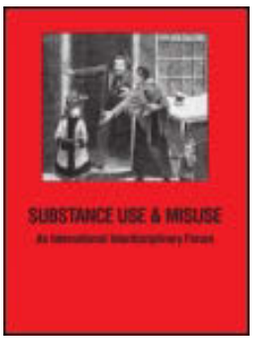

Substance Use & Misuse

ISSN: 1082-6084 (Print) 1532-2491 (Online) Journal homepage: https://www.tandfonline.com/loi/isum20

Program Factors and Treatment Outcomes in Drug Dependence Treatment: An Examination Using Meta-Analysis

Michael L. Prendergast, Deborah Podus & Eunice Chang

Substance Use & Misuse, 35(12–14), 1931–1965, 2000

Program Factors and Treatment Outcomes in Drug Dependence Treatment: An Examination Using Meta-Analysis

Michael L. Prendergast,* Deborah Podus, and Eunice Chang

Drug Abuse Research Center, Neuropsychiatric Institute, University of California, Los Angeles, California 90024, USA

ABSTRACT

In comparison with studies of client characteristics and treatment processes, limited research has been conducted on how program features of drug dependence treatment programs may affect client outcomes. Of particular interest are those characteristics of programs that may have a clinically significant impact on outcomes and that are amenable to change within programs. This study examines the impact of various program factors on client outcomes using data from a meta-analysis of drug dependence effectiveness studies ($n = 143$). Because of heterogeneity among studies, the data are analyzed in terms of type of outcome variable (drug use and crime), type of design (single-group and treatment-comparison group), and type of treatment (methadone maintenance, therapeutic communities, outpatient drug free, and detoxification). For the more valid treatment–comparison group studies, the weighted mean

* To whom correspondence should be addressed at UCLA Drug Abuse Research Center, 1640 Sepulveda Blvd., Suite 200, Los Angeles, CA 90025, USA. Telephone: (310) 445-0874. FAX: (310) 473-7885. E-mail: mlp@ucla.edu

effect size was 0.29 for drug use outcomes and 0.17 for crime outcomes. Program factors found to be significantly correlated with effect size in one or more modalities were decade of treatment, researcher involvement in treatment delivery, maturity of the program, counselor/client ratio, treatment implementation, treatment exposure, and methadone dosage.

Key words. Treatment effectiveness; Program effects; Meta-analysis; Treatment modalities

INTRODUCTION

Numerous reviews have summarized three decades of research that largely supports the effectiveness of drug user treatment generally and specific treatment modalities (Anglin and Hser, 1990; Apsler and Harding, 1991; Berg, 1992; Brown, 1984; Cooper et al., 1983; Gerstein and Harwood, 1990; Kleber, 1989; McLellan et al., 1992; Sisk et al., 1990). Despite evidence of the effectiveness of drug user treatment, substantial variation in outcomes among programs exists, even within the same modality. Researchers have found differences in such program factors as treatment philosophy, operating policies and protocols, program management, professional level of clinical staff, specific services offered, duration and intensity of services, and morale among staff and clients. These program differences often result in substantial variation in treatment outcomes, even after baseline client characteristics are controlled for (Ball and Ross, 1991; Hubbard et al., 1989; McLellan et al., 1993; Moos et al., 1995).

In comparison with studies of client characteristics and treatment processes, research has been limited on the program features of drug dependence treatment programs and how these may affect client outcomes. Of particular interest are those aspects of program structure that have a clinically significant impact on outcomes and that are amenable to change with a reasonable amount of effort and resources. Recently, researchers have focused greater attention on the way programs are organized, administered, and staffed, and the methods of service provision to clients. Measures to describe program dimensions and treatment processes have been developed for alcoholism and drug dependence treatment programs (Allison and Hubbard, 1985; Ball and Ross, 1991; McLellan et al., 1992; Moos et al., 1995; Polinsky et al., 1998; Timko, 1996). These various measures provide objective assessments of the characteristics of programs at a level of detail not previously available in treatment evaluation studies. Despite these

advances, however, identifying salient program factors that influence outcomes remains an elusive goal.

The present study examines the impact of various program factors on client outcomes using data from a meta-analysis of drug-dependence treatment effectiveness studies. Meta-analysis is a set of procedures and techniques used to statistically combine the results of two or more independent studies in order to provide a quantitative answer to a question of interest (for theoretical and practical discussions of meta-analysis, see Cooper, 1989; Cooper and Hedges, 1994; Glass et al., 1981; Hedges and Olkin, 1985; Rosenthal, 1991). This paper is intended to answer two questions based on a preliminary examination of the data gathered for the study.

- What are the average effect sizes of drug dependence treatment for drug use and crime outcomes?
- Within treatment modalities (methadone maintenance treatment [MMT], therapeutic communities [TC], outpatient drug free [ODF] detoxification [DETOX], and others), which selected program factors are correlated with estimated effect sizes for drug use and crime outcomes?

This paper focuses on studies that have assessed treatment effectiveness based on measures of individual client outcomes with respect to specific variables determined by the researcher to be of theoretical or clinical relevance. We focus exclusively on studies conducted in "real world" conditions as opposed to those conducted in controlled, experimental settings. Not included in this study is research that adopts other approaches to effectiveness such as cost–benefit analysis, clinical significance, program quality, or other perspectives. Moreover, although researchers have used numerous outcome variables of potential concern in assessing program effectiveness, in this paper we focus on only two outcome domains: drug use and criminal behavior. We selected these because they are the most commonly used in research over time and are of greatest general interest. Other outcome variables not examined here include education, employment, social and family relationships, psychological status, HIV risk behaviors, and housing.

The paper begins with a review of studies examining the relationship of treatment program factors and client outcomes. The following sections describe the methodology of the meta-analysis and the specific approach for this study of program factors. Findings from the analysis of the impact of program factors on outcomes are then presented. The paper concludes with a summary of findings and a discussion of the methodological issues and limitations.

PREVIOUS STUDIES

The few previous studies of drug dependence treatment that examined the relationship of program factors and client outcomes have generally found that program factors do make a contribution to explaining client outcomes, although the specific factors examined differ from study to study. Using data from a national survey of drug user treatment programs, McCaughrin and Price (1992) analyzed the relationship of measures of organizational characteristics and program quality to treatment effectiveness (based on estimates from clinical supervisors) within outpatient treatment programs. Better outcomes were found in units with a larger number of treatment staff, units with for-profit status, units with sobriety as a condition of treatment, units that provided formal post-treatment referrals, and units that conducted follow-up of clients following discharge. By contrast, poorer outcomes were associated with older program units and with those having a high staff/client ratio. Three measures of quality—accreditation, U.S. Food and Drug Administration (FDA) licensing, and quality assurance plans—yielded mixed results. Accredited programs had better outcomes, but programs licensed by the FDA and programs with quality assurance plans did not. Possible reasons offered by the authors for these latter (counterintuitive) findings are that FDA-licensed units may deal with difficult-to-treat clients who are more likely to continue drug or alcohol use and that quality assurance plans may sensitize staff to continuing drug or alcohol use, resulting in higher levels of client drug use reported by clinical supervisors.

In a study of six methadone programs, which had all originally been selected because of their reputed high quality, Ball and Ross (1991) found clear differences in the characteristics of the programs and in client outcomes among the programs. In examining the relative contributions of client, program, and process variables to outcomes using regression analysis, Ball and Ross found that program variables were better at explaining outcomes than were client or process variables, and that process variables were better than client variables. They concluded that "patients in programs with rehabilitation and long-term maintenance orientation, which also delivered more counseling services to patients and had more effective directors, tended to have better outcomes than patients in programs characterized by favorable staff/patient ratios in better facilities, where there was emphasis on administrative functioning rather than provision of services" (Ball and Ross, 1991, p. 231).

Magura and colleagues (1999) conducted a replication and extension of the Ball and Ross study, using a larger sample of methadone programs (17

separate clinics). Although by using factor analysis they identified different numbers and types of program quality factors than did Ball and Ross (possibly because of the larger sample of programs), they found that a number of the same program variables reported by Ball and Ross were associated with lower levels of patient drug use during treatment: number of counseling contacts, involvement by the clinic director in patient treatment, and experience of the clinic director.

Differences in treatment and administrative policies among programs have also been shown to have differential effects on client performance during and after treatment. For example, McGlothlin and Anglin (1981) found that an MMT program with low-dose methadone policies and strict discharge policies had poorer client retention and outcomes than did another program with policies favoring higher dosages and less strict discharge criteria. Other researchers (Ball and Ross, 1991; Caplehorn and Bell, 1991; Strain et al., 1999) reported similar results regarding the relationship between methadone dosage policy and treatment retention and outcomes.

Moos and collaborators (1999; 1998) also found that program and administrative policies are associated with treatment outcomes. Among the factors that they found to be associated with better outcomes were clear program policies, a 12-step treatment orientation, a more supportive and goal-directed staff environment, and a more supportive and goal-directed treatment milieu. Their work also suggests, however, that the impact of program factors may differ by age (Moos et al., 1995). Among older patients, a more supportive and well-organized program with outpatient aftercare may be more effective, whereas among younger patients, an intensive and directed treatment may be more efficacious. In a study of community residential programs, Moos and colleagues (1997) found that other program factors associated with better outcomes were high expectations for patients' functioning and a high proportion of staff in recovery.

In a somewhat different approach to examining the impact of program characteristics on outcomes, Timko (1996) developed the Physical and Architectural Characteristics Inventory (PACI) to assess the physical characteristics of residential psychiatric and substance dependence programs and their relationship with treatment outcome. The PACI includes seven domains: community accessibility, physical amenities, social–recreational aids, prosthetic aids, safety features, staff facilities, and space availability. In a sample of 94 hospital-based and community-based residential psychiatric and substance dependence programs, all but one of these dimensions (safety features) were positively correlated with at least one of five measures of outcome.

One study found little evidence of a program effect. Joe et al. (1983) assessed the impact of differences in treatment programs (within the same

modality) on outcomes among clients in the Drug Abuse Reporting Program (DARP) study. Within a modality (MMT, TC, and ODF), significant differences in outcomes were found for ODF programs, but not for MMT or TC agencies, and most of the differences found across ODF programs were explained by variation in client background characteristics. Findings indicated that "only a small percentage of variance in post-treatment outcomes... is uniquely related to treatment program differences" (p. 537).

As is evident from this brief summary of previous research, a variety of program characteristics have been examined as possible influences on treatment outcomes, and a number of these have been found to be associated with outcomes (see Table 1). The program characteristics assessed fall into several categories: program type and philosophy, treatment policies, management and administration, physical plant, accreditation and licensure, staff characteristics, types of services, and frequency of counseling contacts. Different researchers have selected different variables and have measured them in different ways, thus making it difficult to compare findings across studies. Each of the studies reviewed above used a sample of programs to attempt to determine the effects of program characteristics on treatment outcomes. An alternative, which is followed in this study, is to aggregate the results of a set of existing outcome studies and examine the relationship of program factors to treatment outcomes. Although this approach has the advantage of including the same set of program (and other) variables in the coding of the eligible studies, it has the disadvantage of having to rely on whatever information about program characteristics is available in a given study, which produces a dataset with high percentages of missing values for some variables. Still, the dataset used here is sufficiently complete to provide at least a preliminary examination of the relationship of selected program factors with treatment outcomes using meta-analysis.

METHODS

Definitions

Modality

For purposes of this study, modality refers to the traditional approaches that have been used to treat drug dependence in the United States—MMT, TC, ODF, and DETOX. Although short-term hospital and chemical dependency programs gained prominence in the 1980s and 1990s, only a few studies of these program types were found to be eligible for coding. These were included in an "other" category, along with multimodality programs

Table 1

Program Factors Found to be Associated with Positive (Better) Treatment Outcomes

Author	Modality	Program factors
Ball and Ross, 1991	Methadone	Rehabilitation orientation Long-term maintenance policy More frequent counseling services Effective clinic director
McCaughrin and Price, 1992	Outpatient	Larger treatment staff For-profit status Abstinence orientation Provision of post-treatment referrals Follow-up of clients after discharge Accreditation
McGlothlin and Anglin, 1981	Methadone	High-dose policy More lenient discharge policy
Magura et al., 1999	Methadone	More frequent counseling contacts Clinic director involvement in patient treatment More experienced clinic director
Moos et al., 1999	Inpatient	12-Step treatment orientation More regular and intensive outpatient mental health care
Moos et al., 1995	Inpatient	Younger patients: Family involvement Community consultation Social and work skills training Older/middle-age patients: Structured program policies Flexible discharge rules Comprehensive assessment Outpatient mental health care
Moos and Moos, 1998	Inpatient	Supportive and goal-directed staff environment 12-Step orientation Supportive and goal-directed treatment milieu
Moos et al., 1997	Residential	High expectations for clients' functioning Clear policies Structured programming High percentage of staff in recovery More emphasis on psychosocial
Timko, 1996	Residential	Community accessibility Physical amenities Social–recreational aids Prosthetic aids Staff facilities Space availability

and a few other programs that did not clearly fall within one of the main modalities. Even though the array of available types of drug dependence treatment has been categorized into these modalities, there is considerable variation within each modality. Because the focus here is on factors that influence outcomes in programs, studies that examined specific techniques that are used within programs (e.g., acupuncture, contingency management, skills training) are not included here.

Study

A study was defined as a comparison of two conditions designed to determine whether a treatment program for drug dependence was effective.[1] The comparison conditions could be either a treatment group and a comparison group (called a "treatment–comparison design" in this paper) or a pretest and a post-test assessment of a single group (called a "single-group design").[2] In treatment–comparison designs, the comparison condition could be either no treatment or some type of minimal treatment. Because it is difficult to justify not providing treatment to drug users in need of treatment, comparison groups in treatment evaluations usually involve minimal treatment or standard treatment, rather than no treatment. Studies in which two types of drug user treatment that are generally considered effective and that are of roughly equal strength or intensity (termed "treatment-by-treatment studies") were not eligible for coding because such studies are concerned with the *relative* effectiveness of different types of treatment (77 studies of this type were retrieved, but not coded). Typically, the eligibility criteria of meta-analyses exclude studies that lack a comparison group, largely for methodological reasons. Single-group designs were included in this meta-analysis, however, because many studies of the effectiveness of drug dependence treatment make use of this design, including major evaluations by the National Institute on Drug Abuse and the Center for Substance Abuse Treatment. For reasons explained below, effect sizes from the two types of design are analyzed separately.

Treatment Outcome

A treatment outcome is a behavior that the treatment is expected to change, measured either during treatment, at the end of treatment, and/or at one or more follow-up points. Most studies include multiple types of outcomes (e.g., drug use, crime, employment) and often multiple measures of a particular type of outcome (e.g., both self-report and urinalysis for drug use). With few exceptions (e.g., client satisfaction), an effect size was

calculated for all outcomes in a study for which sufficient quantitative information was available.

This current analysis focused on two outcomes: drug use and crime. These two variables are typically regarded as the main outcomes of drug user treatment, and most studies reported results for these two outcomes. Also, because this paper focuses on the relationship of program variables to outcomes, selecting two types of outcomes simplified the analysis. Future papers on this dataset will include a broader range of outcomes. Studies measured drug use either by self-report or by drug testing, or by both. Criminal behavior was assessed from self-report or, less often, from official records. Self-reported criminal behavior includes crimes that are not detected by the criminal justice system, whereas official records only include data on criminal justice processing (arrest, conviction, incarceration).

Program Variables

The program variables selected for this analysis (shown in Table 2) are a subset of the program-level variables that were included in the coding of eligible studies. The variables are either continuous or have been dichotomized from ordinal scales. The main reasons for not including other program variables were a large percentage of missing data and high correlations with other variables.

Table 2

Program Variables and Coding Definitions

Variable	Coding
Decade of study	1960s, 1970s, 1980s, 1990s
Role of researcher in designing or delivering treatment	1 = High involvement 0 = Low involvement
Maturity of program	1 = New program (2 years old) 0 = Older program
Counselor/client ratio (treatment group)	Mean number of clients per counselor
Methadone dose (methadone programs only)	Mean milligrams
Implementation of treatment	1 = Well implemented 0 = Poorly implemented
Length of treatment participation	Mean weeks
Amount of contact during treatment	Mean hours
Intensity of treatment contact	1 = High intensity 0 = Low intensity

Meta-Analytic Procedures

The procedure for conducting a meta-analysis consists of five main steps:

- Definition of the research topic, including questions of interest, independent and dependent variables, and eligibility criteria
- Identification and retrieval of studies based on the eligibility criteria
- Systematic coding of substantive and methodological characteristics of each study
- Calculation of effect sizes and direction of the treatment effect for all (or selected) dependent variables in each study
- Calculation of an estimated average effect size and examination of relationships between study characteristics and effect size.

Identification and Selection of Studies

Studies were identified and selected for inclusion in the meta-analysis on the basis of predetermined eligibility criteria. Briefly, a study was eligible for inclusion in the meta-analysis if it was an outcome evaluation of a drug dependence treatment program or technique for adults, conducted in the United States or Canada, and published or dated (for unpublished studies) from 1965 through 1996. A more detailed listing of the eligibility criteria appears in the Appendix.

We used three main strategies to identify relevant literature. First, 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, National Technical Information Service (NTIS), PsychINFO, Public Affairs Information Service (PAIS), Sociological Abstracts, and Social Work Abstracts. Second, we scanned both printed bibliographies that did not have an on-line counterpart and specialized bibliographies in substance dependence. In addition, the proceedings of conferences and professional meetings were reviewed. Finally, letters were sent to researchers, organizations, and agencies in the drug dependence field, requesting references to documents (and copies if available) that were not likely to be identified through standard sources. Requests for studies were also circulated at professional meetings and through drug-use-related newsletters. As documents were retrieved and catalogued, their reference lists were examined for other citations that appeared to be relevant.

Coding of Studies

The codebook consisted of 260 questions organized into five sections: study context, study methodology, subject characteristics, program characteristics, and dependent variable characteristics and effect size calculation. (A copy of the codebook is available from the authors.) Each study was coded by one of seven coders, all masters- or doctoral-level students. Coders attended a 2-day training session on the use of the codebook and on effect size calculation, and coded and discussed three practice studies before beginning work on the project. Coders met with project staff approximately every 2 weeks for the first several months of the study, less often thereafter, to discuss coding questions. In addition, coders and staff periodically all coded the same study and reviewed the results. Coders received a coding policy manual that was continuously updated to reflect new coding decisions. All coded studies were checked for discrepancies by one of two investigators before data entry.

Effect Size Calculation

The statistical methods used to calculate, combine, and analyze effect sizes were those of Hedges and Olkin (1985), supplemented by procedures presented in *The Handbook of Research Synthesis* (Cooper and Hedges, 1994). The most common method for estimating effect sizes in evaluation studies 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 this difference by the pooled standard deviation; that is, $d = M_t - M_c/SD_{pooled}$. For single-group designs, the difference between the post-test mean and the pretest mean is divided by the pretest standard deviation; that is, $d = M_{post} - M_{pre}/SD_{pre}$. If means and standard deviations are not available, effect sizes may be estimated using reported t, F, or χ^2 critical values, using formulas found in standard meta-analysis texts (e.g., Cooper and Hedges, 1994; Hedges and Olkin, 1985). Effect sizes for outcomes reported as proportions or percentages were calculated using the arcsin transformation (Cohen, 1988). For treatment–comparison group designs, the effect size was not corrected for differences in the outcome variable between the two groups at pretest.

Consistent with convention, an effect size in 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. For example, an effect size of 0.25 for a given study indicates that the average score of the treatment subjects on the outcome variable is one-fourth of a standard deviation higher than

the average score of the comparison subjects. By contrast, an effect size of −0.25 would indicate better performance for the comparison group compared with the treatment group. Following Cohen (1988), an effect size of 0.20 is considered small; 0.50 is moderate; and 0.80 is large.

Because the effect size calculated from means and standard deviations (the standardized mean difference) provides an overestimate of the population effect size for small samples, it has become standard practice in meta-analysis to apply a correction to all such effect sizes, regardless of sample size, in order to provide an unbiased effect size estimate (Hedges and Olkin, 1985).[3] A similar correction factor for effect sizes from proportions or percentages is not available.

Because many studies report outcome results for multiple measurement points (e.g., during treatment, end of treatment, post-treatment), we often had more than one measure for the same outcome variable in a particular study. To avoid creating dependencies among outcome variables within studies, we selected the effect size calculated from the first post-treatment assessment point or, if all measures were taken during treatment, from the assessment point nearest to the end of the treatment. To produce a single study-level effect size for each outcome category, effect sizes for multiple measures of either drug use or criminal behavior within a given study were averaged. Because studies with a large sample provide more precise and stable estimates of the population effect size than do studies with a small sample, each effect size estimate was weighted using the inverse of its variance.[4] To prevent studies with very large sample sizes from dominating the effect size averages, sample sizes of large studies were Windsorized at 240 for single-group designs and at 160 for treatment–comparison-group designs (80 for each group).[5]

Weighting each effect size by the inverse of its variance assumes a fixed effects model in which the combined individual effect sizes provide an estimate of a single, or common, population effect size (i.e., the population variance is zero). For this dataset of drug user treatment effectiveness studies, it is unlikely that this assumption is correct, even within the same modality of treatment, because between-study variation is as likely to influence effect size estimates as within-study variation. A more plausible assumption is that the obtained effect sizes are a sample 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 the variability of the population effect sizes, is added to the individual effect size variance. Both fixed effects and random effects weighted means are reported in Table 4.

Study Sample

The full dataset consists of 293 coded studies, but not all studies could be used in this analysis of program factors. Studies were dropped for four reasons. First, as noted above, studies that evaluated specific treatment techniques were not included ($n = 88$). Second, a study had to include at least one outcome variable related to drug use or crime. Any study that had neither a drug use nor a crime outcome variable was dropped ($n = 24$).[6] Third, because the analysis uses weighted effect sizes, each study-level effect size includes two components: the effect size itself and the weight, which is based on sample size. Thus, studies that lacked sufficient data to calculate an effect size (for drug use or crime) were dropped, as were studies (even those with effect sizes) for which sample size could not be determined ($n = 20$). Fourth, for single-group designs, individual effect sizes that must be calculated from within-group t-test statistics (because means and standard deviations are missing) do not estimate the same population parameter as do effect sizes calculated from pretest and post-test means and standard deviations. Thus, studies in which drug use and/or crime outcome variables were calculated from within-group t-tests were dropped ($n = 18$). The final set of studies analyzed in this paper consists of 143 studies of treatment programs. (A list of the references for these studies is available from the authors.)

Table 3 displays the characteristics of the studies included in the current analysis. Studies that used single-group designs were more common than treatment–comparison-group designs. For the latter type of studies, just over two-thirds used an active comparison group, with the remaining using a passive comparison group. More than two-fifths of the treatment–comparison-group studies assigned subjects randomly or quasi-randomly, whereas in half of the studies the assignment procedures involved neither randomization nor matching. Nearly three-quarters of the studies were reported in journal articles or technical reports. With regard to program modalities, one-quarter (26.6%) of the studies assessed outcomes of MMT programs, followed by TC programs (22.4%), ODF programs (20.3%), DETOX programs (20.3%), and other program types (10.5%). The time period of the studies was about equally divided between the 1970s, 1980s, and 1990s. Federal agencies were the primary funding source for more than one-half of the studies, although it was not possible to determine the funding source for about 30% of the studies. Virtually all studies were conducted in the United States, with Canada contributing only two

Table 3

Characteristics of Outcome Studies of Drug User Treatment Programs (N = 143)

	N	%
Type of design		
Single group	115	80.4
Treatment–comparison	28	19.6
Type of comparison conditions[a]		
Passive comparison		
No treatment	2	7.1
Delayed treatment/wait list	2	7.1
Minimal contact	5	17.9
Active comparison		
Routine treatment	10	35.7
Other	9	32.1
Assignment procedure[a]		
Random or quasi-random	12	42.9
Nonrandom; matching	2	7.1
Nonrandom; no matching	14	50.0
Publication type		
Journal article	71	49.7
Book or book chapter	10	7.0
Technical report	34	23.8
Dissertation	6	4.2
Unpublished paper	11	7.7
Other	10	7.0
Missing	1	0.7
Program modality		
Methadone maintenance	38	26.6
Therapeutic community	32	22.4
Outpatient drug free	29	20.3
Detoxification	29	20.3
Other modality	15	10.5
When conducted		
1960s	17	11.9
1970s	42	29.4
1980s	30	21.0
1990s	37	25.9
Missing	17	11.9
Primary funding source		
Federal	75	52.4
Other	23	16.1
Missing	45	31.5
Country		
United States	141	98.6
Canada	2	1.4

Number of Subjects	Single	Tx-Comp[b]
Mean	250.9	286.6
Min/max	8/3440	16/2544
Range	3432	2528
Median	122.0	194.5
SD	402.2	468.1

[a] Treatment–comparison design only.
[b] *N* for treatment and comparison groups combined.

studies. Both types of design have a few studies with very large samples, which resulted in a highly skewed distribution of sample size. The median number of subjects was 122 in the single-group studies and 195 in the treatment–comparison-group studies.

	DRUG USE		CRIME		
	Treatment-Comparison (N=28)	Single Group (N=102)	Treatment-Comparison (N=17)	Single Group (N=42)	
Stem	Leaf				
2.7		9			
2.6					
2.5		3			2.5
2.4		0 2 4 9			
2.3					
2.2		1 1 4 4			
2.1		1 6 6			
2.0		7 9			2.0
1.9		5		1	
1.8		2 7 8		0	
1.7		3 3 5 6 9			
1.6		5 5 7			
1.5		1 1 2 3 5 7			1.5
1.4		2 5 9 9		8	
1.3		4 4 4 6		8	
1.2		2 2 2 3 6 8 8 8		6	
1.1	4	0 0 1 1 1 3 6 6 8		0 5 5 6 6 7 7	
1.0	6	0 0 2 2 2 4 4 6 8 8 9 9	8	6	1.0
0.9	1 5	0 3 5		1 5 8	
0.8	1 5	6 7 7 8	7	0 2 4 6	
0.7		0 3 3 3 6 6 9	2	1 3 9	
0.6	1	1 1 4 4 5 6 7		4	
0.5		3 5 6		0 0 4 7 9	0.5
0.4	1 7	3 5 5 6	0	6	
0.3	4 4		0 2 5	2 6	
0.2	2 2 2 5 7 7	6 9	8	6 6	
0.1	3 5 6			1 5 6 9	
0.0	6 9	0	0 8	8 8	0.0
-0.0	9 3 3		9 6 3 2 1	5	
-0.1			1		
-0.2			5		
-0.3					
-0.4	8				
-0.5					-0.5
-0.6					
-0.7	6				
-0.8					
-0.9					

Fig. 1. Stem-and-leaf plot of study-level effect sizes from outcome studies of drug user treatment programs.

RESULTS

Overall Effect Size Estimates

Figure 1 displays a stem-and-leaf plot of the effect sizes for drug use and crime outcomes within each type of design. The stem-and-leaf plots provide a graphical representation of the distribution of all study-level effect sizes across studies. 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. Table 4 shows summary statistics for the two outcome variables broken down by design type.[7] What is perhaps most evident from the stem-and-leaf plots is that studies using single-group designs tend to have much higher effect sizes than treatment–comparison-group studies. For drug use, the unweighted means are 0.32 for treatment–comparison-group designs and 1.28 for single-group designs; for crime, the figures are 0.23 and 0.76, respectively.

The elevated effect sizes from single-group designs are not peculiar to this study. They are typically observed in meta-analyses that have been conducted on psychological, educational, and behavioral treatments. In examining 45 meta-analyses that included effect sizes for both types of design, Lipsey and Wilson (1993) found an overall mean of 0.76 for single-group designs and 0.47 for treatment–comparison-group designs. The variation in the magnitude of effect sizes from the two designs is understandable conceptually, given the type of difference that each design compares. The single-group design provides an estimate of mean change over time within a single group, whereas the treatment–comparison group design estimates the mean difference between two groups at the end of or following treatment. The single-group estimate includes both the effect of treatment and the effect of nontreatment factors such as history, maturation, selection, placebo effect, etc. In treatment–comparison-group designs, by contrast, these nontreatment factors are presumably captured in the scores of the comparison group, which are subtracted from the scores of the treatment group, resulting in lower average effect sizes than in single-group designs.

For the outcome variables selected for this analysis, two other factors tend to result in large effect sizes for single-group designs. First, scores for drug use and crime tend to be extreme at pretest. Particularly when measured dichotomously, pretest scores for drug use and crime are often close to 100%. Post-test behavior measured at the end of treatment or soon thereafter is likely to be considerably lower than at pretest and will thus yield large effect sizes when pretest and post-test scores are compared. For instance, if the percentage of subjects using drugs drops from 100% at

Table 4

Summary Statistics of Effect Sizes for Drug Use and Crime Outcomes by Design Type

	Drug Use		Crime	
	Treatment–Comparison	Single Group	Treatment–Comparison	Single Group
Number of studies	28	103	17	42
Unweighted mean	0.32	1.27	0.23	0.76
Standard deviation (SD)	0.54	0.59	0.38	0.47
Maximum	1.14	2.79	1.08	1.92
Median	0.26	1.16	0.08	0.76
Minimum	−0.77	0.00	−0.26	−0.06
Fixed effects weighted mean (95% CI)	0.29 (0.21, 0.36)	1.20 (1.18, 1.21)	0.13 (0.03, 0.22)	0.73 (0.72, 0.75)
Homogeneity test (Q)	96.64***	11495.43***	30.62**	3176.18***
Random-effects weighted mean (95% CI)	0.29 (0.14, 0.45)	1.27 (1.15, 1.38)	0.17 (0.00, 0.33)	0.75 (0.61, 0.90)

** $p < .01$.

*** $p < .0001$.

baseline to 25% at follow-up, the effect size, using the arcsine transformation, is 2.10. Although effect sizes over 1.0 are not typical in meta-analyses (see Lipsey and Wilson, 1993), they have ben reported in at least one other meta-analysis of drug dependence treatment. From a meta-analysis of 11 studies of the effectiveness of MMT, Marsch (1998) reported effect sizes for opiate use ranging from 0.19 to 2.47, with a mean of 0.78; for crime, effect sizes ranged from 0.00 to 2.33, with a mean of 0.54. Second, as can be seen in Fig. 1, for single-group designs, it is likely that the direction of effect for drug use and crime outcomes will nearly always be positive. By contrast, the effect sizes for some studies using a treatment and comparison group may be negative, which tends to reduce the overall average. Because at present there is no generally accepted method to combine effect sizes from the two types of design in a way that would make them comparable, we have treated studies using each type of design separately.

As noted above, unweighted effect size averages give equal weight to each study-level effect size. But because studies with larger sample sizes are more precise and stable, effect sizes are typically averaged from individual weighted effect sizes, where the weighting factor is the inverse of the variance. In Table 4, it can be seen that the weighted means, using a fixed effects model, are somewhat smaller than the unweighted means. Because none of the confidence intervals include zero, the null hypothesis that the population effect size is zero can be rejected. Table 4 also shows the results of the homogeneity test of the effect size estimates, which is intended to answer the question: Do all of the effect sizes from the sample of studies represent or estimate, within sampling error, the same population effect size? As can be seen, the homogeneity tests for each of the outcome/design categories are significant. If the Q statistic is not significant, then the estimates of effect size among the studies are considered to differ among themselves by unsystematic sampling error only. If the Q statistic is significant, as is the case here, then the effect sizes from the studies are more heterogeneous than would be expected from sampling error alone and suggests that the effect sizes from each of the categories 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 4, the random effects weighted means for drug use outcomes are 0.29 for treatment–comparison designs and 1.28 for single-group designs; for crime outcome, the figures are 0.17 and 0.75, respectively.

Correlations Between Program Factors and Effect Sizes

Tables 5 and 6 present the results of a weighted correlation analysis of the relationship between the selected program variables and average effect

Table 5

Weighted Correlation of Program Variables with Effect Size for Drug Use Outcomes by Design Type and Modality (Number of Studies in Parentheses)[a]

	Treatment–Comparison				Single Group				
	MMT[b]	TC[c]	ODF[d]	Other[e]	MMT	TC	ODF	DETOX	Other
Weighted effect size	0.49	0.09	0.28	0.24	1.48	1.06	1.12	0.91	1.36
	(8)	(8)	(8)	(4)	(22)	(21)	(20)	(27)	(13)
Program variable									
Decade of treatment	0.47	−0.27	−0.02	0.54	−0.38*	0.25	0.24	0.05	−0.42
	(6)	(8)	(6)	(3)	(21)	(20)	(16)	(21)	(12)
Role of researcher	0.13	0.07	—	—	−0.24	−0.03	0.28	−0.03	0.26
	(7)	(8)	(4)	(3)	(16)	(19)	(17)	(19)	(11)
Maturity of program	0.06	−0.33	0.33	0.52	0.13	−0.11	−0.11	−0.30	0.15
	(7)	(7)	(6)	(4)	(20)	(19)	(15)	(21)	(7)
Counselor/client ratio	—	—	—	—	0.18	0.35	—	—	—
	(0)	(2)	(1)	(1)	(5)	(4)	(1)	(1)	(1)
Average methadone dose (mg)	−0.05	—	—	—	0.89**	—	—	0.81**	—
	(5)	(0)	(0)	(1)	(5)	(0)	(0)	(9)	(0)
Implementation of treatment	−0.88*	0.46	0.36	−0.42	−0.49*	0.85**	0.09	0.18	0.33
	(5)	(7)	(8)	(3)	(14)	(11)	(14)	(18)	(7)
Mean no. of weeks in treatment	0.74	0.54	−0.15	−1.00	0.57*	0.34	0.23	−0.16	0.18
	(3)	(6)	(5)	(2)	(11)	(14)	(16)	(15)	(10)
Mean no. of hours of contact time over treatment	—	0.54	−0.02	—	−0.53	0.33	0.50*	0.38	0.12
	(1)	(6)	(4)	(1)	(5)	(14)	(13)	(10)	(8)
Intensity of treatment contact	−0.57	—	-	−0.03	0.09	−0.12	0.01	0.27	—
	(8)	(8)	(6)	(4)	(16)	(20)	(17)	(23)	(10)

* $p < .10$.

** $p < .05$.

[a] Values for some of the correlations are missing either (1) because no study reported data on the specific program variable or (2) because there was no variation among the scores for the program variable.

[b] MMT = Methadone maintenance.

[c] TC = Therapeutic community.

[d] ODF = Outpatient drug free.

[e] Because of the small number of studies of detoxification using a treatment–comparison design, they have been combined here with the Other category.

Table 6

Weighted Correlation of Program Variables with Effect Size for Crime Outcomes by Design Type and Modality (Number of Studies in Parentheses)[a]

	Treatment–Comparison				Single Group				
	MMT[b]	TC[c]	ODF[d]	Other[e]	MMT	TC	ODF	DETOX	Other
Weighted effect size	0.17	0.04	0.22	0.16	0.80	0.84	0.62	0.60	0.52
	(3)	(7)	(4)	(3)	(16)	(12)	(8)	(3)	(3)
Program variable									
Decade of treatment	−0.81	−0.45	0.36	—	−0.09	−0.63*	−0.62	−0.31	0.53
	(3)	(7)	(3)	(2)	(16)	(12)	(7)	(3)	(3)
Role of researcher	0.81	0.03	—	—	0.45*	—	—	—	—
	(3)	(7)	(3)	(3)	(15)	(11)	(7)	(3)	(2)
Maturity of program	—	−0.73*	−0.99*	—	0.34	—	−0.49	—	—
	(2)	(7)	(3)	(3)	(15)	(11)	(7)	(3)	(1)
Counselor/client ratio	—	—	—	—	−0.80	0.98**	—	—	—
	(0)	(2)	(0)	(0)	(5)	(5)	(0)	(0)	(0)
Average methadone dose (mg)	—	—	—	—	−0.16	—	—	—	—
	(1)	(0)	(0)	(1)	(4)	(1)	(0)	(0)	(0)
Implementation of treatment	−1.00	−0.02	—	−1.00	0.06	0.67	0.97**	—	—
	(2)	(7)	(4)	(2)	(12)	(6)	(6)	(0)	(1)
Mean no. of weeks in treatment	—	0.77*	0.48	—	−0.28	0.13	−0.35	—	1.00
	(0)	(6)	(3)	(1)	(9)	(7)	(6)	(2)	(2)
Mean no. of hours of contact time over treatment	—	0.77*	—	—	−0.70	0.31	−0.73	—	1.00
	(0)	(6)	(1)	(1)	(3)	(7)	(4)	(2)	(2)
Intensity of treatment	—	—	—	0.46	0.21	0.50*	0.44	−0.08	—
	(3)	(7)	(3)	(3)	(13)	(12)	(6)	(3)	(2)

*$p < 0.10$.
**$p < 0.05$.
[a] Values for some of the correlations are missing either (1) because no study reported data on the specific program variable or (2) because there was no variation among the scores for the program variable.
[b] MMT = Methadone maintenance.
[c] TC = Therapeutic community.
[d] ODF = Outpatient drug free.
[e] Because of the small number of studies of detoxification using a treatment–comparison design, they have been combined here with the Other category.

size estimates for drug use (Table 5) and for crime (Table 6) outcomes. Just as each effect size was weighted to take into account the varying sample sizes of each study, so also were the values of each program factor weighted by the inverse of the variance. For each table, the results are broken down by type of design and by modality. The effect sizes for each of the design/modality categories are shown across the top of each table. Although the average effect sizes for the TC programs evaluated in treatment–comparison-group designs are low relative to the effect sizes of the other modalities, the result includes two studies of TC programs with negative effect sizes (see Fig. 1).

For the correlation analysis, when the program variable was continuous, Pearson's correlation coefficient was used. Variables measured on an ordinal scale were dichotomized and analyzed using point biserial correlation. Because so few studies evaluated detoxification using a treatment–comparison design, in both tables DETOX studies were combined with "other" for the treatment–comparison designs; they were kept separate for the single-group designs because of the larger number of available studies. In presenting the results, we have been careful to indicate that the unit of analysis is studies of programs and not programs themselves. Thus, we use such phrasing as "MMT studies" or "studies of TCs" to emphasize that the meta-analysis results are based on available evaluation studies of programs, which may or may not be representative of current or former treatment programs.

Drug Use Outcomes

As can be seen in Table 5, even when the significance level is set at 0.10, most of the correlations within modalities for drug use are not statistically significant. In some cases, too few studies were available for analysis, or there was no variation in the program variable. The decade in which the study was conducted was negatively correlated with effect size for drug use in MMT studies; more recent studies had lower effect sizes. Not surprisingly, average methadone dose was strongly correlated with outcome, although only for single-group studies. Mixed results were found for the relationship of effect size to quality of program implementation (dichotomized as well implemented versus poorly implemented). The average effect size for drug use outcomes was negatively correlated with treatment implementation for MMT studies, but was positively correlated for TC studies. The mean number of weeks that clients participated in treatment was significantly correlated with effect size for MMT in the single-group studies. Another measure of the degree of treatment exposure is the mean number of hours of

contact over the course of treatment, which was found to be significantly related to effect size for studies of ODF programs.

Crime Outcomes

Correlations between program variables and average effect sizes for crime outcomes across design types and modalities are shown in Table 6. As with the drug use outcomes, few of the correlations are statistically significant. Studies conducted more recently had lower effect sizes, mainly among studies of TC programs. A high degree of researcher involvement in the treatment under study (design and/or delivery of treatment) was associated with larger effect sizes in MMT studies. The means for the other modalities were all zero, indicating little or no researcher involvement in treatment delivery. At least in treatment–comparison-group studies, better crime outcomes were found in more mature programs. Most of the included studies did not report information on the counselor caseload. In the five TC studies that reported data on the counselor/client ratio, programs with a large number of clients per counselor had larger effect sizes, although the actual ratios were very low (1 : 6 or less). Well-implemented programs were strongly associated with crime outcome in ODF studies. For TC programs, studies in which clients had greater treatment exposure had higher effect sizes. Finally, a measure of the intensity of treatment contact (how involving, emotional, memorable, or focused each contact was) was positively correlated with crime outcome in the studies of TC programs.

SUMMARY

Previous research has shown that treatment outcomes vary among programs depending on a variety of factors, including the characteristics of programs. If certain program characteristics—and not others—make an important difference in how clients perform during and after treatment, then government agencies, professional organizations, and treatment providers are in a better position to attempt to improve the effectiveness of drug user treatment by focusing training materials, standards, and monitoring systems on those program factors that have an important impact on outcomes. In the absence of such knowledge, efforts to improve treatment must rely on a "scatter shot" strategy. The findings from the present study identified several program factors from a large set of treatment effectiveness studies that may contribute to the development of a more evidence-based foundation for improving treatment outcomes.

Discussion of Findings

The findings presented here provide evidence that some program factors are associated with the average effect size for drug use and crime outcomes, although in some cases the direction of the relationship is contrary to what might be expected. The small number of available studies for some of the cells makes it unlikely that even relatively large coefficients would be significant. This, as well as the lack of variability among some of the program variables, limits what can be derived from these data about the impact of program factors on outcomes.

Decade of Treatment

We selected decade of treatment as a proxy for changes in program characteristics or program quality over time. For MMT and TCs, studies conducted in the 1960s and 1970s tended to have larger effect sizes than studies conducted in the 1980s and 1990s. It could be that treatment programs have declined in quality or have experienced reductions in services over the past three decades, leading to the lower effect sizes found in the more recent outcome studies. With regard to services, national drug treatment outcomes studies (the Treatment Outcome Prospective Study [TOPS] and the Drug Abuse Treatment Outcome Study [DATOS] indicate a decline in the number of services available between the early 1980s and the early 1990s (Etheridge et al., 1997). However, the negative association could also result from changes in client characteristics, including more clients who use multiple drugs, and who therefore have more problems, entering treatment in more recent years, more entrenched drug use patterns in "hard-core" drug users, and changes in funding priorities directed to special populations (e.g., persons with HIV, pregnant women, persons with comorbid psychiatric conditions)—all of which might make clients admitted in recent decades more difficult to treat (Craddock et al., 1997).

Treatment Implementation

A number of researchers have discussed treatment implementation, often in terms of program integrity, as an important (and generally neglected) issue in program evaluation and in moderating the effects of treatment (Gresham et al., 1993; Salend, 1984; Yeaton and Sechrest, 1981). Some evidence for the importance of implementation issues is seen in higher effect sizes found in MMT studies where the researcher was involved in the design and/or delivery of treatment. Compared with studies in which the researcher is an outside evaluator and has no involvement in

the delivery of treatment, researcher-involved treatments tend to follow standard protocols, involve specially trained staff, and provide close monitoring of treatment delivery—all of which are likely to improve treatment implementation and to be associated with stronger treatment effects.[8] Other findings, however, were contradictory. Treatment implementation was positively correlated with effect size in TC and ODF studies; that is, studies in which the treatment was judged to be well implemented tended to have high effect sizes. By contrast, the correlation for the MMT studies was negative: studies of programs rated as well implemented tended to have low effect sizes. This negative association is not what would be expected, but at this point there is no obvious methodological or substantive explanation for this finding.

Treatment Exposure

The longer duration of time that clients participate in treatment has consistently been found to result in better outcomes following treatment (Hubbard et al., 1989; Simpson, 1981; Simpson et al., 1997). In this analysis, treatment exposure was measured in two ways: the mean number of weeks that subjects actually participated in treatment (not the expected length of treatment), and the mean number of hours that subjects were in direct contact with the treatment. The first variable provides a measure of duration, the second of intensity. Significant positive correlations between one or the other measure of treatment exposure and effect size were found for MMT, TC, and ODF. The impact of treatment, however, is not simply a function of duration. For example, clients in MMT tend to stay longer than clients in TCs (Simpson et al., 1997), but the *intensity* of treatment measured in contact hours is greater in TCs than in MMT programs.

Methadone Dosage

Across the studies of methadone programs included in this analysis, the average methadone dose was 44.6 mg (range, 23–80 mg). For studies using a single-group design, average methadone dose was positively associated with effect size for drug use outcomes in both MMT and DETOX programs. These findings support other studies (not duplicated in this meta-analysis) that have found higher methadone dosages to be associated with better outcomes (Caplehorn and Bell, 1993; D'Aunno and Vaughn, 1992; Ling et al., 1976; Strain et al., 1999). For treatment–comparison-group studies, however, dosage was not significantly correlated with drug use outcome. Further analysis is needed to examine this apparent inconsistency. Given the relatively low dosages reported in the studies included here, little can be

said about the effects on outcome of methadone administered in dosages in excess of 80 mg, which has been recommended by some researchers and clinicians.

Methodological Issues and Limitations

Publication Bias

As a method for synthesizing findings from multiple studies, meta-analysis must deal with the potential effects of publication bias. Publication bias arises from the tendency of researchers to submit for publication, and of editors to accept for publication, articles that report statistically significant results, leaving an unknown number of studies that found nonsignificant (or significantly negative) results in the "file drawer" (Rosenthal, 1979; Easterbrook et al., 1991). Potential publication bias in the drug user treatment meta-analysis was handled in two ways. First, during the literature search, we attempted to identify as many unpublished studies as possible by searching databases that include such studies (e.g., Dissertation Abstracts, National Technical Information Service), by contacting funding agencies, and by asking drug dependence researchers to identify unpublished studies. As seen in Table 3, 11 of the studies included in this analysis were based on unpublished papers.

Second, the problem can be addressed statistically through the "fail-safe *N*" technique developed by Rosenthal (1979), which provides an estimate of the minimum number of unpublished or unretrieved studies with nonsignificant results that would need to exist in order to bring the significance level of a set of studies down to a "just significant" level. For the current set of studies, the analysis was conducted for each design/outcome combination. The results are shown in Table 7. Although there are no established guidelines for what constitutes a critical number of unpublished or unretrieved

Table 7

Results of Rosenthal's Fail-Safe N Test for Publication Bias

Design	Outcome	k[a]	*X*[b]
Single group	Substance use	102	1,030,601
Single group	Criminal activity	42	74,755
Treatment–comparison group	Substance use	28	344
Treatment–comparison group	Criminal activity	17	30

[a] Number of studies.
[b] Fail-safe *N*.

studies, the results for this dataset indicate that only for crime outcomes in treatment–comparison-group studies is it plausible that the overall results could be due to publication bias.

Study Design Differences

Differences in study design complicate the analysis of the association between program factors and client outcomes. Average effect sizes from single-group designs almost surely overstate the magnitude of the impact of treatment itself, but because the effect sizes are based on pre- and post-test data within single programs, variation in effect sizes across studies are more likely to reflect program differences. By contrast, although treatment-comparison group designs are more methodologically sound and more likely to provide a more accurate indication of program effectiveness, effect sizes from such designs are based on between-group differences at post-test. Because comparison groups are likely to have received at least some treatment, and because both treatment and comparison groups may be subject to many of the same general program influences, variation in average effect sizes due to program characteristics may be partially obscured. The lack of comparability between the ways effect sizes in the two designs are calculated may at least partially explain why we did not find more consistent findings across study designs. One possible way to address this issue would be to compute pre- and post-test effect sizes for treatment groups that were part of an experimental design and combine them in an analysis with the single-group studies. However, because this approach sacrifices the methodological advantage of the treatment–comparison-group design, we did not do so here.

Study Variability

Caution should be used in generalizing the results from this analysis of drug dependence treatment interventions to the treatment field as a whole. There is considerable variability in methodological features, subject characteristics, and treatment characteristics across the studies even after variables were examined by modality, design type, and type of outcome. We attempted to avoid the "apples and oranges" criticism (Gallo, 1978; Presby, 1978) that has sometimes been levied against meta-analysis by averaging effect sizes within clinically relevant subcategories defined by type of outcome, type of treatment, and type of design. This, of course, does not exhaust the ways in which the full set of studies may differ on important dimensions. In particular, for this analysis, we did not take into account differences among subjects, except indirectly in the sense that clients in

MMT programs are more homogeneous in terms of their drug use than are subjects in other types of treatment. Apart from the general type of study design (single group, treatment–comparison group), other differences in methodology are also likely to be important in accounting for variability in effect size. For the eligible drug user treatment studies, methodological features were coded, permitting an examination of the impact of differences in study quality and design on average effect sizes in future analyses of this dataset.

Missing Data

Another problem in this study is the impact of missing values for both the dependent and independent variables. A number of studies that were located for the meta-analysis were not included in this analysis because of the inability to calculate effect sizes or to determine sample size. Missing data for independent variables was also a problem. The coding of studies in a meta-analysis must rely on "found data," that is, information that is included in the eligible studies. Although each study report is "asked" the same set of questions, the number of questions about which "answers" are available varies greatly across studies. Given the space limitations of the venues for reporting results (mainly articles, book chapters, and conference abstracts), the lack of information available in the reports to code program-level variables is understandable. The result, however, is that many variables have a high percentage of missing values. The degree to which conclusions can be drawn about the differential effects of program factors on treatment outcomes is limited by the large proportion of missing values for many of the coded variables. Even large correlations may not be significant if the number of observations is small. Future analyses will include an imputation strategy for missing values in order to increase the number of usable cases.

CONCLUSIONS

This study of the impact of program factors on treatment outcomes highlights some of the strengths and limitations of meta-analysis to answer questions about drug dependence treatment. The method provides the opportunity to synthesize findings from a large body of studies using a common set of questions to extract information. The aggregated effect sizes provide an estimate of the magnitude and direction of treatment effects overall and for specific modalities. These quantitative estimates are likely

to be more valid and useful as indicators of treatment effectiveness than statements based on a "box score" of significant and nonsignificant findings from individual studies. Finally, the coding of study characteristics provides the opportunity to examine whether certain variables moderate the variation found among effect sizes across studies.

The findings are limited, however, by a number of factors, including the heterogeneity of the drug user treatment literature in terms of modalities, research designs, and multiple outcomes; the failure of some studies to provide sufficient quantitative data to calculate effect sizes and/or variances; and the high percentage of missing data for variables of interest. Some of these problems are inherent in the nature of the field and reporting practices; others can be addressed with multivariate techniques and missing data strategies. The benefits of meta-analysis in understanding the contribution of program factors to treatment outcomes cannot be fully realized as long as study reports vary in the type and amount of information they provide about the programs they are evaluating. One strategy to address the lack of information on important variables of interest would be to contact researchers whose studies are selected for a meta-analysis and ask them to supply the missing data. Although the success of this strategy would largely depend on the willingness of researchers to cooperate and on the accessibility of the needed data, and would likely introduce some response bias into the final dataset, it would be an improvement on the present situation.

Overall, the results of this analysis of drug user treatment outcome studies indicate that the effects are small to moderate when based on treatment–comparison-group designs. There is also wide variability among programs in their effectiveness, and although treatment is generally effective overall, this does not necessarily imply that any given treatment program is effective. Quality assurance is an important issue that continues to be relevant for all treatment programs.

In terms of program factors, within a given modality, better drug use and crime outcomes are likely to be found in those programs that have been established for several years, that closely monitor the integrity of the treatment protocol, that have a lower ratio of counselors to clients, that have higher retention rates, that offer more hours of contact time, and (for MMT and DETOX programs) that provide higher dosages of methadone. These are dynamic features of programs that are amenable to change and should be given greater attention in quality improvement efforts and agency technical assistance. Additional analysis is needed to confirm these results, to control for other variables, to address missing data issues, and to extend the findings to a broader array of program factors.

ACKNOWLEDGMENTS

Support for this study came from Grant R01DA09151 from the National Institute on Drug Abuse. The views expressed in the paper, however, do not necessarily represent the position of NIDA or of the 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 figures. 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.

GLOSSARY

Effect size: A generic term that refers to the magnitude of an effect or more generally the size of the relation between two variables. Special cases include standardized mean difference, correlation coefficient, odds ratio, or the raw mean difference.

Fail-safe N: An estimate of the number of unreported studies averaging magnitude zero that would have to exist in order to bring the significance level for a set of studies down to a "just significant" level.

File-drawer problem: A situation related to publication bias in which findings from a study are not reported or published (i.e., remain in the "file drawer") when results are not significant.

Homogeneity test: A statistical test to determine whether effect size estimates exhibit greater variability than would be expected if their corresponding effect size parameters were identical.

Meta-analysis: The statistical analysis of a collection of results from individual studies for the purpose of integrating the findings on a particular topic using quantitative methods.

Modality: One of several approaches for treating drug problems, traditionally identified as methadone maintenance, therapeutic community, outpatient drug free, and detoxification.

Moderator: Any factor that influences the size of a particular relationship and is itself not a consequence of the relationship.

Single-group design: A study involving one group only that receives the treatment being evaluated.

Treatment–comparison design: A study comparing two groups in which one group receives the treatment being evaluated and the other does not.

The condition received by the comparison group can be either no treatment, minimal treatment, or an alternative treatment.

NOTES

1. The conduct and findings of a study may be reported in one or more documents, either published or unpublished. Thus, operationally, for coding purpose, the data on a study are derived from all retrieved documents that provide information about the study.
2. The term "comparison" is used throughout this paper rather than "control" because it is more inclusive of the types of groups that are compared with the treatment group. Although some studies that used randomization did have a control group, many others used matched comparison groups, convenience samples, or intact groups.
3. For treatment–comparison-group designs, the correction coefficient applied to effect sizes based on means and standard deviations (or their equivalent test statistic) is $1 - (3/(4n_t + 4n_c - 9))$, where n_t is the sample size for the treatment group and n_c is the sample size for the comparison group. For single-group designs, the coefficient is $1 - (3/(4/n - 5))$, where n is the sample size.
4. The formulas for the variance of an effect size from single-group designs (assuming the correlation between pretest and post-test scores is 0.8) are

$$v = \frac{2 \cdot (1 - 0.8)}{n} + \frac{d^2}{2 \cdot n}$$

for effect size from means and standard deviations, where d is the effect size and n is the sample size, and

$$v = \frac{2 \cdot (1 - 0.8)}{n}$$

for effect size from proportions.

The formulas for the variance of an effect size from treatment–comparison designs are

$$v = \frac{n_t + n_c}{n_t \cdot n_c} + \frac{d^2}{2 \cdot (n_t + n_c)}$$

for effect size from means and standard deviations, 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 companion group, and

$$v = \frac{n_t + n_c}{n_t \cdot n_c}$$

for effect size from proportions.

5. These values fall at approximately the 75% percentile of the sample size distributions for the studies in each of the designs, using the entire dataset of 293 studies.
6. Because MMT is directed specifically at heroin and other opiates use, the drug outcome variable for methadone programs was confined to measures of opiate use. Thus, included in this category of excluded studies are six studies of MMT programs that did not have an opiate-related outcome variable.
7. The number of studies in Table 4 does not equal the number of studies in Table 3 because some studies report both drug use and crime outcome variables.
8. Expectancy effects on the part of researchers may also have some influence on the reported outcomes. Evidence for this possibility from the coded studies is mixed. For all studies with

nonmissing data on a question about researcher allegiance, the average effect size for studies in which the researcher was judged to be favorably disposed toward the treatment being evaluated (ES = 1.14; k = 67; 95% confidence interval [CI] = 0.98, 1.29) was similar to that for studies in which the researcher was judged to be neutral or unfavorably disposed (ES = 1.04; k = 63; 95% CI = 0.86, 1.21). Results were similar for single-group designs, but for treatment-comparison-group designs, the average effect size was much larger in those studies where the researcher was judged to be favorable toward the treatment (ES = 0.46; k = 18; 95% CI = 0.30, 0.63) than when judged to be neutral or unfavorable (ES = 0.02; k = 10; 95% CI = −0.19, 0.24).

REFERENCES

ALLISON, M., and HUBBARD, R. L. (1985). Drug abuse treatment process: A review of the literature. *Int. J. Addict.* 20(9): 1321–1345.

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

APSLER, R., and 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, pp. 58–81). Rockville, MD: National Institute on Drug Abuse.

BALL, J. C., and ROSS, A. (1991). *The Effectiveness of Methadone Maintenance Treatment: Patients, Programs, Services, and Outcomes.* New York, NY: Springer-Verlag.

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

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

CAPLEHORN, J. R., and BELL, J. (1991). Methadone dosage and retention of patients in maintenance treatment. *Med. J. Austr.* 154: 195–199.

CAPLEHORN, J. R., BELL, J., KLEINBAUM, D. G., and GEBSKI, V. J. (1993). Methadone dose and heroin use during maintenance treatment. *Addiction* 88(1): 119–124.

COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, (2nd ed.). New York: Academic Press.

COOPER, H. M. (1989). *Integrating Research: A Guide for Literature Reviews* (2nd ed). Newbury Park, CA: Sage Publications.

COOPER, H. M., and HEDGES, L. V. (Eds.) (1994). *The Handbook of Research Synthesis.* New York: Russell Sage Foundation.

COOPER, J. R., ALTMAN, F., BROWN, B. S., and CZECHOWICZ, D. (Eds.) (1983). *Research on the Treatment of Narcotics Addiction: State of the Art* (NIDA Treatment Research Monograph). Rockville, MD: National Institute on Drug Abuse.

CRADDOCK, S. G., ROUNDS-BRYANT, J. L., FLYNN, P. M., and HUBBARD, R. L. (1997). Characteristics and pretreatment behaviors of clients entering drug abuse treatment: 1969 to 1993. *Am. J. Drug Alcohol Abuse* 23(1): 43–59.

D'AUNNO, T., and VAUGHN, T. E. (1992). Variations in methadone treatment practice: Results from a national survey. *JAMA* 267(2; January 8): 253–258.

EASTERBROOK, P. J., BERLIN, J. A., GOPALAN, R., and MATTHEWS, D. R. (1991). Publication bias in clinical research. *Lancet* 337: 867–872.

ETHERIDGE, R. M., HUBBARD, R. L., ANDERSON, J., CRADDOCK, S. G., and FLYNN, P. M. (1997). Treatment structure and program services in the Drug Abuse Treatment Outcome Study (DATOS). *Psychol. Addict. Behav.* 11(4): 244–260.

GALLO, P. (1978). Meta-analysis: Mixed meta-phor. *Am. Psychol.* 33(5): 515–517.

GERSTEIN, D. R., and HARWOOD, H. J. (1990). *Treating Drug Problems*, Vol. 1. Washington, DC: National Academy Press.

GLASS, G. V., McGAW, B., and SMITH, M. L. (1981). *Meta-Analysis in Social Research.* Beverly Hills: Sage Publications.

GRESHAM, F. M., GANSLE, K. A., NOELL, G. N., COHEN, S., and ROSENBLUM, S. (1993). Treatment integrity of school-based behavioral intervention studies: 1980–1990. *School Psychol. Rev.* 22(2): 254–272.

HEDGES, L. V., and OLKIN, I. (1985). *Statistical Methods for Meta-Analysis.* Orlando, FL: Academic Press, Inc.

HUBBARD, R. L., MARSDEN, M. E., RACHAL, J. V., HARWOOD, H. J., CAVANAUGH, E. R., and GINZBURG, H. M. (1989). *Drug Abuse Treatment: A National Study of Effectiveness.* Chapel Hill: University of North Carolina Press.

JOE, G. W., SINGH, B. K., LEHMAN, W., SIMPSON, D. D., and SELLS, S. B. (1983). Agency differences in post-treatment outcomes: A follow-up of drug abuse treatment clients. *Int. J. Addict.* 18(4): 523–538.

KLEBER, H. D. (1989). Treatment of drug dependence: What works. *Int. Rev. Psychiatry* 1: 81–100.

LING, W., CHARUVASTRA, C., KAIM, S. C., and KLETT, C. J. (1976). Methadyl acetate and methadone maintenance treatments for heroin addicts. *Arch. Gen. Psychiatry* 33: 709–720.

LIPSEY, M. W., and WILSON, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *Am. Psychol.* 48(12): 1181–1209.

McCAUGHRIN, W. C., and PRICE, R. H. (1992). Effective outpatient drug treatment organizations. Program features and selection effects. *Int. J. Addict.* 27(11): 1335–1358.

McGLOTHLIN, W. H., and ANGLIN, M. D. (1981). Long-term follow-up of clients of high- and low-dose methadone programs. *Arch. Gen. Psychiatry* 38: 1055–1063.

McLELLAN, A. T., ALTERMAN, A. I., CACCIOLA, J., METZGER, D., and O'BRIEN, C. P. (1992). A new measure of substance abuse treatment: Initial studies of the Treatment Services Review. *J. Nerv. Ment. Dis.* 180(2): 101–110.

McLELLAN, A. T., GRISSOM, G. R., BRILL, P., and DURELL, J. (1993). Private substance abuse treatments: Are some programs more effective than others? *J. Subst. Abuse Treat.* 10(3): 243–254.

McLELLAN, A. T., O'BRIEN, C. P., METZGER, D., ALTERMAN, A. I., CORNISH, J., and URSCHEL, H. (1992). How effective is substance abuse treatment: Compared to what? In C. P. O'Brien and J. H. Jaffe (Eds.), *Addictive States* (pp. 231–252). New York: Raven Press.

MAGURA, S., NWAKEZE, P. C., KANG, S.-Y., and DEMSKY, S. (1999). Program quality effects on patient outcomes during methadone maintenance: A study of 17 clinics. *Subst. Use Misuse* 34(9): 1299–1327.

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.

MOOS, R. H., PETIT, B., and GRUBER, V. A. (1995). Characteristics and outcomes of three models of community residential care for substance abuse patients. *J. Subst. Abuse* 7: 99–116.

MOOS, R. H., FINNEY, J. W., OUIMETTE, P. C., and SUCHINSKY, R. T. (1999). A comparative evaluation of substance abuse treatment: I. Treatment orientation, amount of care and 1-year outcomes. *Alcoholism Clin. Exper. Res.* 23(3): 529–536.

MOOS, R. H., and MOOS, B. S. (1998). The staff workplace and the quality and outcome of substance abuse treatment. *J. Stud. Alcohol* 59(1): 43–51.

MOOS, R. H., KING, M. J., BURNETT, E. B., and ANDRASSY, J. M. (1997). Community residential program policies, services, and treatment orientations influence patients' participation in treatment. *J. Subst. Abuse* 9: 171–187.

MOOS, R. H., MERTENS, J. R., and BRENNAN, P. L. (1995). Program characteristics and readmission among older substance abuse patients: Comparisons with middle-aged and younger patients. *J. Ment. Health Admin.* 22(4): 332–345.

POLINSKY, M. L., HSER, Y., ANGLIN, M. D., and MAGLIONE, M. A. (1998). Drug-user treatment programs in a large metropolitan area. *Subst. Use Misuse* 33(8): 1739–1761.

PRESBY, S. (1978). Overly broad categories obscure important differences between therapies. *Am. Psychol.* 33(5): 514–515.

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 ed.). Newbury Park, CA: Sage Publications.

SALEND, S. J. (1984). Therapy outcome research: Threats to treatment integrity. *Behav. Modif.* 8(2): 211–222.

SIMPSON, D. D. (1981). Treatment for drug abuse: Follow-up outcomes and length of time spent. *Arch. Gen. Psychiatry* 38: 875–880.

SIMPSON, D. D., JOE, G. W., and BROWN, B. S. (1997). Treatment retention and follow-up outcomes in the Drug Abuse Treatment Outcome Study (DATOS). *Psychol. Addict. Behav.* 11(4): 294–307.

SISK, J. E., HATZIANDREU, E. J., and HUGHES, R. (1990). *The Effectiveness of Drug Abuse Treatment: Implications for Controlling AIDS/HIV Infection* (AIDS-Related Issues Background Paper 6). Washington, DC: Office of Technology Assessment, U.S. Congress.

STRAIN, E. C., BIGELOW, G. E., LIEBSON, I., and STITZER, M. L. (1999). Moderate- vs. high-dose methadone in the treatment of opioid dependence: A randomized trial. *JAMA* 281:1000–1005.

TIMKO, C. (1996). Physical characteristics of residential psychiatric and substance abuse programs: Organizational determinants and patient outcomes. *Am. J. Commun. Psychol.* 24(1): 173–192.

YEATON, W. H., and SECHREST, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *J. Consult. Clin. Psychol.* 49(2): 156–167.

APPENDIX

Eligibility Criteria for Studies

1. The study must have a primary focus on treatment for illicit drug dependence.
2. The treatment or intervention must be intended for adults (age 18 years and older), or adults must constitute at least 75% of the study sample.

3. The treatment or intervention must be directed toward changing the drug use and/or related behaviors or attitudes of the subjects. Typical treatment approaches included are MMT, TCs, ODFs, chemical dependency programs, and specific treatment techniques. Excluded are clinical studies of antiaddiction medications that have not been approved by the FDA for general use, studies conducted within the criminal 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 impaired professionals and employee assistance programs.
4. The date of the document reporting on the study must be 1965 or later. The recent cutoff point for documents is December 1996.
5. The document reporting the study can be either published or unpublished.
6. The document(s) reporting on the study must be in English.
7. The setting of the study must be the United States (50 states and District of Columbia) or Canada.
8. The study must include quantitative outcome variables in which one or more treatment conditions is compared with one or more comparison conditions. The comparison condition may be respondents' baseline measurements, as is the case in a single-group pretest/post-test study. The comparison condition may also be "no treatment," "typical or usual treatment," "placebo treatment," or other condition in which the intention is not to produce change (or is intended to produce only minimal change) in the outcome variables for the comparison group. Studies that compare two types of treatment against each other (e.g., MMT vs. TC) are excluded.
9. Random assignment of subjects to treatment and comparison conditions is not necessary, but the study report must include information that will allow an assessment of the comparability of the treatment and comparison groups at baseline. That is, the researcher either used matching in constructing the groups or reported background characteristics that would permit a judgment on initial group equivalence.

THE AUTHORS

Michael Prendergast, Ph.D., is Associate Research Historian at the UCLA Drug Abuse Research Center and is Principal Investigator of the NIDA-funded meta-analysis of treatment effectiveness. His research interests are treatment effectiveness, criminal justice treatment, historical aspects of drug use and treatment, and drug user treatment policy.

Deborah Podus, Ph.D., is Assistant Research Sociologist at the UCLA Drug Abuse Research Center and is project director of the meta-analysis on treatment effectiveness study funded by the National Institute on Drug Abuse. Her research interests include treatment effectiveness, drug user treatment policy, and drug use issues related to welfare reform.

Eunice Chang, Ph.D., is Senior Statistician at the PacifiCare Health Systems. Her statistical expertise focuses on meta-analysis, longitudinal analysis, hierarchical linear modeling, and structure equation modeling.