sizes. The coded variables selected as predictors were those that correlated ($P \leq 0.10$) with effect size in a weighted correlational analysis, plus additional variables that were of theoretical interest. The variables were grouped into three clusters: subject variables, treatment variables, and study context variables. These variables are shown in Appendix B along with information on how they were coded in order to facilitate interpretation of the regression coefficient results. The treatment variable cluster was further broken down into three sub-clusters: amount or intensity of treatment, characteristics of the treatment condition, and treatment context. Thus, there were five variable clusters for this analysis.

Due to the likelihood of high multicolinearity within each cluster, a weighted regression model for the method-adjusted effect sizes was first run separately by each cluster. The statistically significant ($P \leq 0.05$) predictors identified from each model were then entered into a final model in the order of the clusters as listed above. Separate models were run for substance abuse outcomes and crime outcomes.

Results from the cluster models and the final model are shown in Tables 3 and 4. For substance abuse outcomes, the significant predictors of effect size in the final model ($R^2 = 0.34$) were treatment implementation, level of theoretical development, and researcher allegiance. Larger effect sizes tended to be found in studies in which (a) treatment implementation was rated high, (b) the degree of theoretical development of the treatment was rated low, or (c) researcher allegiance to the treatment was rated as favorable.

For crime outcomes, in the final model ($R^2 = 0.34$) only the age variable remained a significant predictor. Effect sizes were smaller in studies in which the average age of the study sample was older. Put differently, studies with samples that included greater numbers of younger adults tended to produce larger effect sizes.

5. Discussion

5.1. Overall treatment effects

In this meta-analysis of the effectiveness of drug abuse treatment based on treatment-comparison studies, the average (weighted) effect size was 0.30 for drug use outcomes and 0.13 for crime outcomes. After adjusting for variations among studies in methodological features, the effect sizes increased somewhat to 0.34 for drug use and 0.16 for crime. The fact that the effect sizes are positive indicates that, on average, clients who participated in treatment had better outcomes than did those who received no treatment or those who received minimal treatment. In addition, the methodologically adjusted mean effect sizes are statistically significant.

Table 3
Weighted multiple regression results for prediction of method-adjusted effect sizes for drug abuse treatment: substance abuse outcomes

Variable cluster	Beta estimate	S.E.	*P*
Subject variables			
Gender	−0.01	0.04	0.82
Race/ethnicity	−0.10	0.09	0.18
Age	0.00	0.01	0.89
Primary drug cocaine	0.12	0.14	0.31
Primary drug heroin	0.12	0.14	0.29
Primary drug polydrug	0.25	0.16	0.06
$R^2 = 0.06$			
$Q = 103.53$, $P = 0.01$			
Treatment integrity			
Assessment of integrity	0.07	0.03	0.01
Delivery of treatment	−0.16	0.04	0.00
$R^2 = 0.21$			
$Q = 86.78$, $P = 0.21$			
Treatment type			
Technique	−0.13	0.09	0.10
Outpatient drug free	−0.05	0.13	0.63
Therapeutic community	−0.19	0.12	0.06
Other	−0.11	0.15	0.37
Structured approach	−0.04	0.09	0.61
$R^2 = 0.05$			
$Q = 104.68$, $P = 0.01$			
Treatment context			
Theoretical development	−0.10	0.02	0.00
Program maturity	−0.12	0.06	0.02
$R^2 = 0.20$			
$Q = 87.85$, $P = 0.06$			
Study context			
Document date	−0.01	0.01	0.03
Document type	0.00	0.07	0.98
Funding source	−0.02	0.08	0.76
Researcher allegiance	0.18	0.07	0.00
$R^2 = 0.12$			
$Q = 96.89$, $P = 0.05$			
Final model			
Assessment of tx integrity	−0.08	0.04	0.07
Delivery of treatment	0.08	0.03	0.01
Theoretical development	−0.07	0.03	0.01
Program maturity	−0.07	0.06	0.19
Researcher allegiance	0.13	0.06	0.05
Document date	0.00	0.00	0.36
$R^2 = 0.34$			
$Q = 72.71$, $P = 0.55$			

Table 4
Weighted multiple regression results for prediction of method-adjusted effect sizes for drug abuse treatment: crime outcomes

Variable cluster	Beta estimate	S.E.	*P*
Subject variables			
Gender	−0.11	0.06	0.01
Race/ethnicity	−0.50	0.28	0.01
Age	0.02	0.02	0.04
Primary drug cocaine	−0.18	0.28	0.36
Primary drug heroin	−0.12	0.28	0.54
$R^2 = 0.28$			
$Q = 38.90$, $P < 0.01$			
Treatment integrity			
Assessment of integrity	−0.02	0.08	0.67
Delivery of treatment	−0.06	0.09	0.33
$R^2 = 0.02$			
$Q = 52.92$, $P < 0.01$			
Treatment type			
Technique	−0.13	0.22	0.36
Therapeutic community	−0.16	0.22	0.26
Methadone maintenance	−0.15	0.24	0.34
Other	−0.11	0.26	0.52
Structured approach	−0.19	0.17	0.07
$R^2 = 0.12$			
$Q = 47.32$, $P < 0.01$			
Treatment context			
Theoretical development	0.00	0.06	0.92
Program maturity	0.07	0.13	0.38
$R^2 = 0.01$			
$Q = 53.12$, $P < 0.01$			
Study context			
Document date	−0.00	0.01	0.99
Document type	−0.25	0.18	0.04
Funding source	−0.38	0.26	0.03
Researcher allegiance	0.02	0.17	0.84
$R^2 = 0.21$			
$Q = 42.83$, $P < 0.01$			
Final model			
Gender	−0.07	0.06	0.12
Race/ethnicity	−0.38	0.28	0.07
Age	0.02	0.01	0.04
Document type	−0.14	0.15	0.18
Funding source	−0.33	0.23	0.06
$R^2 = 0.34$			
$Q = 35.54$, $P = 0.02$			

Other meta-analyses that have examined the effectiveness of drug abuse treatment have also found positive and significant outcomes (Brewer et al., 1998; Glanz et al., 1997; Griffith et al., 2000; Marsch, 1998; Stanton and Shadish, 1997), although differences in methodology across the individual meta-analyses (e.g. selection criteria, effect size index, weighting procedures) make comparison of average effect sizes problematic.

What about the magnitude of the treatment effects found in this meta-analysis? Are they important or meaningful? An effect size is a number whose meaning is not obvious unless placed within some context or translated into an equivalent value that is more understandable and relevant for policy and programming purposes. Technically, an effect size (using the standardized mean difference) is the distance, on some outcome variable, that the average client in the treatment group is from the average client in the comparison group as measured in standard deviation units. But what does an effect size of 0.30 mean? Saying that the average outcome of the treatment group is 0.30 standard deviations above that of the comparison group is probably not very informative or helpful to the non-technical reader. Translating an effect size into a unit that is more useful and meaningful makes it easier for

policymakers, service providers, and others interested in treatment to understand the importance or relevance of the results.

Following the convention established by Cohen (1988), an effect size of 0.20 is considered to be small, 0.50 is medium, and 0.80 is large. But even these magnitude statements may have limited substantive meaning, particularly for policy discussions. Cohen himself offered them reluctantly, noting that the 'size' of an effect of (say) 0.30 may be small, medium, or large depending upon the disciplinary or research context within which it is derived. Alternatively, a number of effect size equivalents, usually expressed as percentages, are available that allow effect sizes to be understood in a more clinically relevant context.

The most common of these for social science oriented meta-analyses is the binomial effect size display (BESD) developed by Rosenthal and Rubin (1979, 1982). The BESD is the percentage of treatment participants and comparison participants who meet a common success criterion, defined arbitrarily as the median of the scores of the combined groups. Specific criteria for 'success' or 'failure' in a body of research studies are rarely available or even easily defined, so the overall median should be regarded as a hypothetical, but useful, representation of success rates. Using the BESD equivalents for the results of this meta-analysis, the effect size for drug-use outcomes translates to a 57% success rate for the treatment groups compared with 42% for the comparison groups. For the crime measure, the success rates are 53% for the treatment groups and 47% for the comparison groups. The difference in percentages for each of the outcomes is not large, but neither is it trivial. In this example, the 15-point difference for drug use outcomes is a 36% improvement in success for the treatment groups over the comparison groups (15/42).

5.2. Moderators of treatment effects

While the average effect sizes found in this meta-analysis for drug use and crime are positive, significant, and clinically meaningful, the individual study-level effect sizes display wide variability, which prompted an examination of factors that may account for differences in effect size across studies, particularly those that may have clinical implications.

The number of variables that were significant predictors of effect size were relatively few. Though perhaps disappointing, other meta-analyses examining predictors of outcomes have also found, at best, modest associations between client and program variables and outcomes (Brewer et al., 1998). Nevertheless, most of the factors that were predictive of effect size in this meta-analysis were clinically relevant.

5.2.1. Implementation

Programs that the coders rated as being well implemented (based on descriptions provided in the study reports) had higher drug abuse effect sizes. Indications of a well-implemented program included the use of a manual, training in the treatment protocol, monitoring of treatment delivery, low dropout rates, and other evidence that treatment was delivered as intended. The importance of program integrity to the effectiveness of human service interventions has been discussed by a number of researchers (e.g. Gresham et al., 1993; Hansen, 1996; Salend, 1984; Yeaton and Sechrest, 1981). Implementation issues are particularly germane when treatments found to be efficacious in research settings are moved out into community-based programs that may lack the resources, highly trained staff, careful monitoring, and carefully selected clients of the original study. While integrity may be difficult to maintain, program directors and staff who are aware of the potential for program drift, dilution of services, and staff burnout can institute policies and procedures to ensure that the program stays on track.

5.2.2. Theoretical development

The negative relationship found between level of theoretical development and drug abuse effect size is puzzling. One might expect that studies that are grounded in explicit hypothesis testing or in well-developed theories would exhibit higher effect sizes. This analysis found the opposite: treatment programs with little or no theoretical grounding had higher effect sizes. This may be indicative of a rift between theory and practice. Theoretically based interventions may not have been adequately developed for the realities of practical application, or the application of these interventions may have diverged from what was theoretically intended. The relationship between implementation and effect size described above lends support to this latter argument. On the other hand, a skeptic might argue that researchers who were aware of relatively weak results may simply have taken more time to buttress the report of their study with a theoretical rationale for their approach, whereas authors with relatively strong results may have been less inclined to thoroughly lay out the theoretical underpinnings of their interventions in favor of focusing on the favorable results. Thus, it could be that the disconnect between theory and effect sizes may be more apparent than real. While our data cannot identify the source of this result, they do point to an important question that calls for further investigation.

5.2.3. Researcher allegiance

The positive correlation between researcher allegiance and average effect size is, perhaps, not surprising. Although it is plausible that a researcher's allegiance to a program or modality may introduce bias into the

results of the evaluation, other explanations are just as likely. The rating of researcher allegiance was a judgment by the coder that the researcher's stance toward the program being evaluated was favorable, neutral, or unfavorable. However, since most researchers write their reports after they have seen the results, researchers with favorable data will likely express greater enthusiasm, and, as a result, these researchers may have been rated as more favorable toward the program. By contrast, researchers with ambiguous or less positive findings may take a more cautious tone and thus have been rated as more neutral. Therefore, it is unclear whether researcher allegiance begets higher effect sizes, or whether higher effect sizes led to stronger apparent researcher allegiance. Further complicating the matter is the fact that most of the ratings were favorable and none was unfavorable. Providing a clearer operational definition of 'researcher allegiance' and a greater response range might have resulted in a more sensitive measure.

5.2.4. Age and gender

Treatment appears to reduce crime to a greater degree among studies with samples consisting of younger adults as opposed to older adults. This makes intuitive sense, as we would expect drug treatment to have its largest impact among those groups that are actively involved in crime. That is, it would be fairly difficult to produce any dramatic reduction among groups that are not already involved in crime due to a 'floor' effect. Since young adults are far more likely to commit crimes than are older adults (Bureau of Justice Statistics, 2000), drug treatment should have its strongest effects in programs that contain large numbers of young adults. The age-related difference in effect size may be reinforced if older adults who do not 'mature out' of criminal activity are among the most refractory clients and, therefore, the least likely to respond to treatment.

Age was the only subject variable related to effect size. Gender was not significant in either model, although this cannot be construed to mean that there is no difference in treatment effectiveness between men and women. Rather, the variable 'gender' is a measure of the mix of males and females in a given study sample. Assuming that the gender mix of the study sample reflects the mix in the program, these findings suggest that programs in which women are a majority of the clients are no more or less effective than those in which women are in the minority. A test of the relative effectiveness of treatment for men and women would require that outcome data be analyzed and reported separately for each group within each program. Such data are not generally presented in treatment outcome studies.

5.2.5. Modality

It is noteworthy that treatment modality was not found to be related to effect size for either drug use or crime outcomes in the regression models, even though there were initial differences in the magnitude of effects across modalities[5]. This finding is consistent with other (narrative) reviews that have argued that none of the drug treatment modalities is clearly superior to any of the others for all drug users (Gerstein and Harwood, 1990; Institute of Medicine, 1990). Clearly, however, certain modalities are more appropriate for some clients than for others. The obvious case is methadone maintenance treatment, which is intended specifically for clients who are opiate dependent. While both therapeutic community and outpatient drug-free programs treat clients with a variety of drug preferences, therapeutic communities are better suited than outpatient programs for clients who require more intensive residential services because of their more severe drug histories.

5.2.6. Methodology

Adjustment of effect sizes by methodological features of the studies is an important factor to consider when examining outcomes across different studies. Certain methodological features tend to bias effect sizes upward; others lead to lower effect sizes. Lipsey has paid particular attention to methodology in his meta-analyses of correctional treatment (Lipsey, 1992; Lipsey and Wilson, 1998) and in a summary of findings from 302 meta-analyses of psychological, educational, and behavioral treatments (Lipsey and Wilson, 1993). Four of the methodological features found in the present meta-analysis to be associated with effect size variation—number of dependent variables, comparability of the treatment and comparison groups, participant attrition, and type of comparison condition—were also reported in the Lipsey studies, with the direction of effect also being the same. The fifth methodological feature found to be significant in this study—whether drug use was determined by drug testing (usually urinalysis)—was not included in the Lipsey analyses, but the positive association observed between using drug tests and effect sizes is consistent with Lipsey's finding (1992) that studies using more reliable and valid measures of outcome have larger effect sizes.

5.2.7. Missing data

In the database developed for this meta-analysis, many of the coded variables had a large percentage of

[5] Method-adjusted effect sizes (fixed effects model) by modality were as follows: treatment technique or service ($k = 50$), 0.31; methadone maintenance ($k = 8$), 0.45; therapeutic community ($k = 8$), 0.25; outpatient drug-free ($k = 8$), 0.37; detoxification ($k = 2$), 0.35; and other ($k = 2$), 0.32).

missing data; that is, the study report(s) did not include the information needed for the coder to calculate effect sizes or to complete specific questions about study or participant characteristics. To a certain extent, the amount of missing data is a function of the number and type of questions included in the codebook. The coding procedure attempted to extract some types of information that are not typically reported in most drug treatment studies, such as the source of funding for treatment slots, the type of ownership of the treatment facility, and the social class of participants. In addition, the type of document necessarily limited the amount of information that was reported, as in the case of conference abstracts. Still, it was surprising how often essential information about a study was unavailable or could only be estimated. The most noteworthy example was the number of studies in which the size of the sample on which the analyses were based could not be determined. Although the problem of missing data can be addressed through various imputation techniques (Little and Rubin, 1987; Pigott, 1994), it is clearly preferable to have as complete a dataset as possible, particularly for the values needed to calculate effect sizes and variances. Ensuring the availability of essential information to describe studies and to calculate effect sizes for future meta-analyses on drug abuse treatment could be accomplished several ways, including improvements in editorial review of submitted manuscripts and creation of a registry of funded studies that would require submission of standard information and data about studies upon their commencement and completion (such as the Cochrane Collaboration on health care intervention, http://www.cochrane.org; see also Tonks, 1999).

5.3. Limitations

Although many disciplines increasingly use meta-analysis to integrate findings from primary research studies, among some researchers meta-analysis remains a controversial, even suspect technique. Sharpe (1997) notes that critics tend to point to three main threats to the internal validity of meta-analytic studies: (1) mixing of dissimilar studies ('apples and oranges' problem), (2) publication bias ('file drawer' problem), and (3) inclusion of poor quality studies ('garbage in, garbage out' problem). In the present meta-analysis, we attempted to mitigate, if not eliminate, these validity problems.

5.3.1. Dissimilar studies

One of the criticisms of meta-analysis has been the 'apple and oranges' problem; meta-analysts have been faulted for defining the subject so broadly that different types of treatments, populations, and outcome variables are mixed indiscriminately and, therefore, important differences among treatments are lost (Eysenck, 1978; Presby, 1978). Meta-analysts have attempted to address this problem in different ways. One response is that 'apples and oranges' can be considered within the larger category of 'fruit' (Smith et al., 1980). Another approach (Lipsey and Wilson, 1995) is to point out that in primary studies, the sample of individual participants frequently includes a wide variety of characteristics that do not result in objections to the validity of the methodology, particularly if proper statistical controls are applied in analyzing the data. A third approach is to define the topic more narrowly (include oranges only), although even this more focused approach might not satisfy the critic who objects to the mixing of Valencia oranges and Navel oranges.

In this study, we addressed the 'apples and oranges' problem in a number of ways. A set of eligibility criteria was used to provide clear guidelines as to which studies were included and which were not, although in reality, some studies fell into a gray area that required a consensus judgment. For this paper, we selected those studies that used a treatment and comparison group design, used two conceptually distinct outcome variables, and included the main treatment modalities as independent variables in the regression analysis.

5.3.2. Publication bias

Publication bias arises from the tendency for researchers to submit for publication, and for editors to accept for publication, studies that report statistically significant results, leaving a large number of reports with statistically nonsignificant (or negative) results in the 'file drawer' (Rosenthal, 1979; Begg, 1994; Begg and Berlin, 1989; Easterbrook et al., 1991). Thus, confining a research synthesis to published literature only is likely to result in a biased sample of studies and to lead to an overestimation of the average effect of treatment.

We used two approaches to mitigating publication bias. First, during the literature search, we attempted to identify fugitive studies and unpublished studies by searching databases that include unpublished studies (e.g. Dissertation Abstracts, National Technical Information Service), by contacting agencies that fund drug treatment research for technical reports, and by asking drug abuse researchers to provide us with unpublished studies. Second, as part of the analysis, we assessed the degree to which publication bias may have affected the average effect size of the studies. One common approach

Table 5
Results of Rosenthal's fail–safe N test for publication bias (based on adjusted effect size)

Outcome	Mean z	Number of studies	Fail–safe N
Substance abuse	1.38	78	2,934
Criminal activity	0.68	25	52

is the 'fail-safe *N*' test developed by Rosenthal (1979), which provides an estimate of the minimum number of unpublished or unretrieved studies with nonsignificant results (i.e. no treatment effect) that would need to exist in order to bring the significance level of the average effect size for a set of studies down to a 'just significant' level.

For the current set of studies, the analysis was conducted for each outcome measure. The results are shown in Table 5. Although there are no established guidelines for what constitutes a critical number of unpublished or unretrieved studies, the obtained fail-safe *N* for this dataset indicates that for drug abuse outcomes it is highly unlikely that publication bias was a factor in this meta-analysis. That is, it is implausible to believe that we were unable to identify 2900 unpublished or fugitive studies reporting a nonsignificant effect size. For crime, however, the situation is different. Fifty unretrieved studies with effect sizes at or near zero would be sufficient to make the average effect size of the total set of drug treatment studies reporting on crime outcomes nonsignificant. Even so, although publication bias could more plausibly influence crime effect size than drug-use outcomes, it seems to be an unlikely explanation for the average crime effect size found in this study.

5.3.3. *Poor quality studies*

The eligibility criteria for this meta-analysis did not include methodological quality as a basis for selecting studies. Early meta-analyses were criticized on the grounds that the results of a research synthesis should be based only on those studies that are methodologically strong in order to avoid the 'garbage-in, garbage-out' problem (Bangert-Drowns, 1986; Slavin, 1986). A comprehensive literature search will undoubtedly turn up a certain percentage of studies that suffer from serious methodological problems. But no study is completely free of limitations or weaknesses in methodology, and any attempt to judge which studies are 'strong' and which are 'weak' is subject to theoretical or other biases of the researcher developing or applying the criteria for assessing study quality. Rather than eliminating methodologically weak studies altogether, an alternative is to test empirically whether methodological characteristics or assessments of quality has an impact on effect size. In this study, we adjusted study-level effect sizes by methodological characteristics that were found to significantly predict effect size. In addition, we used the EM technique to impute missing data values; while this technique allows the use of all studies in the regression analysis, thereby increasing power, it does introduce some additional uncertainty into the results.

Related to study quality, few studies report data on motivation. As a result, this meta-analysis was unable to assess the impact of initial motivation on treatment outcomes. Particularly in nonrandom assignment studies, motivation may be an important factor in explaining post-treatment outcome differences between the groups.

Finally, because meta-analysis is a secondary analysis technique, the generalizability of the results of a meta-analysis is limited to the characteristics of the set of studies included in the analysis. A particular meta-analysis is based on observational data rather than experimental data; that is, its results and conclusions are based on 'found' data and are thus limited to the characteristics of the data available in the studies that it samples. What can be known through meta-analysis largely depends upon the state of the literature regarding the research question. For example, if certain populations are excluded from, or are greatly underrepresented in, drug treatment research studies, the results of a meta-analysis can offer no conclusions regarding the effectiveness of drug treatment with those populations. Furthermore, the degree to which a meta-analysis can provide evidence of causal relationships among variables depends upon the types of design that have been used in the primary research (Miller and Pollock, 1994). Generally, a meta-analysis based wholly or largely on experimental studies using random assignment can make strong statements about causality. In the drug treatment field, however, most program evaluations use quasi-experimental designs or single-group designs; thus, the relationships that can be tested through meta-analysis are mainly correlational, and causal statements that the researcher may make derive more from logic than from the study's design. In most cases, the best that can be done is to identify program and client variables that reliably covary with effect size.

6. Conclusions

With these caveats in mind, the findings from this meta-analysis indicate that drug abuse treatment, as it is practiced in the United States, is effective in reducing drug use and crime. Overall, people with drug abuse problems are better off being in treatment than not. To the extent that the drug treatment modalities examined in this study are not unique to the United States, the findings are likely to have policy and programming implications for treating drug abuse problems in other countries. Considering the positive results from this meta-analysis, as well as the findings from other meta-analyses and narrative reviews of drug treatment, it would seem appropriate to cease asking whether treatment for drug abuse is effective and begin asking instead how treatment can be improved and how it can be tailored to the needs of different types of clients. Although more refined meta-analyses can address some of the issues involved in answering these questions,

they are dependent on the moderating and mediating variables reported in the original studies. The findings from meta-analyses can suggest directions for further research, but the field will advance only as researchers carry out well designed studies that manipulate treatment conditions, rather than examine them post hoc, and that take into account the diversity of drug-abusing populations.

Acknowledgements

Support for this study came from grant R01DA09151 from the National Institute on Drug Abuse (NIDA). The views expressed in the paper, however, do not necessarily represent the position of NIDA or of the US Department of Health and Human Services. Thanks are due to Jane Chen for data management support, to Sheila Brooks for data entry, and to Kai McCormack for literature searches, document retrieval, database management, and other essential support tasks. Thanks are also due to Michael Gregg for preparing the tables and figure. Kris Langabeer deserves acknowledgement for a careful edit of the entire manuscript. The coders of the studies were Chris Aberson, Amy Maher, David Whelan, Xiadong Zhou, Marylynn Russell, and Xuscheng Li. We are also grateful for the advice of Betsy Becker and Vicki Pollack, our meta-analysis consultants. The paper also benefited from a critical review by Lynn Brecht and from the comments of two anonymous reviewers.

Appendix A: Study eligibility criteria for drug abuse treatment meta-analysis project

1) The treatment or intervention is directed toward changing the drug use and/or related behaviors or attitudes of adult (18 or older) illicit drug users.
2) Excluded are clinical studies of anti-addiction medications that have not been approved by the FDA for general use, studies conducted within nal justice system, studies that focus on the processes of treatment (e.g. counselor training, clinic management, assessment techniques), studies of methadone dosage, and studies of improfessionals and employee assistance programs.
3) The date of the document reporting on the study is between 1965 and 1996 (inclusive).
4) The document reporting the study can be either published or unpublished.
5) The document reporting on the study is in English.
6) The setting of the study is the United States (50 states or District of Columbia) or Canada.
7) The study includes quantitative outcome variables in which one or more treatment conditions is compared with one or more comparison conditions. The comparison condition may be 'no treatment,' 'typical or usual treatment,' 'placebo treatment,' or another condition in which the intention is not intended to produce change (or is intended to produce only minimal change) in the outcome variables for the comparison group. The comparison condition may also be respondents' baseline measurements, as in a single-group pretest/posttest study. Studies that compare two types of treatment against each other (e.g. methadone maintenance vs. therapeutic community) are excluded.
8) Random assignment of participants to treatment and comparison conditions is not necessary, but the study report should include data that allows assessment of the comparability of the treatment and comparison groups at baseline. These data could include an indication that the researcher attempted matching in constructing the groups or background characteristics that would permit a judgment on initial group equivalence.
9) Exclude 'treatment-by-treatment' studies, that is, studies in which two or more established treatments, roughly equivalent in strength and intensity, are compared.
10) Exclude 'dosage' studies, that is, studies that compare outcomes by different levels of methadone dose or those that compare outcomes by

Appendix B. Variable clusters

Methodological variables

Number of dependent variables	Continuous variable
Post-test sample size	Continuous variable
Research design	1 = pre experimental, 2 = quasi-experimental/correlational/ex post facto, 3 = true experimental

Assignment of subjects	1 = nonrandom, no matching, 2 = nonrandom, treatment & control matched, 3 = random/quasi random
Comparability of treatment, comparison groups	0 = no differences or differences of no/uncertain importance, 1 = differences judged important by coders
Attrition rate of treatment group	0 = less than 20%, 1 = greater than 20%
Similarity between treatment, control groups	7 point scale: 1 = very different/not equivalent, 7 = very similar/equivalent
Duration of follow-up period	0 = 0–13 weeks, 1 = 14–26 weeks, 2 = 27–52 weeks, 3 = 52+weeks
Measure of drug use	0 = study did not measure drug use, 1 = study measured drug use
Reliability of measures	3 point scale: 1 = poorly conducted, 3 = well conducted
Outcome type	0 = crime variable, 1 = drug variable
Treatment for Control Group	0 = standard, placebo, alternative treatment, 1 = no treatment or minimal treatment
Subject variables	
Gender	% male in treatment group: 1 = 0–49, 2 = 50–70, 3 = 71–89, 4 = 90+
Ethnicity	0 = majority of sample not white, 1 = majority of sample white
Age	Mean age for treatment group
Marital status*	% married or living as married
Primary drug cocaine	0 = not reported as primary drug problem, 1 = reported as primary drug problem
Primary drug heroin	0 = not reported as primary drug problem, 1 = reported as primary drug problem
Primary drug polydrug	0 = not reported as primary drug problem, 1 = reported as primary drug problem
Frequency of drug use*	1 = less than daily, 2 = daily
Treatment variables	
Treatment intensity and integrity	
Length of treatment (weeks)*	continuous variable
Hours of contact time*	continuous variable
Assessment of treatment integrity	variation/degradation in implementation/delivery of treatment. 5 point scale: 1 = implementation of treatment was entirely complete, 5 = implementation of treatment was too flawed to provide any confidence in treatment implementation
Integrity of treatment delivery	possibility that treatment was not fully delivered, weak, crosses over to contaminate controls, etc. 3 point scale: 1 = poorly conducted, 3 = well conducted
Characteristics of treatment condition	
Treatment type	
Detoxification	0 = no, 1 = yes
Technique	a specific technique was being evaluated, such as acupuncture, anger management, contingency management, counseling, relapse prevention, skills training, etc. 0 = no, 1 = yes
Methadone maintenance	0 = no, 1 = yes
Outpatient drug free	0 = no, 1 = yes
Therapeutic community	0 = no, 1 = yes
Other	0 = no, 1 = yes
Role of researcher*	1 = Delivered treatment, 2 = involved in planning, 3 = influential but no direct role in treatment, 4 = affiliated with treatment, capacity not specified, 5 = independent of the treatment
Use of structured approach	0 = no, 1 = yes
Program control*	1 = University, 2 = VA Medical Center, 3 = Community facility, 4 = Other
Drug test frequency*	9 point scale: 1 = daily, 9 = no drug testing done

different duration of treatment.

Treatment context	
Level of theoretical of the development	5 point scale: 1 = atheoretical, 5 = integrated theory
Maturity of program	1 = relatively new, 2 = established, 3 = defunct
Study context variables	
Document date	Continuous variable (2-digit year)
Published document	0 = not published, 1 = published
Federal funding	0 = not federally funded, 1 = Federally funded
Researcher allegiance	3 point scale: 1 = unfavorable, 3 = favorable

This variable was not included in the regression model due to substantial occurrence of missing data.

References

Andrews, D.A., Zinger, I., Hoge, R.D., Bonta, J., Gendreau, P., Cullen, F.T. 1990. Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. Criminology 28, 369–404.

Anglin, M.D., Hser, Y.-I. 1990. Treatment of drug abuse. In: Tonry, M., Wilson, J.Q. (Eds.), Drugs and Crime. University of Chicago Press, Chicago, pp. 393–460.

Apsler, R., 1994. Is drug abuse treatment effective? American Enterprise, March/April, 46–53.

Apsler, R., Harding, W.M. 1991. Cost-Effectiveness Analysis of Drug Abuse Treatment: Current Status and Recommendations for Future Research. In Background Papers on Drug Abuse Financing and Services Research (NIDA Drug Abuse Services Research Series 1). National Institute on Drug Abuse, Rockville, MD, pp. 58–81.

Bangert-Drowns, R.L. 1986. Review of developments in meta-analytic method. Psychol. Bull. 99, 388–399.

Begg, C.B. 1994. Publication bias. In: Cooper, H., Hedges, L.V. (Eds.), The Handbook of Research Synthesis. Russell Sage Foundation, New York, pp. 399–409.

Begg, C.B., Berlin, J.A. 1989. Publication bias and dissemination of clinical research. J. Nalt. Cancer Inst. 81, 107–115.

Bennett, W.J., DiIulio, J., Jr, Walter, J.P. 1996. Body Count: Moral Poverty...and How to Win America's War Against Crime and Drugs. Simon and Schuster, New York.

Berg, W.E. 1992. Evaluation of community-based drug abuse treatment programs: a review of the research literature. In: Freeman, E.M. (Ed.), The Addiction Process: Effective Social Work Approaches. Longman, New York, pp. 81–95.

Brewer, D.D., Catalano, R.F., Haggerty, K., Gainey, R.R., Fleming, C.B. 1998. A meta-analysis of predictors of continued drug use during and after treatment for opiate addiction. Addiction 93, 73–92.

Brown, B.S. 1984. Treatment of non-opiate dependency: Issues and outcomes. In: Smart, R.G., Glaser, F.B., Israel, Y., et al. (Eds.), Research Advances in Alcohol and Drug Problems, vol. 8. Plenum Press, New York.

Bureau of Justice Statistics, 2000, November 29. Criminal Offenders Statistics. Washington, DC: Author. Retrieved December 27, 2000, from the World Wide Web: http://www.ojp.usdoj.gov/bjs/crimoff.htm#data.

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences, second ed.. Academic Press, New York.

Cole, S.G., James, L.R. 1975. A revised treatment typology based on the DARP. Am. J. Drug Alcohol Abuse 2, 37–50.

Cole, S.G., Waterson, O. 1976. A treatment typology for drug abuse in the DARP: 1971–1972 admissions. In: Sells, S.B., Simpson, D.D. (Eds.), The Effectiveness of Drug Abuse Treatment, vol. 3. Ballinger, Cambridge, MA, pp. 201–252.

Cooper, H.M. 1984. The Integrative Research Review: A Systematic Approach. Sage Publications, Beverly Hills.

Cooper, H.M. 1989. Integrating Research: A Guide for Literature Reviews, second ed.. Sage Publications, Newbury Park, CA.

Cooper, H.M., Hedges, L.V. (Eds.),The Handbook of Research Synthesis. Russell Sage Foundation, New York 1994..

Cooper, J.R., Altman, F., Brown, B.S., Czechowicz, D. (Eds.),Research on the Treatment of Narcotics Addiction: State of the Art (NIDA Treatment Research Monograph). National Institute on Drug Abuse, US Department of Health and Human Services, Rockville, MD 1983..

Crits-Christoph, P., Siqueland, L. 1996. Psychosocial treatment for drug abuse: selected review and recommendations for national health care. Arch. Gen. Psychiatry 53, 749–756.

De Leon, G., Staines, G., Sacks, S. 1997. Passages: a therapeutic community oriented day treatment model for methadone maintained clients. J. Drug Issues 27, 341–366.

Dempster, A.P., Laird, N.M., Rubin, D.B. 1997. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Stat. Soc. 39, 1–38.

Easterbrook, P.J., Berlin, J.A., Gopalan, R., Matthews, D.R. 1991. Publication bias in clinical research. Lancet 337, 867–872.

Eysenck, H. 1978. An exercise in mega-silliness. Am. Psychol. 33, 517.

Farabee, D., Fredlund, E. 1996. Self-reported drug use among recently admitted jail inmates: Estimating prevalence and treatment needs. Subst. Use Misuse 31, 423–435.

Gerstein, D.R., Harwood, H.J. 1990. Treating Drug Problems, vol. 1. National Academy Press, Washington, DC.

Glanz, M., Klawansky, S., McAullife, W., Chalmers, T. 1997. Methadone versus l-alpha-acetylmethadol (LAAM) in the treatment of opiate addiction: a meta-analysis of the randomized, controlled trials. Am. J. Addict. 6, 339–349.

Glass, G.V., McGaw, B., Smith, M.L. 1981. Meta-Analysis in Social Research. Sage Publications, Beverly Hills.

Gresham, F.M., Gansle, K.A., Noell, G.N., Cohen, S., Rosenblum, S. 1993. Treatment integrity of school-based behavioral intervention studies: 1980–1990. Sch. Psychol. Rev. 22 (2), 254–272.

Griffith, J.D., Rowan-Szal, G.A., Roark, R.R., Simpson, D.D. 2000. Contingency management in outpatient methadone treatment: a meta-analysis. Drug Alcohol Depend. 58 (1–2), 55–66.

Hansen, W.B. 1996. Pilot test results comparing the all star program with seventh grade D.A.R.E.: program integrity and mediating variable analysis. Subst. Use Misuse 31, 1359–1377.

Harrison, L. 1995. The validity of self-reported data on drug use. J. Drug Issues 25, 91–111.

Hedges, L.V., Olkin, I. 1985. Statistical Methods for Meta-Analysis. Academic Press, Orlando, FL.

Hubbard, R.L. 1992. Evaluation and treatment outcome. In: Lowinson, J.H., Ruiz, P., Millman, R.B., Langrod, J.G. (Eds.), Substance Abuse: A Comprehensive Textbook, second ed.. Williams and Wilkins, Baltimore, pp. 596–611.

Hunter, J.E., Schmidt, F.L. 1990. Methods of Meta-Analysis: Correcting Error and Bias in Research Findings. Sage Publications, Newbury Park.

Institute of Medicine, 1990. Broadening the Base of Treatment for Alcohol Problems. National Academy of Sciences, Washington, DC.

Kleber, H.D. 1989. Treatment of drug dependence: what works. Int. Rev. Psychiatry 1, 81–100.

Krauthammer, C., 1997. Drug-treatment hustle. Washington Post, February 20, 1997.

Landry, M.J. 1995. Overview of Addiction Treatment Effectiveness. Officer of Applied Statistics, Substance Abuse and Mental Health Services Administration, Rockville, MD.

Lipsey, M.W. 1992. Juvenile delinquency treatment: a meta-analytic inquiry into the variability of effects. In: Cook, T.D., et al. (Ed.), Meta-Analysis for Explanation: A Casebook. Russell Sage Foundation, New York, pp. 83–127.

Lipsey, M.W., Wilson, D.B. 1993. The efficacy of psychological, educational, and behavioral treatment: confirmation from meta-analysis. Am. Psychologist 48, 1181–1209.

Lipsey, M.W., Wilson, D.B. 1995. Reply to comments on Lipsey and Wilson 1993. Am. Psychol. 50, 113–115.

Lipsey, M.W., Wilson, D.B. 1998. Effective intervention for serious juvenile offenders. In: Loeber, R., Farrington, D.P. (Eds.), Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions. Sage Publications, Thousand Oaks, CA.

Little, R.J.A., Rubin, D.B. 1987. Statistical Analysis with Missing Data. Wiley, New York.

Marsch, L.A. 1998. The efficacy of methadone maintenance interventions in reducing illicit opiate use, HIV risk behaviors and criminality: a meta-analysis. Addiction 93 (4), 515–532.

McLellan, A.T., O'Brien, C.P., Metzger, D., Alterman, A.I., Cornish, J., Urschel, H. 1992. How effective is substance abuse treatment: compared to what. In: O'Brien, C.P., Jaffe, J.H. (Eds.), Addictive States. Raven Press, New York, pp. 231–252.

McLellan, A.T., Woody, G.E., Metzger, D., McKay, J., Durrell, J., Alterman, A.I., O'Brien, C.P. 1996. Evaluating the effectiveness of addiction treatments: reasonable expectations, appropriate comparisons. Milbank Quart. 74 (1), 51–85.

Messina, N., Wish, E.D., Nemes, S., Wraight, B. 2000. Correlates of underreporting of post-discharge cocaine use among therapeutic community clients. J. Drug Issues 30, 119–132.

Miller, N., Pollock, V. 1994. Meta-analytic synthesis for theory development. In: Cooper, H., Hedges, L.V. (Eds.), The Handbook of Research Synthesis. Russell Sage Foundation, New York, pp. 457–483.

National Association of State Alcohol and Drug Abuse Directors, 1990. Treatment works: The tragic cost of undervaluing treatment in the "drug war". NASADAD, Washington.

National Institute on Drug Abuse and Office for Treatment Improvement, 1991. Uniform program reporting: Guidelines for providers of alcohol and other drug abuse treatment services. National Institute on Drug Abuse and Office for Treatment Improvement, Rockville, MD.

Nurius, P.S., Yeaton, W.H. 1987. Research synthesis reviews: an illustrated critique of 'hidden' judgments, choices, and compromises. Clin. Psychol. Rev. 7, 695–714.

Pearson, F.S., Lipton, D.S. 1999. A meta-analytic review of the effectiveness of corrections-based treatments for drug abuse. Prison J. 79, 384–410.

Pigott, T.D. 1994. Methods for handling missing data in research synthesis. In: Cooper, H.M., Hedges, L.H. (Eds.), The Handbook of Research Synthesis. Russell Sage Foundation, New York, pp. 163–176.

Prendergast, M., Podus, D., Chang, E. 2000. Program factors and treatment outcome in drug dependence treatment: an examination using meta-analysis. Subst. Use Misuse 35 (12–14), 1931–1965.

Presby, S. 1978. Overly broad categories obscure important differences between therapies. Am. Psychologist 33, 514–516.

Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. Psychol. Bull. 86, 638–641.

Rosenthal, R. 1991. Meta-Analytic Procedures for Social Research (Revised Edition). Sage Publications, Newbury Park, CA.

Rosenthal, R. 1995. Writing meta-analytic reviews. Psychol. Bull. 118, 183–192.

Rosenthal, R., Rubin, D.B. 1979. A note on percent variance explained as a measure of the importance of effects. J. Appl. Soc. Psychol. 9, 395–396.

Rosenthal, R., Rubin, D.B. 1982. Comparing effect sizes of independent studies. Psychol. Bull. 92, 500–504.

Salend, S.J., 1984. Therapy outcome research: threats to treatment integrity. Behav. Mod. 8, 211–222.

Schafer, J. 1999. NORM (Version 2.02 for WINDOWS 95/98/NT) [Computer software]. Department of Statistics, Pennsylvania State University, University Park.

Sharpe, D. 1997. Of apples and oranges, file drawers and garbage: why validity issues in meta-analysis will not go away. Clin. Psychol. Rev. 17, 881–901.

Sisk, J.E., Hatziandreu, E.J., Hughes, R. 1990. The Effectiveness of Drug Abuse Treatment: Implications for Controlling AIDS/HIV Infection (AIDS-Related Issues Background Paper 6; OTA-BP-H-73). Office of Technology Assessment, US Congress, Washington, DC.

Slavin, R.E. 1986. Best-evidence synthesis: an alternative to meta-analytic and traditional reviews. Educ. Res. 15, 5–11.

Smith, M.L., Glass, G.V., Miller, T.I. 1980. The Benefits of Psychotherapy. Johns Hopkins University Press, Baltimore.

Sorensen, J.L., Copeland, A.L. 2000. Drug abuse treatment as an HIV prevention strategy: a review. Drug Alcohol Depend 59, 17–31.

Sorensen, J.L., Acampora, A.P., Trier, M., Gold, M. 1987. From maintenance to abstinence in a therapeutic community: follow-up outcomes. J. Psychoactive Drugs 19, 345–351.

Stanton, M.D., Shadish, W.R. 1997. Outcome, attrition, and family-couples treatment for drug abuse: a meta-analysis and review of the controlled, comparative studies. Psychol. Bull. 122, 170–191.

Tobler, N.S. 1997. Meta-analysis of adolescent drug prevention programs: results of the 1993 meta-analysis. In: Bukoski, W.J. (Ed.), Meta-Analysis of Drug Abuse Prevention Programs (NIDA Research Monograph 170). National Institute on Drug Abuse, Rockville, MD, pp. 5–68.

Tonks, A. 1999. Registering clinical trials. Br. Med. J. 319, 1565–1568.

Wells, E.A., Hawkins, J.D., Catalano, R.F. 1988a. Choosing drug use measures for treatment outcome studies. I. The influence of measurement approach on treatment results. Int. J. Addict. 23, 851–873.

Wells, E.A., Hawkins, J.D., Catalano, R.F. 1988b. Choosing drug use measures for treatment outcome studies. II. Timing baseline and follow-up measurement. Int. J. Addict. 23, 875–885.

Yeaton, W.H., Sechrest, L. 1981. Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. J. Consult. Clin. Psychol. 49, 156–167.

Documents used in the meta-analysis*

Abramowitz, S.I., Pantleo, P.M., 1972. The effectiveness of brief methadone withdrawal among urban opiate addicts. Int. J. Addict. 7, 629–635.

Ashery, R.S., 1983. Psychotherapy for methadone maintained opiate addicts: A report of two studies. (NIDA Treatment Research Report Series). National Institute on Drug Abuse, Rockville, MD.

Avants, S.K., Margolin, A., Chang, P., Kosten, T.R., Birch, S., 1995. Acupuncture for the treatment of cocaine addiction: Investigation of a needle puncture control. J. Subst. Abuse Treat 12, 195–205.

Azrin, N.H., McMahon, P.T., Donohue, B., Besalel, V.A., Lapinski, K.J., Kogan, E.S., Acierno, R.E., Galloway, E., 1994. Behavior therapy for drug abuse: a controlled treatment outcome study. Behav. Res. Ther. 32, 857–866.

Baganz, P.C., Jarvis, J.J., 1971. Work and study release in the rehabilitation of the narcotic addict. Rehab. Lit. 32, 354–364.

Bale, R.N., Van Stone, W.W., Kuldau, J.M., Engelsing, T.M., Elashoff, R.M., Zarcone, V.P., Jr, 1980. Therapeutic communities vs. methadone maintenance: A prospective controlled study of narcotic addiction treatment—Design and 1-year follow-up. Arch. Gen. Psych. 37, 179–193.

Bleiberg, J.L., Devlin, P., Croan, J., Briscoe, R., 1994. Relationship between treatment length and outcome in a therapeutic community. Int. J. Addict 29, 729–740.

Burling, T.A., Seidner, A.L., Robbins-Sisco, D., Krinsky, A., Hanser, S.B., 1992. Batter up! Relapse prevention for homeless veteran substance abusers via softball team participation. J. Subst. Abuse 4, 407–413.

Caldwell, H.L., 1982. The use of cognitive behavior modification in the treatment of heroin addicts on methadone maintenance. Unpublished doctoral dissertation, California School of Professional Psychology at Fresno. UMI ProQuest Digital Dissertations AAT8324468.

Catalano, R.F., Hawkins, J.D., 1985. Project skills: Preliminary results from a theoretically based aftercare experiment. R. S. Ashery. Progress in the development of cost-effective treatment for drug abusers. Chap. NIDA Research Monograph 58, pp. 157–181. National Institute on Drug Abuse, Rockville, MD.

Catalano, R.F., Gainey, R.R., Fleming, C.B., Haggerty, K.P., Johnson, N.O., 1996. Unpublished manuscript, University of Washington at Seattle, ERIC Document Reproduction Service No. 0200.

Chang, G., Carroll, K.M., Behr, H.M., Kosten, T.R., 1992. Improving treatment outcome in pregnant opiate-dependent women. J. Subst. Abuse Treat. 9, 327–330.

Clark, W., 1991. Trial of acupuncture detoxification: a comparison of acupuncture to methadone for narcotic addiction treatment. State of California: Department of Alcohol and Drug Programs. (ERIC Document Reproduction Service No. 221).

Coletti, S.D., Kim, S., Newel, R., 1995. A long-term outcome evaluation of a drug treatment program for women with children: The PAR village experiment. Unpublished manuscript. Operation PAR, St. Petersburg, FL.

Craig, R.J., 1980. Effectiveness of low-dose methadone maintenance for the treatment of inner city heroin addicts. Int. J. Addict. 15, 701–710.

Dansereau, D.F., Joe, G.W., Simpson, D.D., 1995. Attentional difficulties and the effectiveness of a visual representation strategy for counseling drug-addicted clients. Int. J. Addict. 30, 371–386.

De Leon, G., Staines, G.L., Perlis, T.E., Sacks, S., McKendrick, K., Hilton, R., Brady, R., 1995. Therapeutic community methods in methadone maintenance (passages): an open clinical trial. Drug Alcohol Depend. 37, 45–57.

Desmond, D.P., Maddux, J.F., 1983. Optional versus mandatory psychotherapy in methadone maintenance. Int. J. Addict. 18, 281–290.

Dole, V.P., Robinson, J.W., Orraca, J., Towns, E., Searcy, P., Caine, E., 1969. Methadone treatment of randomly selected criminal addicts. New Engl. J. Med., 280, 1372–1375.

Fals-Stewart, W., Schafer, J., 1992. The treatment of substance abusers diagnosed with obsessive-compulsive disorder: An outcome study. J. Subst. Abuse Treat. 9, 365–371.

Fiorentine, R., Anglin, M.D. Enhancing Drug Treatment: Evaluation of the Los Angeles Target Cities Project, Year 01–03. Los Angeles: Drug Abuse Research Center, University of California, Los Angeles, ERIC Document Reproduction Service No. 0112, 0528.

Frawley, P., Smith, J., 1992. One-year follow-up after multimodal inpatient treatment for cocaine and methamphetamine dependencies. J. Subst. Abuse Treat. 9, 271–286.

Gariti, P., Auriacombe, M., Incmikoski, R., McLellan, A.T., Patterson, L., Dhopesh, V., Mezochow, J., Patterson, M., O'Brien, C., 1992. A randomized double-blind study of neuroelectric therapy in opiate and cocaine detoxification. J. Subst. Abuse 4, 299–308.

Gurevich, M., Duckworth, D., Imhof, J., Katz, J., 1996. Is auricular acupuncture beneficial in the inpatient treatment of substance-abusing patients? J. Subst. Abuse Treat. 13, 165–172.

Hall, S.M., Bass, A., Hargreaves, W.A., Loed, P.C., 1979. Contingency management and information feedback in outpatient heroin detoxification. Behav. Ther. 10, 443–451.

Havassy, B.E., Hall, S.M. 1981. Efficacy of urine monitoring in methadone maintenance. Am. J. Pychiatry 38, 1497–1500.

Hawkins, D., Catalano, R.F., Wells, E.A., 1986. Measuring effects of a skills training intervention for drug abusers. J. Consult. Clin. Psychol. 54, 661–664.

Higgins, S.T., Budney, A.J., Bickel, W.K., Badger, G.J., Foerg, F.E., Ogden, D., 1995. Outpatient behavioral treatment for cocaine dependence: One-year outcome. Exp. Clin. Psychopharm. 3, 205–212.

Higgins, S.T., Delaney, D.D., Budney, A.J., Bickel, W.K., Hughes, J.R., Foerg, F., Fenwick, J.W., 1991. A behavioral approach to achieving initial cocaine abstinence. Am. J. Psychiatry 148, 1218–1224.

Higgins, S.T., Stitzer, M.L., Bigelow, G.E., Liebson, I.A., 1986. Contingent methadone delivery: Effects on illicit-opiate use. Drug Alcohol Depend. 17, 311–322.

Iguchi, M.Y., Stitzer, M.L., Bigelow, G.E., Liebson, I.A., 1988. Contingency management in methadone maintenance: Effects of reinforcing and aversive consequences on illicit polydrug use. Drug Alcohol Depend. 22, 1–7.

Joe, G.W., Dansereau, D., Pitre, U., Simpson, D., 1997. Effectiveness of node-link mapping to enhance counseling for opiate addicts: a 12-month posttreatment follow-up. J. Nerv. Ment. Dis. 185, 306–313.

Johnson, B., Jainchill, N., Liberty, H., Ryder, J., Messina, M., Reynolds, S., Hossain, M., 1996. Dynamic recovery: Comparative study of a therapeutic community for homeless, substance abusing men. Unpublished manuscript. New York: National Development and Research Institutes, Inc., ERIC Document Reproduction Service No. 0474, 0559.

Khatami, M., Woody, G., O'Brien, C., Mintz, J., 1982. Biofeedback treatment of narcotic addiction: A double-blind study. Drug Alcohol Depend. 9, 111–117.

Kidorf, M., Stitzer, M.L., 1993. Contingent access to methadone maintenance treatment: Effects on cocaine use of mixed opiate-cocaine abusers. Exp. Clin. Psychopharm. 1, 200–206.

Konefal, J., Duncan, R., Clemence, C., 1994. The impact of the addition of an acupuncture treatment program to an existing metro-dade county outpatient substance abuse treatment facility. J. Addict. Diseases, 13, 71–99.

Kosten, T.R., Jalali, B., Hogan, I., Kleber, H.D., 1983. Family denial as a prognostic factor in opiate addict treatment outcome. The Journal of Nervous and Mental Disease, 171, 11–616.

Lam, J., Jekel, J., Thompson, K., Leaf, P., Hartwell, S., Florio, L., 1995. Assessing the value of a short-term residential drug treatment program for homeless men. J. Addict. Diseases 14, 21–39.

Lipton, D.S., Brewington, V., Smith, M., 1994. Acupuncture for crack-cocaine detoxification: Experimental evaluation of efficacy. J. Subst. Abuse Treat. 11, 205–216.

Maddux, J.F., Desmond, D.P., Vogtsberger, K.N., 1995. Patient-regulated methadone dose and optional counseling in methadone maintenance. Am. J. Addict. 4, 18–32.

Maddux, J.F., Prihoda, T.J., Desmond, D.P., 1994. Treatment fees and retention on methadone maintenance. J. Drug Issues 24, 429–444.

Magura, S., Rosenblum, A., Lovejoy, M., Handelsman, L., Foote, J., Stimmel, B., 1994. Neurobehavioral treatment for cocaine-using methadone patients: A preliminary report. J. Addictive Diseases 13, 143–160.

Margolin, A., Avants, S.K., Chang, P., Kosten, T.R., 1993. Acupuncture for the treatment of cocaine dependence in methadone maintained patients. Am. J. Addict. 2, 194–201.

McAuliffe, W.E., 1990. A randomized controlled trial of recovery training and self-help for opioid addicts in New Engl. and Hong Kong J. Psychoactive Drugs 22, 197–209.

McCarthy, J.J., Borders, O.T., 1985. Limit setting on drug abuse in methadone maintenance patients. Am. J. Psychiatry 142, 1419–1423.

McLellan, A.T., Arndt, I.O., Metzger, D.S., Woody, G.E., O'Brien, C.P., 1993. The effects of psychosocial services in substance abuse treatment. J. Am. Med. Assoc. 269, 1953–1960.

Milby, J.B., Clarke, C., Toro, C., Thornton, S., Rickert, D., 1980. Effectiveness of urine surveillance as an adjunct to outpatient psychotherapy for drug abusers. Inter. J. Addictions 15, 993–1001.

Miller, N., Pollock, V., 1994. Meta-analytic synthesis for theory development. In H. Cooper, L. V. Hedges (Eds.), The handbook of research synthesis (pp. 457–483). New York: Russell Sage Foundation.

Monahan, R.J., 1977 Secondary prevention of drug dependence through the transcendental meditation program in metropolitan Philadelphia. Inter. J. Addictions 12, 729–754.

Monti, P.M., Rohsenow, D.J., Michalec, E., Martin, R.A., Abrams, D.B., 1996. Coping skills treatment for cocaine abuse: Substance use outcomes at 3 months. Unpublished manuscript. Providence, RI: Center for Alcohol and Addiction Studies, Brown University.

Prendergast, M., Podus, D., McCormack, K. 1998. Bibliography of literature reviews on drug abuse treatment effectiveness. J. Subst. Abuse Treat. 15, 267–270.

Resnick, R., Washton, A., Stone-Washton, N., 1981. Psychotherapy and naltrexone in opioid dependence. L.S. Harris Problems of Drug Dependence, 1980: Proceedings of the 42nd Annual Scientific Meeting, The Committee on Problems of Drug Dependence, Inc. (Chap. NIDA Monograph 34, pp. 109–115). National Institute on Drug Abuse, Rockville, MD.

Shafii, M., Lavely, R., Jaffe, R., 1974. Meditation and marijuana. Am. J. Psychiatry 131, 60–63.

Silverman, K., Chutuape, M.A.D., Bigelow, G.E., Stitzer, M.L., 1997. Reinforcement of cocaine abstinence in treatment-resistant patients: Effects of reinforcer magnitude. L.S. Harris Problems of Drug Dependence, 1996: Proceedings of the 58th Annual Scientific Meeting, The College on Problems of Drug Dependence, Inc. (Chap. NIDA Research Monograph 174, p. 74). National Institute on Drug Abuse, Washington, DC.

Silverman, K., Higgins, S.T., Brooner, R.K., Montoya, I.D., Cone, E.J., Schuster, C.R., Preston, K.L., 1996. Sustained cocaine abstinence in methadone maintenance patients through voucher-based reinforcement therapy. Arch. Gen. Psychiatry 53, 409–415.

Silverman, K., Wong, C.J., Umbricht-Schneider, A., Montoya, I. D., Schuster, C.R., Preston, K.L., 1996. Voucher-based reinforcement of cocaine abstinence: Effects of reinforcement schedule. L.S. Harris Problems of Drug Dependence, 1995: 57th Annual Scientific Meeting, The College on Problems of Drug Dependence, Inc. (Chap. NIDA Research Monograph 162, p. 97). Rockville, MD: National Institute on Drug Abuse.

Sosin, M.R., Bruni, M., Reidy, M., 1995. Paths and impacts in the progressive independence model: A homelessness and substance abuse intervention in Chicago. J. Addict. Diseases 14, 1–20.

Sowder, B., 1992. Summary report: Waiting list community intervention project. Unpublished manuscript. Bethesda, MD: NOVA Research Company, Inc.

Stanton, M.D., Shadish, W.R., 1997. Outcome, attrition, and family-couples treatment for drug abuse: A meta-analysis and review of the controlled, comparative studies. Psychol. Bull. 122, 170–191.

Stanton, M.D., Todd, T.C., Steier, F., 1979. Outcome for structural family therapy with drug addicts. NIDA Research Monograph 27-Problems of Drug Dependence 1979. Proceedings Of The 41st Annual Scientific Meeting. The Committee On Problems Of Drug Dependence, Inc. (US Department of Commerce National Technical Information Service (NTIS), 415–421.

Stark, M., He, H., Booth, R., Smith, R., McAnulty, J., Stevens, S., Ashery, R., Beardsley, M., 1996. Correlates and consequences of entering drug treatment: A study of the NIDA cooperative agreement national database. Drugs Soc. 9, 199–212.

Sterling, R., Gottheil, E., Glassman, S., Weinstein, S., Serota, R., 1997. Patient treatment choice and compliance: Data from a substance abuse treatment program. Am. J. Addict. 6, 168–176.

Stone-Washton, N., Resnick, R.B., Washton, A.M., 1982. Naltrexone and psychotherapy. L. S. Harris Problems of Drug Dependence, 1981: Proceedings of the 43rd Annual Scientific Meeting, The Commitee on Problems of Drug Dependence, Inc. (Chap. NIDA Research Monograph 41, pp. 505–507). National Institute on Drug Abuse, Rockville, MD.

Wells, E.A., Jackson, R., Diaz, O.R., Stanton, V., Saxon, A.J., Krupski, A., 1995. Acupuncture as an adjunct to methadone treatment services. Am. J. Addict. 4, 198–214.

Wells, E.A., Peterson, P.L., Gainey, R.R., Hawkins, J.D., Catalano, R.F., 1994. Outpatient treatment for cocaine abuse: A controlled comparison of relapse prevention and Twelve-Step approaches. Am. J. Drug Alcohol Abuse 20, 1–17.

Yancovitz, S.R., Des Jarlais, D.C., Peyser, N.P., Drew, E., Friedmann, P., Trigg, H.L., Robinson, J.W., 1991. A randomized trial of an interim methadone maintenance clinic. Am. J. Public Health 81, 1185–1191.

*Because a given document may have reported on more than one study, the number of documents listed below is fewer than the number of studies used in the meta-analysis.