Contingency management for treatment of substance use disorders: a meta-analysis

Michael Prendergast[1], Deborah Podus[1], John Finney[2], Lisa Greenwell[1] & John Roll[3]

Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA,[1] Center for Health Care Evaluation, Veterans Affairs Palo Alto Health Care System, CA, USA[2] and Washington State University, Spokane, USA[3]

ABSTRACT

Aims   To examine the effectiveness of contingency management (CM) techniques in treating substance use disorders (i.e. illicit drugs, alcohol, tobacco). **Design**   Meta-analysis was used to determine the average effect size and potential moderators in 47 comparisons of the effectiveness of CM from studies based on a treatment–control group design and published between 1970 and 2002. **Findings**   The mean effect size (ES) of CM was positive, with a magnitude of $d = 0.42$ using a fixed effects model. The magnitude of the ES declined over time, following treatment. CM was more effective in treating opiate use ($d = 0.65$) and cocaine use ($d = 0.66$), compared with tobacco ($d = 0.31$) or multiple drugs ($d = 0.42$). Larger effect sizes were associated with higher researcher involvement, earlier studies and shorter treatment duration. **Conclusions**   Study findings suggest that CM is among the more effective approaches to promoting abstinence during the treatment of substance use disorders. CM improves the ability of clients to remain abstinent, thereby allowing them to take fuller advantage of other clinical treatment components.

Keywords   Contingency management, meta-analysis, substance abuse treatment.

INTRODUCTION

Contingency management (CM) interventions to treat people dependent on psychoactive substances are based on a robust theoretical and empirical science literature that regards drug use and addiction as a form of operant conditioning in which behavior is controlled or shaped by its consequences [1,2]. As such, the likelihood of substance use should be influenced by the context in which use occurs. More specifically, alternative non-drug reinforcers should decrease substance use if they are available in sufficient magnitude and according to a schedule that is incompatible with drug use [3–5]. These observations form the theoretical basis for the use of contingency-management approaches to treat substance use disorders [6].

Researchers have tested various types of CM procedures to address dependence on a range of psychoactive substances (alcohol, tobacco, illicit drugs). One commonly used type of CM has been popularized by Higgins and colleagues (e.g. [7–9]). In this procedure, often called voucher-based reinforcement therapy (VBRT), patients receive 'vouchers' having various monetary values for providing biological samples (urine or breath) that are negative for the tested drugs. These vouchers are withheld when the biological sample indicates recent drug use. Once earned, vouchers are exchanged for goods or services that are compatible with a drug-free life-style.

Numerous variations of VBRT reinforcement schedules are possible, primarily in an effort to promote continuous abstinence: the value of the voucher escalates with each successive negative urine, the value of the voucher is reset to a lower value following a positive urine, or a bonus is provided after a certain number of negative urines are produced (e.g. [5,10,11]).

Another CM technique was developed by Petry [12–14] and has been termed colloquially the 'fishbowl' procedure; however, it is more properly called the variable magnitude of reinforcement procedure. This technique was selected recently for inclusion in the first wave of protocols for NIDA's Clinical Trials Network [15,16]. Participants receive draws, often from a number of slips of paper kept in a fishbowl (hence the name), for providing a negative biological specimen. Provision of a specimen

indicating recent drug use results in the withholding of draws. Each draw has a chance of winning a 'prize', the value of which varies. Typically, about half the draws say 'Good Job!' The other half of the draws result in the earning of a prize, which may range in value from $1 to $100. Similar to the VBRT technique, clients receive bonus draws after submitting a given number of consecutive negative specimens. This approach was designed as a way to curtail the cost of voucher-based interventions. Recent research suggests that this technique is of comparable efficacy to the VBRT technique [17].

Other variations of CM using different reinforcers to promote abstinence or reduce substance use have also been used. One type of CM makes the receipt of take-home doses of methadone contingent upon the provision of one or more drug-free urine tests (e.g. [18]). Similarly, access to affordable housing and work opportunities have been made contingent on negative drug tests [19,20]. Contracts have been made with clients stipulating that verified drug use will result in a letter detailing their drug use being sent to the client's professional organization (e.g. state medical association [21,22]).

Regardless of the specific procedure, researchers have used CM to promote abstinence from many types of drugs, including benzodiazepines [23], cocaine [9], nicotine [24], alcohol [13], opiates [25–27], marijuana [28,29] and methamphetamine [11]. Because CM has been found to assist clients in initiating and maintaining abstinence, it may also help them to engage in counseling and other services in a more productive fashion than they would if they were intoxicated. Indeed, tests of CM have nearly always been conducted on a 'platform' of another treatment, such as methadone maintenance or outpatient drug-free treatment, rather than as a stand-alone treatment.

In addition to drug use itself, CM has been used to change behaviors that are associated with improved drug use and other outcomes, such as treatment attendance and retention [30,31]. Reinforcement has also been used to promote such activities as education, employment, family reunification, non-drug-related leisure activities and self-help group attendance—all intended to assist clients develop life-styles that enable them to resist future drug-use episodes more readily and thus prevent relapse [32–34].

There have been several literature reviews (e.g. [35,36]) and two meta-analyses of CM treatment. One meta-analysis [37] focused on the use of CM within the context of methadone maintenance treatment. A second, more recent meta-analysis [38] examined the most commonly used approach to CM, voucher-based reinforcement therapy. Both these meta-analyses found CM to be an effective technique, with small to medium effect sizes depending on the outcome variable. The current study is more comprehensive than the two previous meta-analyses in including studies of all types of CM techniques published since 1970.

This paper focuses on treatment–control group studies of CM with respect to drug use outcomes. We report characteristics of the studies, present descriptive data on effect sizes and examine moderators that may account for variation in effect sizes across CM interventions.

METHODS

The procedures for this meta-analysis included five steps: (1) specification of study eligibility criteria; (2) search, retrieval and selection of studies based on the eligibility criteria; (3) coding of substantive and methodological characteristics of each study; (4) calculation of the effect size for the outcome variables in each study; and (5) calculation of the mean effect size across studies and examination of the relationship between selected study characteristics and effect size. Because of its nature, the study was exempt from Institutional Review Board review.

Eligibility criteria

The eligibility criteria were similar to those used in other meta-analyses of substance use disorder treatment and similar interventions (e.g. [39–41]). Studies were eligible if they were outcome evaluations of CM treatment for dependence on alcohol, tobacco or illicit drugs delivered to juveniles or adults. The studies were limited to those published in English between 1970 and 2002. Studies had to have used a treatment–control group design, although both experimental and quasi-experimental designs were eligible. A study needed to contain the quantitative data needed to calculate an effect size (e.g. means, standard deviations, percentages, number of subjects) on at least one outcome variable. A study had to have a total sample size of 10 or higher to be eligible because CM studies with smaller sample sizes tend to be case studies or single-group studies. Given our focus on the effectiveness of CM procedures to treat substance use disorders, we did not select studies of CM in which, for example, different magnitudes of reinforcers or different schedules of reinforcement were compared with one another and in which there was not a no-CM control group. In other words, this meta-analysis addressed the research question: 'Is CM effective relative to no treatment or when added to standard treatment?', not 'What is the relative effectiveness of different methods of delivering CM?'

Search strategy and document selection

The following bibliographic databases were searched: Current Contents, BIOSIS, Embase, MEDLINE, PsychInfo,

Sociological Abstracts, Cork Database and Cochrane Library. We constructed Boolean searches that paired terms pertaining to a particular technique (e.g. 'contingency management', 'voucher based reinforcement', 'behavioral contracting' or 'token economy') with terms referring to specific problems or substances that were the object of treatment (e.g. 'addiction', 'drug abuse', 'alcoholism', 'cocaine', 'opiates' or 'tobacco'). As empirical studies and literature reviews were retrieved and cataloged, we examined their reference lists for other potentially relevant documents.

Approximately 1150 documents were identified during the literature search. An initial screen of titles and abstracts using the eligibility criteria listed above determined that 203 documents were potentially eligible. After a second-level screening of retrieved documents, 113 documents remained potentially eligible. Finally, coders identified documents that did not meet eligibility criteria, including studies that had insufficient data to calculate an effect size. Senior research staff confirmed their ineligibility. A total of 81 eligible documents were identified. Because a single study was sometimes reported in multiple documents, all retrieved documents associated with a given study were grouped together before coding. Alternatively, a single document might report results on more than one independent comparison. In such cases, each comparison was coded as a separate study.

Coding procedures

The codebook consisted of questions from previous meta-analyses on related subjects [40,42–45] and questions specifically relevant to CM treatment (the codebook is available from the first author). Items in the codebook covered five areas: study context, methodology, participant characteristics, treatment characteristics, dependent variable characteristics and effect size calculation. Project staff pilot-tested the initial version of the codebook. Further revisions occurred during the early stage of coding. Decisions and clarifications regarding specific questions were recorded in a policy manual that was regularly updated for use by the coders.

Each study was coded by one of five masters- or doctoral-level coders. They received training on use of the codebook and effect size calculation; they then coded and discussed three practice studies. Once coding began, coders met with senior project staff to discuss questions about the codebook or specific studies every 2 weeks for the first several months of coding and monthly thereafter. In addition, before data entry, coding for each study was reviewed for accuracy by a postdoctoral fellow with expertise in contingency management. Proposed changes were discussed with the original coder and with the senior investigators before being adopted.

Effect size calculation and aggregation

The most common type of effect size for treatment evaluation studies is the standardized mean difference: $d = (M_t - M_c)/SD_{pooled}$, where d is the effect size estimate, M_t is the mean of the treatment group, M_c is the mean of the comparison group and SD_{pooled} is the pooled standard deviation of the two groups. If means and standard deviations were not available, effect sizes were estimated from the reported value of the t, F or χ^2 statistic (see formulae in Cooper & Hedges [46]; Hedges & Olkin [47]). Effect sizes from proportions were calculated using the arcsine transformation [48]: $d = \arcsin(p_1) - \arcsin(p_2)$, where p_1 is the proportion of 'success' for the treatment group and p_2 is the proportion of 'success' for the comparison group. The correction factor for small sample size recommended by Hedges & Olkin [47] was applied to effect sizes. An effect size for which the treatment group showed more success than the comparison group is indicated by a positive sign, and an outcome favoring the comparison group, by a negative sign. Coders used the meta-analysis software program DSTAT version 1.1 [49] to calculate effect sizes.

We followed standard practice in weighting each effect size by the inverse of its variance (see formulae in Lipsey & Wilson [50]). Weighting each effect size in this way assumes a fixed effects model in which the combined individual effect sizes estimate a single population effect size (i.e. the population variance is zero). An alternative assumption is that the effect sizes for the included studies consist of a sample that has been drawn from a random distribution of study population effect sizes, which requires the use of a random-effects model in averaging effect sizes. In applying this model, a 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 (i.e. they have wider confidence intervals). Both fixed-effects and random-effects weighted means are shown in the descriptive statistics (Table 1). The variability in effect sizes was examined through homogeneity analysis, which compares the observed variation in the effect sizes with the variation that would be accounted for by sampling error alone.

Although we calculated an effect size for each outcome reported in a study, the most common outcome variable was drug use (alcohol, tobacco and/or illicit drugs). Other types of outcomes (e.g. crime, employment, housing) were reported in 10 or fewer studies. Because drug use was the most commonly reported outcome variable and because it is of greatest relevance to assessing the effectiveness of CM, we limited our analysis of effect sizes to drug use variables. For those studies that included more than one drug use outcome variable (e.g. cocaine

Table 1 Summary statistics of comparison-level effect sizes for drug use outcomes measured at the end of intervention or averaged over the intervention.

Number of comparisons	47
Unweighted mean effect size ($n = 47$)	0.57
Standard deviation	0.43
Maximum (truncated)	1.54
Median	0.47
Minimum	0.00
Fixed effects weighted mean (95% CI)	0.42 (0.35, 0.50)
Homogeneity test (Q; df = 46)	86.89*
Random-effects weighted mean (95% CI)	0.49 (0.38, 0.59)
Homogeneity test (Q; df =46)	48.35

*$p < 0.05$.

use measured by self-report and by urinalysis), we averaged the effect sizes for each drug use measure to create a study-level effect size.

Finally, studies varied in the time-point(s) at which outcomes were assessed. Although many studies reported a summary measure (usually a mean) based on urine test results conducted at multiple points during treatment, some studies assessed drug use only once at the end of treatment. A smaller number of studies reported outcomes at various time-points following the intervention. The main outcome variable for this analysis was a measure of drug use (usually determined from urine test results) that was taken multiple times during treatment and averaged to produce a single value or, if multiple assessments were not available, a measure of drug use determined at the end of treatment. For studies that did report post-treatment follow-ups, we categorized the time into 3-month periods up to 12 months, and then averaged effect sizes within each quarter in order to determine to what extent the treatment effect changed over the months following the end of treatment.

In order to examine the likelihood of publication bias on the overall effect size estimate, we used the trim-and-fill method developed by Duval & Tweedie [51–53]. Specifically, the trim-and-fill method assumes that the funnel plot of the 'true' set of effect sizes is symmetrical and that any observed asymmetry may be due to omitted studies. Based on these assumptions, the procedure imputes the number of studies that might be 'missing' from the analysis and their effect sizes—i.e. the data points that would be required to make the funnel plot appear symmetrical. The procedure then adjusts the overall effect size obtained from the data on the assumption that the imputed effect sizes were actual data points.

Moderator variables

In addition to calculating a mean effect size across studies, we examined moderators that might explain variation across CM studies in the main effect size of interest

(i.e. substance use averaged over the course of treatment or assessed at the end of treatment). While many design, client and treatment factors could moderate effect size, we selected four variables that are relevant to clinical practice or that previous meta-analyses have found to be important to outcomes: the decade when the study was conducted, the role of the researcher in the intervention, the drug targeted for reinforcement and the planned duration of the intervention.

The decade in which the study was conducted (dichotomized between the 1970s–1980s and the 1990s) was included to determine whether mean effect sizes have changed over time.

Degree of researcher involvement has been found in previous meta-analyses [43,54] to be a significant moderator of effect sizes. Using the description of the intervention in the document(s), the coders rated how involved the researcher was in developing and/or delivering the intervention. For analysis purposes, the responses were dichotomized into less involved or more involved.

A third potential moderator was the drug targeted for reinforcement. This was not necessarily the client's primary drug, but the drug for which abstinence was reinforced as part of the CM protocol. The studies were grouped into four categories: cocaine only, opiates only, tobacco only and polydrug (more than one drug). Alcohol only was not one of the categories because it was the target of reinforcement in only one study. The mean effect size for the analysis of this moderator was based on the targeted drug for a given study and did not include any other drugs that may have been reported (except for studies in which multiple drugs were targeted).

Finally, the planned duration of the intervention was examined to assess whether effect size varied by the number of weeks over which the intervention was administered. Planned duration was categorized into three levels: 1–11 weeks, 12–25 weeks and 26 or more weeks (with the longest intervention being 142 weeks).

Examination of the effectiveness of CM involved two stages. The overall results were first summarized in terms of descriptive statistics using inverse-weighted techniques for combining effect sizes [47]. The second stage focused on evaluating selected potential moderators of effect size [50]. The analyses were generated using SAS/STAT software, version 9.1.3. The trim-and-fill method to examine publication bias [51,52] was conducted using the user-written program *metatrim* [55] in STATA/SE software, version 8.2.

RESULTS

It was determined that the 81 eligible documents were reports of 75 comparisons. Three studies [56–58] did not have the required data to calculate effect sizes using the

techniques described above. In addition, although studies that used an ABA design were coded ($k = 15$), these types of studies were not included in the analysis reported here because the within-group change score from such studies does not represent the same type of effect size as the between-group difference score from treatment–control group studies. Finally, 10 comparisons that did not have at least one effect size for a drug-related dependent variable were not included in the analysis. Thus, for purposes of this review, the analyses reported below consist of 47 comparisons.

Characteristics of study comparisons

Characteristics of each comparison included in the analysis are presented in Table 2. In terms of aggregate characteristics, 91.5% of the studies used experimental designs employing random assignment procedures. Thus, overall the studies used rigorous research designs that were, in principle, capable of controlling various threats to internal validity.

For all comparisons, the coded document was a published journal article. More than two-thirds of the comparisons (70.2%) were based on research conducted during the 1990s. All of the studies took place in the United States. The median sample size was 69, ranging from 12 to 844 (as noted above, studies with fewer than 10 participants were not eligible). Nearly all the studies were funded by the National Institute on Drug Abuse or other agency within the National Institutes of Health.

Studies varied on the type of reinforcement and the combination of reinforcers used. The three most frequently used reinforcers were vouchers (55.3%), methadone take-home doses or dosage adjustments (23.4%) and cash (21.3%). Other types of reinforcers included prizes, recreational activities, housing and social reinforcement. Some studies used more than one type of reinforcer.

Descriptive results

Figure 1 displays a stem-and-leaf plot of the drug use effect sizes for each comparison included in the analysis. The stem consists of the first digit(s) of an effect size and the leaf identifies the final digit of an effect size. The stem-and-leaf plot shows that all of the effect sizes for CM studies are positive, but also that the effect sizes are widely distributed. As the values of the two largest effect sizes (1.94 and 2.96) are substantial outliers, in subsequent analyses we truncated them to the next highest effect size value (1.54).

Table 1 presents summary statistics of the effect sizes for the 47 treatment–control group comparisons that had drug use outcomes measured during or at the end of the intervention. The median effect size was $d = 0.47$. The fixed-effects weighted mean effect size was $d = 0.42$, with

Stem	Leaf
2.9	6
2.8	
2.7	
2.6	
2.5	
2.4	
2.3	
2.2	
2.1	
2.0	
1.9	4
1.8	
1.7	
1.6	
1.5	4
1.4	
1.3	5
1.2	4
1.1	258
1.0	2
0.9	
0.8	09
0.7	067
0.6	12489
0.5	448
0.4	178
0.3	44688
0.2	0246789
0.1	0555889
0.0	002

Note: The stem-and-leaf plot provides a graphical representation of the distribution of effect sizes across studies. The stem identifies the first digit(s) of an effect size and the leaf identifies the final digit.

Figure 1 Stem-and-leaf plot of comparison-level effect sizes for drug use outcomes in contingency management studies

a 95% confidence interval ranging from 0.35 to 0.50 and a Q (homogeneity statistic) value of 86.89 (df = 46; $P < 0.05$). Compared with the fixed-effects mean effect size, the random-effects weighted mean effect size was higher ($d = 0.49$) but, as expected, the confidence interval was wider (0.38, 0.59); the Q statistic was 48.35 (df = 46; NS).

The statistical interpretation of the standardized mean difference effect size index is the distance, measured in standard deviation units, of the average client in the treatment group from the average client in the comparison group on the outcome variable. A more clinically relevant equivalent of the standardized mean difference is the binomial effect size display (BESD; [59,60]). In calculating the BESD, the threshold of success is assumed to be the grand median of the combined treatment group and comparison group scores. Because specific a priori criteria for 'success' or 'failure' for a particular treatment are rarely available or easily defined, the overall median can be regarded as a hypothetical, but none the less useful, representation of success rates. For this meta-analysis, the effect size of $d = 0.42$ is equivalent to a success rate of 61% for the treatment group and 39% for the comparison group, or a difference of 22%. (Note that for typical

Table 2 Characteristics of eligible contingency management comparisons (*n* = 47).

Study (author/year)	Decade conducted	Design type	Planned duration of treatment (weeks)	No. of subjects (baseline)	Attrition (%)	Longest follow-up (months)	Source of dependent variable(s)	Target drug(s)	Type of reinforcer(s)	Non-contingent reinforcement for control group	Mean effect size d	95% CI
Azrin [76] (1994)	1990s	EXP	52	104	21.15	0	Both	NA	Cash; other	No	0.59	0.12, 1.05
Brooner [77] (1998)	1990s	EXP	13	43	25.58	0	ST	Polydrug	Methadone take-homes; methadone dosage decrease; methadone dosage increase; graduation to next treatment phase; other	No	0.18	−0.44, 0.80
Budney [28] (2000)	1990s	EXP	12	60	25.00	0	Both	Marijuana	Vouchers	No	0.26	−0.48, 1.00
Calsyn [78] (1994)	1980s	EXP	52	360	65.00	0	ST	Polydrug	Methadone dosage decrease; methadone dosage increase; program discharge	No	0.36	0.08, 0.64
Carroll [79] (2001)	1990s	EXP	12	127	62.20	0	ST	Polydrug	Vouchers	No	0.27	−0.18, 0.72
Chutuape [80] (1999)	1990s	EXP	12	14	14.29	0	ST	Polydrug	Vouchers; methadone take-homes	No	1.54	0.03, 3.05
Chutuape [81] (1999)	1990s	EXP	16	29	10.34	0	ST	Polydrug	Methadone take-homes	No	0.76	−0.18, 1.70
Donatelle [82] (2000)	1990s	EXP	32	220	41.82	3	ST	Tobacco	Vouchers	No	0.29	0.01, 0.57
Downey [83] (2000)	1990s	EXP	12	41	48.78	0	ST	Polydrug	Vouchers	Yes	0.38	−0.80, 1.55
Elk [30] (1998)	1990s	EXP	142	12	25.00	0	Both	Cocaine	Vouchers	No	0.19	−0.94, 1.32
Glasgow [84] (1993)	1990s	EXP	52	474	19.62	12	Both	Tobacco	Cash	No	0.00	−0.28, 0.28
Glosser [85] (1983)	1970s	QEXP	142	117	17.09	0	ST	Polydrug	Methadone take-homes; methadone dosage decrease; methadone dosage increase	No	0.69	−0.10, 1.48
Gruber [86] (2000)	1990s	EXP	4	52	0.00	2	Both	Polydrug	Vouchers; other	No	0.64	0.07, 1.21
Hall [87] (1979)	1970s	EXP	2	81	55.56	0	ST	Opiate	Cash	No	0.80	0.36, 1.24
Higgins [88] (1995)	1990s	EXP	24	40	42.50	6	Both	Cocaine	Vouchers	No	0.47	−0.19, 1.12
Higgins [25] (1986)	1980s	EXP	7	39	15.38	0	ST	Opiate	Methadone dosage increase	Yes	1.02	0.20, 1.84
Higgins [7] (1991)	1980s	EXP	12	25	36.00	0	ST	Cocaine	Vouchers	No	1.18	0.33, 2.03
Higgins [89] (2000)	1990s	EXP	24	70	45.71	6	Both	Cocaine	Vouchers	Yes	0.24	−0.24, 0.72

Table 2 *Cont.*

Study (author/year)	Decade conducted	Design type	Planned duration of treatment (weeks)	No. of subjects (baseline)	Attrition (%)	Longest follow-up (months)	Source of dependent variable(s)	Target drug(s)	Type of reinforcer(s)	Non-contingent reinforcement for control group	Mean effect size d	95% CI
Iguchi [90] (1988)	1980s	EXP	20	16	43.75	0	ST	Polydrug	Methadone take-homes; methadone decrease; methadone dosage increase	No	0.02	−1.03, 1.07
Jason [91] (1997)	1990s	QEXP	26	844	0.00	12	ST	Tobacco	Cash	No	0.21	−0.07, 0.48
Jason [92] (1990)	1980s	QEXP	26	95	3.16	6	ST	Tobacco	Cash	No	0.70	0.27, 1.13
Jones [31] (2001)	1990s	EXP	2	85	5.88	0	ST	Polydrug	Vouchers	No	0.35	−0.10, 0.79
Kassaye [93] (1984)	1980s	EXP	NS	109	7.34	0	SR	Tobacco	Cash	No	1.54	0.99, 2.09
Katz [94] (2002)	1990s	EXP	12	52	0.00	0	ST	Polydrug	Vouchers	No	0.22	−0.33, 0.77
Kidorf [95] (1993)	1990s	EXP	7	44	0.00	3	ST	Cocaine	Graduation to next treatment phase	Yes	0.48	−0.14, 1.09
Koffman [96] (1998)	1990s	QEXP	52	187	0.00	0	ST	Tobacco	Cash	No	0.15	−0.17, 0.47
Lando [97] (1976)	1970s	EXP	17	49	0.00	0	SR	Tobacco	Fee reduction	No	0.00	−0.69, 0.69
McCarthy [98] (1985)	1980s	EXP	52	69	57.97	0	ST	Polydrug	Program discharge	No	0.55	−0.26, 1.35
McCaul [99] (1984)	1980s	EXP	12	20	55.00	0	ST	Opiate	Cash; methadone take-homes; reduction in clinic responsibilities	Yes	1.35	0.38, 2.32
Milby [100] (1978)	1970s	EXP	7	75	8.00	0	ST	Opiate	Methadone take-homes	No	0.28	−0.26, 0.82
Milby [101] (2000)	1990s	EXP	26	110	20.00	12	ST	Polydrug	Vouchers; other	No	0.89	0.49, 1.29
Petry [13] (2000)	1990s	EXP	8	42	7.14	0	Both	Alcohol	Vouchers	No	0.68	−0.02, 1.37
Petry [12] (2002)	1990s	EXP	12	42	11.90	3	Both	Polydrug	Vouchers	No	0.39	−0.24, 1.01
Piotrowski [102] (1999)	1990s	EXP	26	102	6.86	0	ST	Polydrug	Vouchers	No	0.55	−0.72, 1.81
Preston [103] (2000)[a]	1990s	EXP	8	120	6.67	0	Both	Polydrug	Vouchers	Yes	0.18	−0.36, 0.71
Preston [103] (2000)[a]	1990s	EXP	8	120	6.67	0	Both	Polydrug	Vouchers	Yes	0.15	−0.35, 0.65
Preston [104] (2002)	1990s	EXP	12	110	24.55	0	Both	Polydrug	Vouchers; methadone take-homes	Yes	0.10	−0.27, 0.47
Rawson [68] (2002)[a]	1990s	EXP	16	120	15.83	0	ST	Cocaine	Vouchers	No	0.77	0.25, 1.29
Rawson [68] (2002)[a]	1990s	EXP	16	120	15.83	0	ST	Cocaine	Vouchers	No	0.34	−0.17, 0.85
Roll [10] (1996)	1990s	EXP	1	61	1.64	0	ST	Tobacco	Cash	Yes	1.25	0.41, 2.08
Shoptaw [105] (2002)[a]	1990s	EXP	12	175	26.86	9	Both	Tobacco	Vouchers	No	0.41	−0.09, 0.91
Shoptaw [105] (2002)[a]	1990s	EXP	12	175	26.86	9	Both	Tobacco	Vouchers	No	0.15	−0.34, 0.64
Silverman [106] (1996)	1990s	EXP	12	37	13.51	0	ST	Polydrug	Vouchers	Yes	1.13	0.40, 1.85
Silverman [107] (1998)	1990s	EXP	12	59	0.00	1	ST	Cocaine	Vouchers	Yes	1.54	0.85, 2.23
Silverman [20] (2001)	1990s	EXP	24	40	0.00	0	ST	Polydrug	Vouchers	No	0.61	−0.03, 1.25
Stitzer [108] (1985)	1980s	EXP	6	28	0.00	0	Both	Tobacco	Cash	No	1.15	0.14, 2.16
Stitzer [109] (1992)	1980s	EXP	48	53	32.08	0	ST	Polydrug	Methadone take-home	Yes	0.62	0.07, 1.17

EXP = experimental design; QEXP = quasi-experimental design; SR = self-report; ST = specimen testing; NS = not specified; NA = not available. [a]Multiple comparisons.

values of d, the BESD translates to a percentage difference between the treatment group and the control group of approximately one-half of d).

Because decisions about coding and analysis may have affected the effect size estimate, several analyses examined whether alternative decisions would produce different results. First, although all the studies included in this analysis used treatment–comparison group design, not all of them allocated subjects randomly to study conditions. Four of the studies were quasi-experimental and the others used random assignment. When we removed the four quasi-experimental designs from the analysis, the weighted mean effect size (fixed effect model) increased slightly from $d = 0.42$ to $d = 0.44$ ($k = 43$; 95% CI = 0.36, 0.52). Secondly, the substance use outcome measure included both self-report and testing of biological specimens (urine, blood, carbon monoxide). When we limited the measure of substance use to biological testing, the mean effect size increased to $d = 0.48$ ($k = 40$; 95% CI = 0.39, 0.56). Finally, because drug use was defined in two different ways—a mean during-treatment measure (16 studies) and an end-of-treatment measure (31 studies), we compared the mean effect size of each measure. Although the effect size for drug use averaged during treatment was higher than that for drug use assessed at the end of treatment ($d = 0.49$ and $d = 0.40$, respectively), the difference between the two values was not significant ($Q_{between} = 1.21$; df = 1; $P = 0.27$).

Some studies reported drug use outcomes measured at various points following the end of the CM intervention. In analyzing these follow-up data, we averaged the effect sizes (using a fixed-effects model) for the during-treatment period and within each quarter of the follow-up year, being sure that a study was included at only one time-point in the analysis (hence the lower number of studies included in the during-treatment mean effect size than the number shown in Table 1). If a study reported data at multiple time-points, the time-point of the last follow-up was selected. Although the number of studies reporting outcomes within each quarter is small, the general trend in effect size is downward. There is a drop from $d = 0.52$ ($k = 33$; 95% CI = 0.42, 0.62) for the during-treatment effect size to $d = 0.37$ ($k = 6$; 95% CI = 0.14, 0.59) in the 3 months following treatment. For the second and third quarters, the mean effect sizes are based on only two studies for each quarter, so the estimates are imprecise and unstable: second quarter, $d = 0.45$ ($k = 2$; 95% CI = 0.22, 0.69); third quarter, $d = 0.34$ ($k = 2$; 95% CI = −0.06, 0.74). For the fourth quarter, the effect size drops to $d = 0.12$ ($k = 5$; 95% CI = −0.06, 0.29). The 95% CI indicate that the third-and fourth-quarter estimates are not significantly different from zero.

Because unpublished studies were not included in the set of eligible studies, the overall effect size may be subject to publication bias, which arises from the tendency of journals to publish studies with statistically significant findings and hence overestimates of treatment effects [61,62]. Applying the trim-and-fill method (using the *metatrim* procedure in STATA/SE [55]) to our data, it was estimated that 15 comparisons with negative effect sizes may have been omitted from our analysis—i.e. that would make the funnel plot symmetrical. Recalculation of the mean effect size using the estimated effect sizes from these imputed studies resulted in an adjusted effect size (and confidence interval) of 0.27 (0.21, 0.34, $P < 0.01$; fixed effects) and 0.29 (0.17, 0.42, $P < 0.01$; random effects). While these results suggest that inclusion of missing studies would lead to a reduction in the overall effect size, the lower effect size is still significantly different from zero.

Moderator analysis

As noted above, we examined several variables that might moderate the relationship between CM and effect size for drug use reported in Table 1. Table 3 reports findings of ANOVA analog analysis comparing the effect of four moderator variables. Results were computed using fixed-effect statistics and indicate the extent to which each moderator accounts for variability in effect sizes with respect to drug use outcomes. A significant value of Q-between indicates significant differences among effect sizes between the categories of the moderator variable. The *post-hoc* contrasts were tested using a 0.05 significance level.

With respect to when studies were conducted, studies conducted during the 1990s had a significantly smaller effect size ($d = 0.35$) than those conducted in the 1970s and 1980s ($d = 0.64$). Studies in which the researcher was more involved in the design or delivery of treatment yield a significantly larger mean effect ($d = 0.46$) than the few studies ($k = 4$) in which the researcher was less involved ($d = 0.14$). The type of drug targeted for reinforcement moderated the effects of CM treatment. CM was more effective in treating opiate use ($d = 0.65$) and cocaine use ($d = 0.66$), compared with tobacco ($d = 0.31$) or multiple drugs ($d = 0.42$). In *post hoc* contrasts, CM approaches that specifically targeted either cocaine use or opiate use had significantly higher effect sizes than those that targeted tobacco use.

Finally, analysis of planned duration of the intervention indicated that longer duration of treatment was associated with lower effect size. Studies in which the intervention lasted from 1 week to 11 weeks had an effect size of $d = 0.58$. For interventions of 12–25 weeks, the effect size was $d = 0.44$ and for 26 weeks and over, $d = 0.34$. In *post hoc* contrasts between pairs of categories,

Table 3 ANOVA analog results: effect sizes for contingency management versus comparison conditions by selected moderators.

Moderator	k	Effect size [a]	95% CI	Q [b]	P of Q	Q-within (df)	P of Q-within	Q-between (df)	P of Q-between
Decade						76.39 (45)	< 0.01	10.50 (1)	< 0.01
1970s–1980s	14	0.64	0.49–0.79	26.59	0.01				
1990s	33	0.35	0.27–0.44	49.80	0.02				
Role of researcher						79.01 (45)	< 0.01	7.88 (1)	< 0.01
Less involved	4	0.14	−0.07–0.35	2.79	0.43				
More involved	43	0.46	0.38–0.54	76.22	0.001				
Target drug[c]						74.19 (39)	< 0.01	8.72 (3)	0.04
Cocaine only	8	0.66	0.44–0.87	11.32	0.13				
Opiates only	3	0.65	0.25–1.06	4.60	0.10				
Tobacco only	11	0.31	0.19–0.43	37.55	< 0.01				
Polydrug	21	0.42	0.30–0.54	20.71	0.41				
Planned treatment duration[d]						79.5 (44)	< 0.01	7.39 (2)	0.03
1–11 weeks	12	0.58	0.45–0.78	24.83	< 0.01				
12-25 weeks	22	0.44	0.29–0.55	333.27	0.04				
26+ weeks	13	0.34	0.23–0.45	21.40	0.05				

[a]Weighted fixed effects. [b]The Q (homogeneity) statistic tests whether the individual effect sizes varies among each other by more than would be expected from sampling error. The degrees of freedom for the Q's of each moderator category (5th column) are equal to $k-1$. [c]*Post-hoc* pairwise contrasts of the mean weighted effect sizes of cocaine versus tobacco and opiates versus tobacco are significant at $P < 0.05$. [d]*Post-hoc* pairwise contrast of the mean weighted effect sizes of 1–11 weeks versus 26+ weeks is significant at $P < 0.05$.

the effect sizes for 1–11 weeks and for 26 + weeks were significantly different.

DISCUSSION

In this meta-analysis of the effectiveness of CM techniques in treatment–comparison studies, the primary outcome was drug use (alcohol, tobacco or illicit drugs), either the average of multiple drug use measures taken during the intervention or the outcome measure at the end of the intervention. The mean (weighted) effect size was positive, with a magnitude of $d = 0.42$ using a fixed-effects estimation model and $d = 0.49$ using a random-effects model. All the effect sizes included in the analysis were positive or zero.

In an earlier meta-analysis of studies that evaluated a variety of types and techniques of drug abuse treatment, Prendergast and colleagues [41] found a weighted mean effect size of $d = 0.30$ (fixed effects), which is equivalent to a success rate of 57% for the treatment group and 42% for the control group. With respect to specific types of treatment, this previous meta-analysis found the following effect sizes for the main treatment modalities: methadone maintenance, $d = 0.45$ $(k = 8)$; Therapeutic Community, $d = 0.25$ $(k = 8)$; and out-patient drug-free, $d = 0.37$ $(k = 8)$. It is perhaps not surprising that the mean effect size for the CM studies $(d = 0.42)$ was larger than the overall effect size in the previous meta-analysis $(d = 0.30)$. First, there was a much wider range of treatment types and design types in the previous review, which might have reduced the overall effect size. Sec-

ondly, the dependent variable for the previous meta-analysis was drug use measured at the end of treatment or (more commonly) at the first follow-up assessment following the end of treatment, which could have been up to 12 months. For the present meta-analysis, the dependent variable was drug use measured during treatment or at the end of treatment. Thus, the effect size for CM would be higher than other meta-analyses of drug treatment in which outcomes were measured post-treatment.

Other meta-analyses on substance use disorder treatment effectiveness have also found positive and significant outcomes [37,41,63–65], but methodological differences across the individual meta-analyses (e.g. selection criteria, type of treatment, effect size index, weighting procedures) make direct comparison of effect sizes across the meta-analyses difficult. With regard to the two previous meta-analyses of CM, the study by Griffith and colleagues [37] focused on methadone maintenance clients. The authors found a mean end-of-treatment effect size (using the *r* index of correlation) of $r = 0.25$ among 25 studies, which is a moderate effect size based on Cohen's criteria [48]. The study by Lussier *et al*. [38] found that the CM approach that it examined (i.e. VBRT) yielded a high moderate effect size for drug use abstinence of $r = 0.32$ over 30 studies. As noted above, the current meta-analysis of CM differed from the other two in including a broader range of approaches to CM and on including a longer time-period for study inclusion. All three meta-analyses of contingency management found a significant, positive treatment effect of relatively high magnitude.

For the studies examined here, the magnitude of the effect declined over time. However, as few CM studies reported outcomes following treatment, the post-treatment effect sizes should be interpreted with caution. Nevertheless, a decline in the positive effects of CM after reinforcement has been discontinued is to be expected for CM specifically [66], and for other types of treatments [67]. For most interventions, the magnitude of the effect observed at the end of treatment is not maintained in the months following treatment. In order to continue the gains produced by CM during treatment, it might be useful to combine CM with others interventions that provide more sustained effects (e.g. cognitive behavioral therapy, see Epstein *et al.* [66]; Rawson *et al.* [68]).

Perhaps the most counter-intuitive finding in the moderator analysis was that studies conducted in the 1970s and 1980s yielded a larger mean effect size than studies conducted in the 1990s ($d = 0.64$ versus $d = 0.35$). Given the theoretical and empirical development of CM over that period, it is unlikely that the true effect of CM declined over time. A more plausible explanation is that the rigor of CM studies has improved. For example, more recent studies may be more likely to employ intent-to-treat analyses than was true in earlier research. Some meta-analyses have found that better quality studies that control for threats to internal validity and limit bias have smaller effect sizes than studies of lower quality [69,70]. Thus, while more recent studies of CM have lower effect sizes than earlier studies, they may also reflect the true (population) effect size more accurately; but other explanations are possible. For example, changes in client characteristics, including more poly-drug use, more entrenched patterns of drug use and a larger number of problems among clients entering treatment in more recent years [71] might mean that clients admitted to treatment in recent decades are more difficult to treat, even with a specialized technique such as CM.

The finding that greater researcher involvement was associated with higher effect sizes is consistent with that found in other meta-analyses [43,54]. Researcher involvement in design or delivery of treatment is related to the issue of attention to fidelity of program implementation, which a number of studies have found to be an important moderator of the effects of treatment [72–74]. 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 better treatment effects.

In terms of specific types of drug use, CM appears to be least effective in changing tobacco use (effect size of $d = 0.31$), and much more effective with opiates

($d = 0.65$) and cocaine ($d = 0.66$). CM is moderately effective when abstinence from several drugs is the target of reinforcement. The relatively large effect size for opiate-focused interventions may reflect the greater power of methadone take-home doses or dose adjustments as reinforcers.

While it might be expected that longer time in treatment would yield larger effects, longer CM interventions in this set of studies had lower effect sizes on average than did shorter interventions. It may, indeed, be the case that CM is more effective when administered in shorter rather than longer interventions. Given the chronic relapsing nature of substance abuse, it might be more difficult for clients to maintain abstinence in studies of CM that are of longer duration, although one recent study (too recent to be included in the set of articles coded) did find that a 52-week voucher-based intervention helped clients to remain abstinent from cocaine use for up to 1 year [75]. The finding may also result from the characteristics of the studies in this meta-analysis or of methodological features, or both.

Limitations

We took several steps to reduce the 'apple and oranges' problem through the eligibility criteria and as part of analysis. A set of eligibility criteria was used to provide guidelines as to which studies would be included and which not. The main selection criterion was that the study must focus on an empirical test of the use of CM techniques in the treatment of drug dependence (alcohol, tobacco and illicit drugs). We limited the outcome variable to drug use, the time-period for the main analysis to the time during treatment, and the design to treatment–comparison studies. Some of the heterogeneity that remained was examined in analyses of particular variables of interest with respect to their impact on effect size variability. However, had we used different definitions or categories, results might have varied from what we reported here.

The second limitation is publication bias, which arises from the greater likelihood that studies with significant findings will be published. The result is a set of unpublished reports with statistically non-significant (or negative) findings that remain in the 'file drawer' [61,62]. The trim-and-fill analysis indicated that 15 studies may be missing from the current meta-analysis. Because we did not include unpublished studies in the set of eligible studies, publication bias is a likely explanation for these missing 15 studies. However, there may be other factors that account for the missing studies. Also, the algorithm for calculating the adjusted effect size assumes that the missing effect sizes are those with the most extreme values on the left-side of the funnel plot distribution [53]. This assumption results in a conservative re-estimated

effect size. Nevertheless, the likelihood of publication bias in the current set of studies needs to be considered in interpreting the results of this meta-analysis.

With respect to methodology, studies included in the analysis were of high quality: nearly all used experimental designs with random assignment, sample sizes were relatively large for treatment evaluation studies, most were conducted in controlled settings and nearly all were funded by the National Institute on Drug Abuse or other Public Health Service agency (indicating a proposed design meeting peer-review standards). Eligible studies were those that used a treatment and a control (or comparison) group, although non-randomized studies were included. The treatment magnitude changed slightly (upward) when only experimental designs were included in the analysis. The mean effect size also increased somewhat when the analysis was limited to substance use outcome measured using biological specimens, but in neither case did these effect sizes differ significantly from weighted mean effect size shown in Table 1.

A final limitation is that the findings and conclusions of this (or any) meta-analysis depend on the decisions made by the researchers with respect to eligibility criteria, document selection, variables to be coded, rules used during coding, definitions of variables used in analysis and the approach to analyzing the data. Different decisions about any of these elements of the meta-analysis process could lead to different conclusions. We examined several of these decisions and found that the choices that we made in selecting studies and calculating the overall effect size produced lower effect sizes than other choices that we could have made. However, within the decision framework that informed this meta-analysis, which is typical of social science meta-analysis generally, the conclusions should have sufficient validity to provide guidance to the field.

CONCLUSION

The findings from this meta-analysis support the effectiveness of various CM techniques when used during treatment for treating clients who are dependent on alcohol, tobacco or illicit drugs. After clients are no longer subject to contingencies, the magnitude of the treatment effect begins to decline, although it appears to decay relatively slowly over time. The findings indicate that CM is able to establish and maintain abstinence for many clients during treatment, thereby permitting clients to engage more productively in treatment services that promote the broader psychosocial aspects of recovery. From this perspective, CM may be viewed as an adjunct to standard treatment, enhancing its effectiveness. Whether CM can serve as a stand-alone treatment is not known. It may be that for certain types of drug users,

mainly those early in their use or those with less severe problems, a protocol that reinforces abstinence only, provides few other services and has limited staff requirements might be effective. For clients with more severe problems, the limited data on effect sizes following CM suggest that continuing care is warranted.

Overall, the relatively large number of empirical studies on CM, the variety of drugs and client populations with which it has been used, the high methodological quality of CM studies and the relatively high mean effect size provide strong support for CM as being among the more effective approaches to promoting abstinence during and after the treatment of drug dependence disorders. Further research on this topic should include examination of the relative effectiveness of different types of CM, further investigation of moderators of the impact of CM and comparison of the effects of CM and other treatment approaches.

Acknowledgements

References

1. Bigelow G., Silverman K. Theoretical and empirical foundations of contingency management treatments for drug abuse. In: Higgins S. T., Silverman K., editors. *Motivating behavior change among illicit drug abusers: research on contingency management interventions*, 1st edn. Washington, DC: American Psychological Association; 1999, p. 15–31.
2. Higgins S. T. Applying learning and conditioning theory to the treatment of alcohol and cocaine abuse. In: Johnson B. A., Roache J. D., editors. *Drug addiction and its treatment: nexus of neuroscience and behavior*. Philadelphia: Lippincott, Williams & Wilkins; 1997, p. 367–85.
3. Carroll M. E., Lac S. T., Nygaard S. L. A concurrently available nondrug reinforcer prevents the acquisition or decreases the maintenance of cocaine-reinforced behavior. *Psychopharmacology* 1989; **97**: 23–9.

4. Higgins S. T., Bickel W. K., Hughes J. R. Influence of an alternative reinforcer on human cocaine self-administration. *Life Sci* 1994; **55**: 179–87.

5. Roll J. M., Higgins S. T. A within-subject comparison of three different schedules of reinforcement of drug abstinence using cigarette smoking as an exemplar. *Drug Alcohol Depend* 2000; **58**: 103–9.

6. Higgins S. T., Alessi S. M., Dantona R. L. Voucher-based incentives: a substance abuse treatment innovation. *Addict Behav* 2002; **27**: 887–910.

7. Higgins S. T., Delaney D. D., Budney A. J., Bickel W. K., Hughes J. R., Foerg F. *et al.* A behavioral approach to achieving initial cocaine abstinence. *Am J Psychiatry* 1991; **148**: 1218–24.

8. Higgins S. T., Budney A. J., Bickel W. K., Hughes J. R., Foerg F., Badger G. J. Achieving cocaine abstinence with a behavioral approach. *Am J Psychiatry* 1993; **150**: 763–9.

9. Higgins S. T., Budney A. J., Bickel W. K., Foerg F. E., Donham R., Badger G. J. Incentives improve outcome in outpatient behavioral treatment of cocaine dependence. *Arch Gen Psychiatry* 1994; **51**: 568–76.

10. Roll J. M., Higgins S. T., Badger G. J. An experimental comparison of three different schedules of reinforcement of drug abstinence using cigarette smoking as an exemplar. *J Appl Behav Anal* 1996; **29**: 495–505.

11. Roll J. M., Petry N. M., Stitzer M., Brecht M. L., Peirce J. M., McCann M. *et al.* Contingency management for the treatment of methamphetamine abuse. *Am J Psychiatry* in press.

12. Petry N. M., Martin B. Low-cost contingency management for treating cocaine- and opioid-abusing methadone patients. *J Consult Clin Psychol* 2002; **70**: 398–405.

13. Petry N. M., Martin B., Cooney J. L., Kranzler H. R. Give them prizes, and they will come: contingency management for treatment of alcohol dependence. *J Consult Clin Psychol* 2000; **68**: 250–7.

14. Petry N. M., Martin B., Finocche C. Contingency management in group treatment: a demonstration project in an HIV drop-in center. *J Subst Abuse Treat* 2001; **21**: 89–96.

15. Petry N. M., Peirce J. M., Stitzer M., Blaine J., Roll J. M., al. e. Prize-based incentives increase retention in outpatient psychosocial treatment programs: results of the national drug abuse treatment clinical trials network study. *Arch Gen Psychiatry* 2005; **62**: 1148–56.

16. Peirce J. M., Petry N. M., Stitzer M., Blaine J., Kellogg S., Satterfield F. *et al.* Effects of lower-cost incentives on stimulant abstinence in methadone maintenance treatment: a National Drug Abuse Treatment Clinical Trials Network Study. *Arch Gen Psychiatry* 2006; **63**: 201–8.

17. Petry N. M., Alessi S. M., Marx J., Austin M., Tardif M. Vouchers versus prizes: contingency management treatment of substance abusers in community settings. *J Consult Clin Psychol* 2005; **73**: 1005–14.

18. Stitzer M., Bigelow G. Contingency management in a methadone maintenance program: availability of reinforcers. *Int J Addict* 1978; **13**: 737–46.

19. Milby J. B., Schumacher J. E., Wallace D., Frison S., McNamara C., Usdan S. *et al.* Day treatment with contingency management for cocaine abuse in homeless persons: 12-month follow-up. *J Consult Clin Psychol* 2003; **71**: 619–21.

20. Silverman K., Svikis D., Roble E., Stitzer M., Bigelow G. A reinforcement-based therapeutic workplace for the treatment of drug abuse: six-month abstinence outcomes. *Exp Clin Psychopharmacol* 2001; **9**: 14–23.

21. Anker A. L., Crowley T. J. Use of contingency contracts in specialty clinics for cocaine abuse. In: Harris L. S., editor. *Problems of drug dependence: Proceedings of the 43rd Annual Scientific Meeting, the Committee on Problems of Drug Dependence.* NIDA Research Monograph no. 41. Rockville, MD: National Institute on Drug Abuse; 1982, p. 452–9.

22. Crowley T. J. Contingency contracting treatment of drug-abusing physicians, nurses, and dentists. In: Grabowski J., Stitzer M., Henningfeld J. F., editors. *Behavioral Intervention Techniques in Drug Abuse Treatment.* NIDA Research Monograph no. 46. Rockville, MD: National Institute on Drug Abuse; 1984, p. 68–83.

23. Stitzer M., Bigelow G., Liebson I. A. Reinforcement of drug abstinence: a behavioral approach to drug abuse treatment. In: Krasnegor N. A., editor. *Behavioral analysis and treatment of substance abuse*. NIDA Research Monograph no. 25. Rockville, MD: National Institute on Drug Abuse; 1979, p. 68–91.

24. Stitzer M. L., Rand C. S., Bigelow G. E., Mead A. M. Contingent payment procedures for smoking reduction and cessation. *J Appl Behav Anal* 1986; **19**: 197–202.

25. Higgins S. T., Stitzer M. L., Bigelow G. E., Liebson I. A. Contingent methadone delivery: effects on illicit-opiate use. *Drug Alcohol Depend* 1986; **17**: 311–22.

26. Robles E., Stitzer M. L., Strain E. C., Bigelow G. E., Silverman K. Voucher-based reinforcement of opiate abstinence during methadone detoxification. *Drug Alcohol Depend* 2002; **65**: 179–89.

27. Preston K. L., Silverman K., Umbricht A., DeJesus A., Montoya I. D., Schuster C. R. Improvement in naltrexone treatment compliance with contingency management. *Drug Alcohol Depend* 1999; **54**: 127–35.

28. Budney A. J., Higgins S. T., Radonovich K. J., Novy P. L. Adding voucher-based incentives to coping skills and motivational enhancement improves outcomes during treatment for marijuana dependence. *J Consult Clin Psychol* 2000; **68**: 1051–61.

29. Sigmon S. C., Steingard S., Badger G. J., Anthony S. L., Higgins S. T. Contingent reinforcement of marijuana abstinence among individuals with serious mental illness: a feasibility study. *Exp Clin Psychopharmacol* 2000; **8**: 509–17.

30. Elk R., Mangus L., Rhoades H., Andres R., Grabowski J. Cessation of cocaine use during pregnancy: effects of contingency management interventions on maintaining abstinence and complying with prenatal care. *Addict Behav* 1998; **23**: 57–64.

31. Jones H. E., Haug N., Silverman K., Stitzer M., Svikis D. The effectiveness of incentives in enhancing treatment attendance and drug abstinence in methadone-maintained pregnant women. *Drug Alcohol Depend* 2001; **61**: 297–306.

32. Bickel W. K., Amass L., Higgins S. T., Badger G. J., Esch R. Behavioral treatment improves outcomes during opioid detoxification with buprenorphine. *J Consult Clin Psychol* 1997; **65**: 1091–118.

33. Iguchi M. Y., Belding M. A., Morral A. R., Lamb R. J., Husband S. Reinforcing operants other than abstinence in drug abuse treatment: an effective alternative for reducing drug use. *J Consult Clin Psychol* 1997; **65**: 421–8.

34. Petry N. M., Tedford J., Martin B. Reinforcing compliance with non-drug-related activities. *J Subst Abuse Treat* 2001; **20**: 33–44.

35. Higgins S. T., Silverman K. *Motivating Behavior Change Among Illicit-Drug Abusers: Research on Contingency Management Interventions*. Washington, DC: American Psychological Association; 1999.

36. Silverman K. Exploring the limits and utility of operant conditioning in the treatment of drug addiction. *Behav Anal* 2004; **27**: 209–30.

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

38. Lussier J. P., Heil S. H., Mongeon J. A., Badger G. J., Higgins S. T. A meta-analysis of voucher-based reinforcement therapy for substance use disorders. *Addiction* 2006; **101**: 192–203.

39. Dumaine M. L. Meta-analysis of interventions with co-occurring disorders of severe mental illness and substance abuse: implications for social work practice. *Res Soc Work Pract* 2003; **13**: 142–65.

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

41. Prendergast M. L., Podus D., Chang E., Urada D. The effectiveness of drug abuse treatment: a meta-analysis of comparison group studies. *Drug Alcohol Depend* 2002; **67**: 53–72.

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

43. Lipsey M. W. Juvenile delinquency treatment: a meta-analytic inquiry into the variability of effects. In: Cook T. D., editor. *Meta-Analysis for Explanation: a Casebook*. New York: Russell Sage Foundation; 1992, p. 83–127.

44. Moyer A., Finney J. W. Randomized versus nonrandomized studies of alcohol treatment: participants, methodological features, and post-treatment functioning. *J Stud Alcohol* 2002; **63**: 542–50.

45. Tobler N. S. Meta-analysis of adolescent drug prevention programs: results of the 1993 meta-analysis. In: Bukoski W. J., editor. *Meta-analysis of drug abuse prevention programs*. NIDA Research Monograph no. 170. Rockville, MD: National Institute on Drug Abuse; 1997, p. 5–68.

46. Cooper H. M., Hedges L. V. *The Handbook of Research Synthesis*. New York: Russell Sage Foundation; 1994.

47. Hedges L. V., Olkin L. *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press; 1985.

48. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. New York: Academic Press; 1988.

49. Johnson B. T. *DSTAT 1.10: Software for the Meta-Analytic Review of Research Literature*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1989.

50. Lipsey M. W., Wilson D. B. *Practical Meta-Analysis*. Thousand Oaks, CA: Sage Publications; 2001.

51. Duval S., Tweedie R. A non-parametric 'trim and fill' method of accounting for publication bias in meta-analysis. *J Am Statist Assoc* 2000; **95**: 89–98.

52. Duval S., Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000; **56**: 455–63.

53. Duval S. The trim and fill method. In: Rothstein, H., Sutton, A., Borenstein, M., eds. *Publication bias in meta-analysis: prevention, assessment and adjustments.* Hoboken, NJ: John Wiley & Sons, Inc.; 2005, p. 127–44.

54. Prendergast M. L., Podus D., Chang E. Program factors and treatment outcomes in drug dependence treatment: an examination using meta-analysis. *Subst Use Misuse* 2000; **35**: 1931–65.

55. Steichen T. J. Nonparametric 'trim and fill' analysis of publication bias in meta-analysis. *Stata Techn Bull* 2000; **57**: 8–14.

56. Havassy B., Hargreaves W. A., DeBarros L. Self-regulation of dose in methadone maintenance with contingent privileges. *Addict Behav* 1979; **4**: 31–8.

57. Jones H., Haug N., Stitzer M., Svikis D. Improving treatment outcomes for pregnant drug-dependent women using low-magnitude voucher incentives. *Addict Behav* 2000; **25**: 263–7.

58. Melin L., Gotestam K. A contingency management program on a drug-free unit for intravenous amphetamine addicts. *J Behav Ther Exp Psychiatry* 1973; **4**: 331–7.

59. Rosenthal R., Rubin D. B. A note on percent variance explained as a measure of the importance of effects. *J Appl Soc Psychol* 1979; **9**: 395–6.

60. Rosenthal R., Rubin D. B. Comparing effect sizes of independent studies. *Psychol Bull* 1982; **92**: 500–4.

61. Rosenthal R. The 'file drawer problem' and tolerance for null results. *Psychol Bull* 1979; **86**: 638–41.

62. Begg C. B. Publication bias. In: Cooper H., Hedges L. V., editors. *The handbook of research synthesis*. New York: Russell Sage Foundation; 1994, p. 399–409.

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

64. Marsch L. A. The efficacy of methadone maintenance interventions in reducing illicit opiate use, HIV risk behaviors and criminality: a meta-analysis. *Addiction* 1998; **93**: 515–32.

65. Stanton M. D., Shadish W. R. Outcome, attrition, and family-couples treatment for drug abuse: a meta-analysis and review of the controlled, comparative studies. *Psychol Bull* 1997; **122**: 170–91.

66. Epstein D. H., Hawkins W. E., Covi L., Umbricht A., Preston K. L. Cognitive-behavioral therapy plus contingency management for cocaine use: findings during treatment and across 12-month follow-up. *Psychol Addict Behav* 2003; **17**: 73–82.

67. Moyer A., Finney J. W., Swearingen C. E., Vergun P. Brief interventions for alcohol problems: a meta-analytical review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction* 2002; **97**: 279–92.

68. Rawson R. A., Huber A., McCann M., Shoptaw S., Farabee D., Reiber C. *et al.* A comparison of contingency management and cognitive-behavioral approaches during methadone maintenance treatment for cocaine dependence. *Arch Gen Psychiatry* 2002; **59**: 817–24.

69. Dowden C., Andrews D. A. Does family intervention work for delinquents? Results of a meta-analysis. *Can J Criminol Crim Just* 2003; **45**: 327–42.

70. Weisburd D., Lum C. M., Petrosino A. Does research design affect study outcomes in criminal justice? *Ann Am Acad Polit Soc Sci* 2001; **578**: 50–70.

71. Craddock S. G., Roundsbryant J. L., Flynn P. M., Hubbard R. L. Characteristics and pretreatment behaviors of clients entering drug abuse treatment: 1969–93. *Am J Drug Alcohol Abuse* 1997; **23**: 43–59.

72. Gresham F. M., Gansle K. A., Noell G. N., Cohen S., Rosenblum S. Treatment integrity of school-based behavioral intervention studies: 1980–90. *School Psychol Rev* 1993; **22**: 254–72.

73. Salend S. J. Therapy outcome research: threats to treatment integrity. *Behav Modif* 1984; **8**: 211–22.

74. Yeaton W. H., Sechrest L. Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *J Consult Clin Psychol* 1981; **49**: 156–67.

75. Silverman K., Robles E., Mudric T., Bigelow G., Stitzer M. A randomized trial of long-term reinforcement of cocaine abstinence in methadone-maintained patients who inject drugs. *J Consult Clin Psychol* 2004; **72**: 839–54.

76. Azrin N. H., McMahon P. T., Donohue B., Besalel V. A., Lapinski K. J., Kogan E. S. *et al.* Behavior therapy for drug abuse: a controlled treatment outcome study. *Behav Res Ther* 1994; **32**: 857–66.

77. Brooner R. K., Kidorf M., King V. L., Stoller K. Preliminary evidence of good treatment response in antisocial drug abusers. *Drug Alcohol Depend* 1998; **49**: 249–60.

78. Calsyn D. A., Wells E. A., Saxon A. J., Jackson T. R., Wrede A. F., Stanton V. *et al.* Contingency management of urinalysis results and intensity of counseling services have an interactive impact on methadone maintenance treatment outcome. *J Addict Dis* 1994; **13**: 47–63.

79. Carroll K. M., Ball S. A., Nich C., O'Connor P. G., Eagan D. A., Frankforter T. L. *et al.* Targeting behavioral therapies to enhance naltrexone treatment of opioid dependence: efficacy of contingency management and significant other involvement. *Arch Gen Psychiatry* 2001; **58**: 755–61.

80. Chutuape M. A., Silverman K., Stitzer M. Contingent reinforcement sustains post-detoxification abstinence from multiple drugs: a preliminary study with methadone patients. *Drug Alcohol Depend* 1999; **54**: 69–81.

81. Chutuape M. A., Silverman K., Stitzer M. L. Use of methadone take-home contingencies with persistent opiate and cocaine abusers. *J Subst Abuse Treat* 1999; **16**: 23–30.

82. Donatelle R. J., Prows S. L., Champeau D., Hudson D. Randomized controlled trial using social support and financial incentives for high risk pregnant smokers: significant other supporter (SOS) program. *Tob Control* 2000; **9**: 67–9.

83. Downey K. K., Helmus T. C., Schuster C. R. Treatment of heroin-dependent poly-drug abusers with contingency management and buprenorphine maintenance. *Exp Clin Psychopharmacol* 2000; **8**: 176–84.

84. Glasgow R. E., Hollis J. F., Ary D. V., Boles S. M. Results of a year-long incentives-based worksite smoking-cessation program. *Addict Behav* 1993; **18**: 455–64.

85. Glosser D. S. The use of taken economy to reduce illicit drug use among methadone maintenance clients. *Addict Behav* 1983; **8**: 93–104.

86. Gruber K., Chutuape M. A., Stitzer M. L. Reinforcement-based intensive outpatient treatment for inner city opiate abusers: a short-term evaluation. *Drug Alcohol Depend* 2000; **57**: 211–23.

87. Hall S. M., Bass A., Hargreaves W. A., Loeb P. Contingency management and information feedback in outpatient heroin detoxification. *Behav Ther* 1979; **10**: 443–51.

88. Higgins S. T., Budney A. J., Bickel W. K., Badger G. J., Foerg F. E., Ogden D. Outpatient behavioral treatment for cocaine dependence: one-year outcome. *Exp Clin Psychopharmacol* 1995; **3**: 205–12.

89. Higgins S. T., Wong C. J., Badger G. J., Ogden D. E., Dantona R. L. Contingent reinforcement increases cocaine abstinence during outpatient treatment and 1 year of follow-up. *J Consult Clin Psychol* 2000; **68**: 64–72.

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

91. Jason L. A., Salina D., McMahon S. D., Hedeker D., Stockton M. A worksite smoking intervention: a 2 year assessment of groups, incentives and self-help. *Health Educ Res* 1997; **12**: 129–38.

92. Jason L. A., Jayaraj S., Blitz C. C., Michaels M. H., Klett L. E. Incentives and competition in a worksite smoking cessation intervention. *Am J Public Health* 1990; **80**: 205–6.

93. Kassaye W. W. Assessment of the impact of monetary incentives upon smoking behavior. *J Drug Educ* 1984; **14**: 109–18.

94. Katz E. C., Chutuape M. A., Jones H. E., Stitzer M. L. Voucher reinforcement for heroin and cocaine abstinence in an outpatient drug-free program. *Exp Clin Psychopharmacol* 2002; **10**: 136–43.

95. Kidorf M., Stitzer M. L. Contingent access to methadone maintenance treatment: effects on cocaine use of mixed opiate-cocaine abusers. *Exp Clin Psychopharmacol* 1993; **1**: 200–6.

96. Koffman D. M., Lee J. W., Hopp J. W., Emont S. L. The impact of including incentives and competition in a workplace smoking cessation program on quit rates. *Am J Health Prom* 1998; **13**: 105–11.

97. Lando H. A. Aversive conditioning and contingency management in the treatment of smoking. *J Consult Clin Psychol* 1976; **44**: 312.

98. McCarthy J. J., Borders O. T. Limit setting on drug abuse in methadone maintenance patients. *Am J Psychiatry* 1985; **142**: 1419–23.

99. McCaul M. E., Stitzer M. L., Bigelow G. E., Liebson I. A. Contingency management interventions: effects on treatment outcome during methadone detoxification. *J Appl Behav Anal* 1984; **17**: 35–43.

100. Milby J. B., Garrett C., English C., Fritschi O., Clarke C. Take-home methadone: contingency effects on drug-seeking and productivity of narcotic addicts. *Addict Behav* 1978; **3**: 215–20.

101. Milby J. B., Schumacher J. E., McNamara C., Wallace D., Usdan S., McGill T. *et al.* Initiating abstinence in cocaine abusing dually diagnosed homeless persons. *Drug Alcohol Depend* 2000; **60**: 55–67.

102. Piotrowski N. A., Tusel D. J., Sees K. L., Reilly P. M., Banys P., Meek P. *et al.* Contingency contracting with monetary reinforcers for abstinence from multiple drugs in a methadone program. *Exp Clin Psychopharmacol* 1999; **7**: 399–411.

103. Preston K. L., Umbricht A., Epstein D. H. Methadone dose increase and abstinence reinforcement for treatment of continued heroin use during methadone maintenance. *Arch Gen Psychiatry* 2000; **57**: 395–404.

104. Preston K. L., Umbricht A., Epstein D. H. Abstinence reinforcement maintenance contingency and one-year follow-up. *Drug Alcohol Depend* 2002; **67**: 125–37.

105. Shoptaw S., Rotheram-Fuller E., Yang X., Frosch D., Nahom D., Jarvik M. E. *et al.* Smoking cessation in methadone maintenance. *Addiction* 2002; **97**: 1317–28.

106. Silverman K., Higgins S. T., Brooner R. K., Montoya I. D., Cone E. J., Schuster C. R. *et al.* Sustained cocaine abstinence in methadone maintenance patients through voucher-based reinforcement therapy. *Arch Gen Psychiatry* 1996; **53**: 409–15.

107. Silverman K., Wong C. J., Umbricht-Schneiter A., Montoya I. D., Schuster C. R., Preston K. L. Broad beneficial effects of cocaine abstinence reinforcement among methadone patients. *J Consult Clin Psychol* 1998; **66**: 811–24.

108. Stitzer M. L., Bigelow G. E. Contingent reinforcement for reduced breath carbon monoxide levels: target-specific effects on cigarette smoking. *Addict Behav* 1985; **10**: 345–9.

109. Stitzer M. L., Iguchi M. Y., Felch L. J. Contingent take-home incentive: effects on drug use of methadone maintenance patients. *J Consult Clin Psychol* 1992; **60**: 927–34.