The Effects of Behavioral/Cognitive-Behavioral Programs on Recidivism

Frank S. Pearson
Douglas S. Lipton
Charles M. Cleland
Dorline S. Yee

The CDATE project coded studies of treatment/intervention programs in prison, jail, probation, or parole settings reported from 1968 through 1996. Meta-analyses were conducted on the 69 primary research studies on the effectiveness of behavioral and cognitive-behavioral treatment in reducing recidivism for offenders. Results on this heterogeneous collection of studies show that this treatment is associated with reduced recidivism rates. However, this effect is mainly due to cognitive-behavioral interventions rather than to standard behavior modification approaches. The specific types of programs shown to be effective include cognitive-behavioral social skills development programs and cognitive skills (Reasoning and Rehabilitation) programs.

In 1994 the National Institute on Drug Abuse (NIDA) funded the Correctional Drug Abuse Treatment Effectiveness (CDATE) project for 4 years to develop a database and conduct meta-analyses on correctional treatment evaluation studies completed between January 1, 1968 and December 31, 1996. Researchers have, over the past 20 years, begun to use meta-analysis to assess the research literature on corrections-based treatment programs. Glass, McGaw, and Smith (1981, p. 21) distinguish primary analysis (the original research study's statistical analysis of data), secondary analysis (a subsequent study's reanalysis of the data from the original study, usually with better statistical techniques or to ask new questions of the original data), and

FRANK S. PEARSON: National Development and Research Institutes, Inc. **DOUGLAS S. LIPTON**: National Development and Research Institutes, Inc. (retired). **CHARLES M. CLELAND**: National Development and Research Institutes, Inc. **DORLINE S. YEE**: National Development and Research Institutes, Inc.

This project was supported under grant number R01 DA08607 from the National Institute on Drug Abuse at the National Institutes of Health, U.S. Department of Health and Human Services. Points of view in this document are those of the authors and do not necessarily reflect the official position of any U.S. Government agency. We are very grateful to the many individuals who worked on the CDATE project over the years, particularly those who did the initial coding of the studies. We particularly want to thank Dr. Robert Rosenthal who provided sound advice to us over the years in his role as meta-analysis consultant on the project.

CRIME & DELINQUENCY, Vol. 48 No. 3, July 2002 476-496

meta-analysis (the statistical analysis of the summary findings of several independent research studies—a few investigators using these methods prefer the term *quantitative research synthesis* to *meta-analysis*).

Garrett (1985) conducted meta-analyses on studies (reported between 1960 and 1983) of the effects of residential treatment on adjudicated delinquents in terms of various outcome measures (e.g., psychological adjustment). She analyzed 43 effect sizes with recidivism as the outcome for behavioral treatment as one of four general treatment categories[1] from 16 separate studies and found a Glass's Δ effect size of +.18 (this approximates a correlation coefficient of $r = +.09$). Specific subcategories within behavioral treatment were as follows: (a) contingency management, (b) cognitive behavioral, (c) guided-group interaction/positive peer culture, and (d) milieu therapy. (Many investigators would not have coded the last two treatment types as behavioral/cognitive-behavioral.) Garrett reported that the positive effect of behavioral approaches was attributable primarily to the cognitive-behavioral and contingency-management programs (the effect size for each would approximate a correlation of $r = +.12$). However, when she focused her analysis on the six studies with more rigorous designs, there was no average positive effect at all ($r = -.04$).

At about the same time, Gottschalk, Davidson, Mayer, and Gensheimer (1987) analyzed recidivism as the outcome variable for behavioral treatment programs in studies published from 1967 through 1983 having a comparison group as well as a treatment group. In their words, "Further, only studies that specifically described the use of behavioral procedures were included, for instance, those studies reporting the systematic application of behavioral procedures such as contingency management, positive reinforcement, token economies, behavioral contracts, or modeling" (p. 406). They coded 14 separate studies and found a mean unweighted Glass's Δ effect size of +.25. Because the 95% confidence interval included zero, however, they did not conclude that the effectiveness of these types of programs was verified.

Izzo and Ross (1990) conducted a meta-analysis of studies published in refereed journals from 1970 to 1985 dealing with the effects on recidivism produced by correctional treatment programs for juveniles. A total of 46 studies met their criteria, and they coded a total of 68 comparisons from those studies. They found that only two variables were related to successful recidivism outcomes: programs having been conducted in a community setting (rather than in an institution) and cognitive treatment. Programs were coded as cognitive if they included such modalities as problem solving, negotiation, skills training, interpersonal skills training, rational-emotive therapy, role-playing and modeling, or cognitively mediated behavior modification. They

found that cognitive programs were more than twice as effective as noncognitive programs.

It should be pointed out that the primary research reports, which provide the "data" for the meta-analyses, typically present terms referring to types of behavioral and cognitive-behavioral treatments without definition, assuming that the reader has a background understanding of what is meant. Consequently, different meta-analysts might disagree on the way a particular treatment in a study has been coded. CDATE's definitions of behavioral interventions and specific subtypes of behavioral and cognitive-behavioral interventions are presented below in the Behavioral and Cognitive-Behavioral Therapy and the Measurement of Variables sections and assessments of reliability are provided in the Quality Control and Reliability section.

Whitehead and Lab (1989) conducted a meta-analysis of 50 studies published in professional journals from 1975 to 1984 dealing with the effects on recidivism produced by correctional treatment programs for juveniles. Their overall conclusion was that "interventions have little positive impact on recidivism and many appear to exacerbate the problem" (1989, p. 276). When they focused on behavioral interventions, including contingency management, skills training, behavior contracting, and token economies, their conclusion was that "behavioral interventions fare no better than other types of treatment at reducing recidivism for their experimental clients as compared to control group subjects" (p. 286).

Andrews et al. (1990) followed with a meta-analysis of 45 of the 50 studies Whitehead and Lab (1989) had included plus another 35 studies (including adult programs) published from 1950 through 1989. They computed 154 effect sizes from these 80 studies and found that appropriate correctional treatment did reduce recidivism; appropriateness was defined as (a) delivery of treatment to higher risk cases, (b) targeting of criminogenic needs, and (c) use of styles and modes of treatment (e.g., cognitive and behavioral) matched with and responsive to client needs and learning styles. Among their specific findings, they mentioned that the 41 correlation coefficient effect sizes for behavioral interventions they found (in 23 studies) yielded a significantly greater mean effect size ($r = .29$) than did nonbehavioral treatment ($r = .04$).[2] Several years afterward, in reporting on their expanded meta-analytic database, it was noted that the "short-hand phrase 'behavioral' may be better described as 'behavioral/social learning/cognitive behavioral' " and that the 60 effect sizes now recorded continue to show significantly better outcomes than nonbehavioral treatment (Andrews & Bonta, 1998, p. 270).

Lipsey (1992) conducted a meta-analysis of reports from 1950 through 1987, including unpublished as well as published documents, on evaluations

of treatment programs for juveniles (including programs to prevent youth from becoming delinquent as well as programs to treat adjudicated delinquents) yielding 443 effect sizes. Of these, 24 effect sizes were for behavioral therapy. However, the effectiveness of behavioral and cognitive-behavioral interventions was not a focus of this study. Lipsey's report does mention that the cluster including behavioral, skill-oriented, and multimodal treatment "was associated with larger effect sizes than other treatment approaches" (p. 120).

THE CDATE PROJECT

In 1994 the NIDA funded the CDATE project for 4 years to develop a comprehensive information database of correctional treatment evaluation studies appearing in published and unpublished research reports between January 1, 1968 and December 31, 1996. CDATE coded over 2,176 research comparisons of experimental groups with comparison groups assessing the impact of the various interventions on several outcome measures, particularly drug abuse and recidivism. Of these, 69 were of programs in which the most important treatment was behavioral or cognitive behavioral.

BEHAVIORAL AND COGNITIVE-BEHAVIORAL THERAPY

We consider behavioral and cognitive-behavioral therapy as a general category comprising two subcategories of treatments: (a) behavior modification/behavior therapy and (b) cognitive-behavioral treatments. Behavior modification/behavior therapy focuses on arranging contingencies of positive reinforcement to develop and maintain appropriate patterns of behavior (Bandura, 1969; Skinner, 1953). In addition to standard behavior modification procedures, we included contingency contracting: A contract is written under which specific desirable behaviors by the client earn specific rewards. (In some contingency contracts there are also punishments specified for specific named undesirable behaviors by the client.[3]) Also included in the behavior modification/behavior therapy subcategory were token economies (i.e., contingencies of reinforcements applied to groups of people—such as inmates in a prison dormitory) whereby specific desirable behaviors (e.g., cleaning a particular area) earn tokens that can be exchanged later for goods or privileges.

The other subcategory includes cognitive-behavioral treatments, treatments that include attention to cognitive and emotional processes that function between the stimuli received and the overt behaviors enacted. For example, social learning theory is broader than behaviorist reinforcement theory because it includes as variables cognitions, verbalizations, and social modeling to explain (and to change) behavior patterns (Akers, 1977).

As McGuire (1996) points out, "there is no single cognitive-behavioural method or theory. Work of this kind is best thought of as a 'family' or collection of methods rather than any single technique easily and clearly distinguished from others" (p. 7). McGuire includes in his discussion of cognitive-behavioral approaches: social skills training (which uses modeling, role-play practice, and feedback), social problem-solving training, rational-emotive therapy, the cognitive skills program (also known as the Reasoning and Rehabilitation program), and the relapse prevention model (pp. 42-49, 58-59, 65, 105-106). He also mentions aggression replacement training for violent offenders as a program with cognitive-behavioral components (p. 106).

Hypotheses

The hypotheses presented here are limited to treatments rated as the most important treatment modality in a program.

1. Behavioral/cognitive-behavioral programs (a broad category encompassing programs ranging from token economy programs up to cognitive-behavioral programs) are more effective than treatment-as-usual comparison group interventions at reducing recidivism.
   - 1.1. Behavioral reinforcement and incentive programs are more effective than treatment-as-usual comparison group interventions at reducing recidivism.[4]
   - 1.2. Cognitive-behavioral programs are more effective than treatment-as-usual comparison group interventions at reducing recidivism.[5]

METHOD

Searching for Primary Research Studies

CDATE's methods of searching for primary research studies included searches of 24 computerized bibliographic databases, searches of bibliographies listed in books and articles, hand searches of 17 major journals within the 1968-1996 period, searches of the books and monographs available at several large libraries.[6] Another search method involved requests for documents from authors and organizations.

Criteria for Inclusion and Exclusion

CDATE included a study if it (a) used behavioral/social science research methods, (b) examined the effect or impact of any rehabilitation/intervention program, (c) had as subjects either adult or juvenile offenders in custody (whether in a residential correctional facility or on probation or parole), (d) used any outcome measures (including, but not restricted to, effects on drug use and recidivism), (e) was in any language and from any country, (f) was either published or unpublished, (g) was produced between January 1, 1968, and December 31, 1996. CDATE coded 2,176 research comparisons of experimental groups with comparison groups; of these, 1,606 were distinct studies or substudies (i.e., independent comparisons). For the analyses presented in this article, two more conditions also had to be met: (h) the CDATE researchers had identified a behavioral or cognitive-behavioral treatment type as the most important treatment of the experimental program, and (i) there was information about the outcomes of the study available in the report that made it possible to compute an effect size with recidivism as the dependent variable. There were 69 independent comparisons meeting these conditions.

Measurement of Variables

The numbers and specific types of behavioral/cognitive-behavioral programs identified in CDATE as independent comparisons of an experimental group and comparison group are as shown in Table 1.

Within the general category of behavioral/cognitive-behavioral programs, the specific subcategories coded were as follows:

Behavioral.

- Standard behavior modification. (Arranging contingencies of positive reinforcement to develop and maintain appropriate patterns of behavior.)
- Contingency contracting. (System in which the offender signs a contract with the person supervising him or her in which specific behaviors by the offender are linked with particular punishments, e.g., a stricter curfew for a positive urine test, and other behaviors are linked with rewards, e.g., good time credits for satisfactory work performance.)
- Token economy. (A reinforcement system in which offenders/inmates who perform specific behaviors satisfactorily, e.g., cleaning their living area, helping other inmates, etc., are rewarded with tokens, which can later be exchanged for privileges, e.g., more time to watch television, or desired goods, e.g., snacks from the canteen.)

TABLE 1: All Behavioral or Cognitive-Behavioral Programs

Program	*Count*
Behavioral	
Standard behavior modification	5
Aversive-conditioning focused	1
Contingency contracting	13
Token economy	6
	(subtotal = 25)
Cognitive behavioral	
Social skills–development training	14
Problem-solving skills training	1
Cognitive skills training	7
Thinking errors approach	2
Other social skills training	2
Social-learning focused	3
Cognitive behavioral	10
Self-control training	1
Training in anger management	1
Relapse prevention model	3
	(subtotal = 44)
Total	69

Cognitive Behavioral

These are counseling/training programs to develop one or more important cognitive skills that are important to avoid serious problems or deficits in society.

- Social skills development training (e.g., developing skills in communication, giving and receiving positive and negative feedback, assertiveness, conflict resolution, etc.).
- Problem solving skills training (e.g., D'Zurilla & Goldfried, 1971).
- Cognitive skills training, such as Ross, Fabiano, and Diemer-Ewles's (1988) program materials on cognitive skills and/or reasoning and rehabilitation.
- Thinking errors approach (e.g., Yochelson & Samenow, 1977).
- Other social skills training.
- Social learning focused. (These approaches include as variables cognitions, verbalizations, and social modeling to explain—and to change—behavior patterns, e.g., Akers [1977].)
- Cognitive behavioral. (Other counseling/training programs that teach self-reinforcement, self-instruction, self-rehearsal, role-taking, self-control, and problem solving.)
- Self-control training.
- Training in anger management. (Training in anger management or aggression management.)

- Relapse prevention. (Programs preparing the offender to deal with cravings, peer pressure, etc., to prevent relapse to the illicit behavior,[8] e.g., Marlatt and Gordon, 1985.)

In the analyses presented here, programs were considered to be of the relevant treatment type if that treatment was rated by CDATE coders as the most important treatment difference between the experimental and the comparison groups. Because a large percentage of studies in this field have features of the research methods used that seem likely to bias the results in favor of the experimental group, we include a rating of the overall quality of the research methods used in the study as a (potential) moderator variable. The dependent (outcome) variable in these studies was recidivism, consisting mainly of rearrest and/or reincarceration. When findings were adequately reported for both recidivism measures, the measure used in the CDATE meta-analysis was arrest because it is procedurally and temporally closer to the crime event (Maltz, 1984, pp. 138 ff).

Quality Control and Reliability

In the standard coding procedure that was followed for all studies, documents were assigned to one individual coder for a full coding. Once a full, initial coding was produced, documents for a given study proceeded to a "quality controller" who reviewed the initial codings and documents. If necessary, changes were made to the initial codings by the quality controller, and reasons for the changes were discussed with the original coder. More important, only the final quality controlled codings for each study were included as observations in the CDATE data set.

In the reliability assessment, both of the quality controllers had the study documents and the initial codings given to them by their respective initial coders, but each remained unaware of codings and comments made by the other initial coder and quality controller. The analysis of intercoder reliability was based on the second quality controller coding the same set of 30 studies as the first quality controller, producing 30 pairwise comparisons for each variable coded. The rationale for estimating the intercoder reliability between the project's two quality controllers was based on the fact that the final database used for meta-analyses only included records that had undergone a thorough check by the quality controllers.

Reliabilities of the variables coded as "raw material" for an effect size, r, were quite satisfactory: reliability, $r = .99$, suggesting very high intercoder reliability on calculated effect sizes. Also, coders agreed on the specific type of outcome in 87% of the 30 comparisons (further details are available from the authors). High reliability was found for the length of treatment in days

and the total sample size. The percentage agreement on treatment categories coded as the most important treatment type (e.g., cognitive behavioral) was 87%. The percentage agreement on the exact type of treatment coded as the most important specific type of correctional treatment/intervention was 70%. The reliability for the rating of the overall quality of research methods used in the study was adequate: the Pearson coefficient *r* was .62.

Methods of Statistical Analysis

Independence of comparisons in these analyses. The vast majority of studies in this body of research literature report only independent comparisons in which a particular experimental group is analytically compared with just one comparison group. This is fortunate because there are important statistical problems to grapple with when one experimental group is analytically compared with more than one comparison group or when one comparison group is compared with more than one experimental group. For example, if a study used two experimental groups, E1 and E2, and two comparison groups, C1 and C2, and reported analytical comparisons among all combinations, E1 vs. C1, E1 vs. C2, E2 vs. C1, and E2 vs. C2, the four comparisons are not independent. When confronted with nonindependent comparisons, CDATE tried to pick the fairest, most pertinent comparison group relative to each particular experimental group and used those independent comparisons (e.g., E1 vs. C1 and E2 vs. C2). We also concentrated on comparisons that the treatment group would naturally be expected to surpass, that is groups receiving only treatment as usual or treatment thought to be irrelevant to the outcome variable under focus.

Effect size estimates. The two most common measures of the size of the effect that an intervention has are Pearson's correlation coefficient *r* (including its variations such as the phi coefficient) and Hedges's *g*. Fortunately, there are formulas that allow transformation from one metric to the other. We use *r* as the effect size, in part because it is widely known and thus easily interpretable. To deal with departures from normality, however, meta-analysts prefer to carry out certain statistical operations on the Fisher's Z_r transformation. Following that advice, *r* is first transformed to Z_r before doing the statistical computations; after obtaining the results, the transformation is reversed to display the results in terms of *r* itself (Rosenthal, 1994).

Statistical procedures. CDATE used the random-effects, inverse-variance weighted regression approach in these meta-analyses. The advantages of using regression models in research synthesis have been noted by Hedges and

Olkin (1983, 1985). Because effect size estimates are more reliable if they are based on larger primary research samples, studies are weighted in relation to the number of subjects used to test a given hypothesis (using the inverse-variance method), and these weighted effect sizes are used in the regression analyses. In the random-effects model, the assumption is that there is not one true effect size, but a distribution of effect sizes. True effect sizes vary "precisely because sources of influence on the outcome are both numerous and unidentifiable" (Raudenbush, 1994, p. 302). The random-effects model "permits generalization to other studies from the same population from which the retrieved studies were sampled" (p. 187). The analyses were done using hierarchical linear modeling (HLM) software (Bryk & Raudenbush, 1992; Bryk, Raudenbush, & Congdon Jr., 1996). This was preceded by exploratory data analysis with graphical displays to search for outliers.

RESULTS

As background for the meta-analyses, Figure 1 shows a correlation coefficient, *r*, effect size with recidivism as the outcome variable for all 69 studies (undifferentiated by specific treatment type). Each *r* is plotted in the middle of a vertical bar representing the 95% confidence interval for that *r*. The studies are sorted by the overall rating of the research methods used in the study (1 = *poor, barely acceptable, very low confidence, 16 studies*; 2 = *fair, a low level of confidence, 29 studies*; 3 = *good, a mid-level of confidence, 17 studies*; 4 = *excellent, a high level of confidence, 7 studies*) from left to right on the graph and, within that rating, by the favorability of the outcome. The study with the largest effect size ($r = +.81$) appears to be an outlier that does not seem to "belong" with this collection of studies. This was a study with especially weak methods: It did not have a separate set of subjects to serve as a comparison group, so it used, for the experimental group only, arrest rates before and after the program. We exclude this study (Davidson & Robinson, 1975) from all subsequent analyses.[9]

There is an association between the research design used and the CDATE rating of the overall quality of the research methods: The poor studies typically used classical or quasi-experimental research designs and the good and excellent studies typically used true experimental designs. However, the preference in CDATE was not to use research design itself as the moderator variable because there are many flaws it does not capture (e.g., preexisting differences between the experimental and comparison subjects that can persist in any research design, differential attrition of research susbjects, biases in the statistical analyses conducted, etc.). The overall flaws (and strengths)

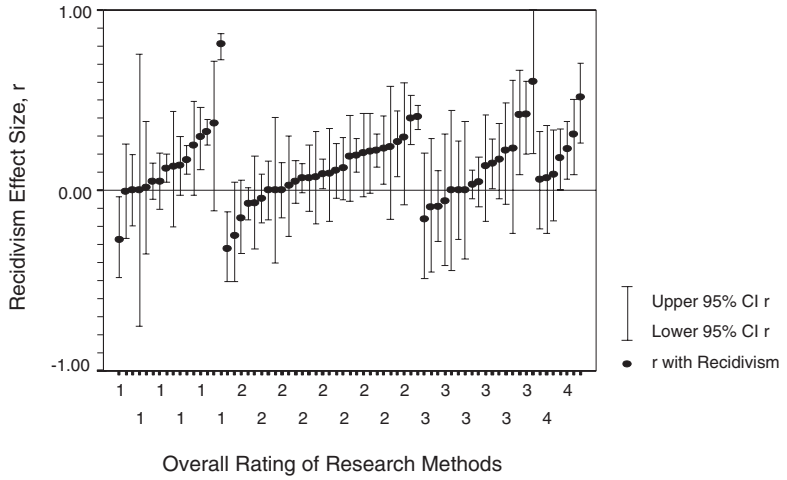

Figure 1: Recidivism Effect Sizes for 69 Behavioral or Cognitive-Behavioral Studies Sorted by Overall Rating of Research Methods

of the research are more likely to be accurately reflected in the rated quality of research methods variable.

We considered an hypothesis to be verified when the inverse-variance-weighted mean effect size, *r*, was greater than or equal to .05 (with no clear evidence of research method artifact) and the *t* test resulted in a one-tailed probability less than .05.[10] If *r* was less than .05 we considered the hypothesis to be disconfirmed. There is a conversion relationship, the Binomial Effect Size Display (BESD), which provides some indication of the practical importance of the effect size (Rosenthal, 1991, pp. 132-136). A BESD relates a Pearson correlation coefficient, *r*, to a percentage differential between the experimental and comparison group, using 50% as a midpoint anchor. For example, a correlation of $r = .05$ can be thought of as the experimental group being 5 percentage points better than the comparison group, using 50% as a midpoint anchor. Thus, the BESD would be 52.5% successes in the experimental group versus 47.5% successes in the comparison group. We set this as our minimal criterion of practical significance.

In our assessments there was also a possibility for results to fall in a borderline "gray area." One reason for this designation could be that, although the *r*, was greater than or equal to .05, the *t* test did not result in a one-tailed probability less than .05. Another reason could be that there are indications of a research method artifact in which the poorer quality studies show a substan-

TABLE 2: Hypothesis 1: Recidivism by All Behavioral or Cognitive-Behavioral Programs

Verified?	*Yes*
Number of studies, *k*	68
Total *N*, Windsorized	10,428
Weighted mean of *r*	.118
One-tailed probability	0.0000003
Homogeneity	0
Method rating beta	.012

Method Rating	k *Studies*	*Mean* r	*Median* r	SD
Excellent	7	.207	.177	.165
Good	17	.119	.046	.209
Fair	29	.091	.092	.173
Poor	15	.110	.123	.164
Total	68	.114	.092	.179

NOTE: This test excludes the outlier study, Davidson and Robinson (1975).

tial effect whereas the better quality studies do not show any substantial effect.

Analyses relating to Hypothesis 1, the effectiveness of all behavioral/cognitive-behavioral programs combined, are presented in Table 2. The number of studies (here, $k = 68$) refers to the number of independent comparisons (one experimental group relative to one comparison group. The total *N* refers to the number of individual subjects (persons) in the experimental plus comparison groups of each study, Windsorized at 600.[11] The summary table also includes the mean inverse-variance-weighted Pearson correlation coefficient,[12] *r*, and the null-hypothesis exact probability associated with the *t* test used.

There are three conventionally used indicators of homogeneity: (a) at least 75% of the observed variance is accounted for by sampling error, (b) the "Q" chi-square test is not significant, and (c) the amount of residual variance is less than 25% of the estimated population effect size (Hunter, Schmidt, & Jackson, 1982; Schwarzer, 1989). The entry for homogeneity in the summary tables designates how many of these three criteria indicate homogeneity. As expected, the general category of behavioral and cognitive behavioral studies was not homogeneous; none of the three criteria indicated homogeneity.

The "Method Rating beta" entry in the table gives the inverse-variance-weighted hierarchical linear regression beta coefficient for the overall rating of research methods. The beta coefficient of +.012 indicates that the effect sizes show a very slight (negligible) positive linear relationship with the quality of research methods used. That there is no research method artifact can

TABLE 3: Hypothesis 1.1: Recidivism by Behavioral Reinforcement/Incentive Programs

Verified?	*Borderline*
Number of studies, *k*	23
Total *N*, Windsorized	1,935
Weighted mean of *r*	.066
One-tailed probability	.0686
Homogeneity	0
Method Rating beta	.042

Method Rating	k *Studies*	*Mean* r	*Median* r	SD
Excellent	4	.171	.159	.117
Good	3	.181	.000	.366
Fair	11	.048	.074	.201
Poor	5	–.017	.000	.154
Total	23	.073	.060	.206

NOTE: This test excludes the aversive conditioning study and the outlier study, Davidson and Robinson (1975).

also be seen in the lower section of Table 2, in which the unweighted correlation effect sizes are shown within each of the research method rating categories.

Table 3 summarizes the results of the hypothesis test for the subcategory of treatments comprising behavioral programs, including contingency contracting, token economy programs, and other standard behavior modification programs.[13] Although the result was not quite statistically significant at the .05 level, the inverse-variance weighted mean *r* of .066 does exceed our (rather lenient) .05 criterion.[14] The corresponding BESD is 53.3% success in the experimental groups and 46.7% success in the comparison groups. The research method rating moderator variable showed a slight positive linear relationship with the effect sizes (beta = +.042), indicating that the better-quality studies found slightly larger effect sizes, on the average. In the detail method rating panel in the lower section of the same table the studies rated as excellent and good have unweighted mean *r*s of .17 and .18. There are two weaknesses, however, that should be pointed out. First, post hoc exploration of the cases revealed that three of the four excellent entries are actually three independent comparisons from one overarching study (Davidson, Redner, & Amdur, 1990). Second, the median unweighted *r* in the good category is .00. Therefore, we are unwilling to state that the effectiveness of behavioral reinforcement/incentives programs has been confirmed or that it has been disconfirmed but, rather, we characterize it as being on the borderline of verified effectiveness.

TABLE 4: Hypothesis 1.2: Recidivism by Cognitive-Behavioral Programs

Verified?			*Yes*	
Number of studies, *k*			44	
Total *N*, Windsorized			8,435	
Weighted mean of *r*			.144	
One-tailed probability			0.0000002	
Homogeneity			0	
Method Rating beta			–0.003	
Method Rating	k *Studies*	*Mean* r	*Median* r	SD
Excellent	3	.254	.177	.234
Good	14	.106	.091	.179
Fair	17	.129	.093	.149
Poor	10	.173	.150	.133
Total	44	.140	.127	.161

As Table 4 summarizes, our meta-analyses showed the other subcategory (i.e., cognitive-behavioral treatments) to be effective in reducing recidivism. For the 44 studies the weighted mean *r* was .144. The corresponding BESD is 57.2% successes in the experimental group versus 42.8% successes in the comparison group. The correlation coefficients are above .05 in both the good and excellent research method rating categories. However, as the summary table showed, these 44 studies are not statistically homogeneous, suggesting that some moderator variable or variables might further partition the studies into those with relatively stronger effect sizes and those with weaker effect sizes.

After seeing the results of the above three hypothesis tests, the following post hoc, exploratory analyses were conducted. First, because there was enough overlap in the cognitive-behavioral studies by age, this variable was explored as a possible moderator variable. However, the results did not differ by age for the cognitive-behavioral studies. Still restricting the analyses to independent comparisons, the mean *r* for juveniles (19 effect sizes) is .14 and that for adults (25 effect sizes) is also .14.

Next, specific treatment types were explored. Within behavioral and incentives treatments, there were five or more studies in each of the three specific types of treatment (i.e., contingency contracting, token economies, and other standard behavior modification), but none of these specific behavioral treatments could be verified (post hoc) as effective in reducing recidivism.

Within the cognitive-behavioral subcategory, only three specific types of treatment included five or more studies: social skills development training, cognitive skills training, and studies coded as other cognitive-behavioral pro-

grams. Based on 14 primary research studies, social skills development programs do meet our criteria as verified effective in reducing recidivism. The inverse-variance-weighted mean *r* is .17 ($p = .0036$), which corresponds to a BESD of 58.5% success in the experimental group and 41.5% in the comparison group. The lone study rated as having excellent methods had an $r = .07$, and the seven studies rated as having good methods had a mean $r = .16$.

There are only seven research studies dealing with the cognitive skills programs developed by Ross and his colleagues (1988) (also known as Reasoning and Rehabilitation programs), but they still meet our criteria of verified effectiveness. The weighted mean *r* is .147, which corresponds to a BESD of 57.4% successes in the experimental group and 42.7% successes in the comparison group. The one study rated as having excellent methods had an $r = .52$, and the three studies rated as having good methods had a mean $r = .15$.

Last, there are 10 primary research studies of other programs CDATE researchers coded as specifically cognitive behavioral.[15] Simply relying on our criteria for verification, these studies would be considered effective. The weighted mean *r* of .114 corresponds to 55.7% successes in the experimental and 44.3% successes in the comparison groups. The method rating beta for this set of studies is just about zero. However, there were no excellent quality studies and only one good study—and that study has a negative effect size ($r = –.16$). At this time, these other cognitive-behavioral programs might be considered to be in a "gray area" of verification.

DISCUSSION

Although Garrett (1985) and Gottschalk et al. (1987) found that they could not reject the null hypothesis for behavioral/cognitive-behavioral programs and recidivism, these constituted the first wave of meta-analyses in this area that, necessarily, relied on only the 16 or so studies available to them. On the one hand, the CDATE meta-analyses drawing from a total of 69 independent comparisons are in line with the general findings of Andrews et al. (1990) in support of "behavioral/social learning/cognitive behavioral" treatment (as a general category). The CDATE findings confirm and expand Izzo and Ross's (1990) meta-analysis of cognitive-behavioral treatment (as a specific category of treatment), showing that cognitive-behavioral programs can reduce recidivism rates by significant amounts. This was found to be true for the overall collection of cognitive-behavioral studies and also for the subcategories social skills development training and cognitive skills training. On the other hand, the CDATE meta-analyses, like that of Whitehead and Lab

(1989), did not allow us to reject the null hypothesis for contingency contracting, token economies, and other standard behavior modification as effective in reducing recidivism (see following).

We present these results with confidence. However, it would be prudent to bear in mind two practical problems. First, some program developers who refer to their programs as "cognitive behavioral" may not be thinking of the elements used in the cognitive-behavioral models discussed in the studies cited here. (Indeed, the subcategory we referred to as other cognitive-behavioral programs was not convincingly verified because not enough good quality research has been conducted.) If programs do not incorporate the types of treatment found in the primary research studies examined here, they may not be as effective. Second, even when the models intended are the same as those discussed here, some program directors may find themselves unable to implement adequately the cognitive-behavioral model in their particular correctional program; for example, they may not be able to obtain the right kind of training in cognitive-behavioral methods for their treatment staff. Naturally, to the extent that implementation of cognitive-behavioral programming may fall short, the results may fall short as well.

These cautions lead to a call for more research on cognitive-behavioral programs to provide more specific information needed about the programming and its effects. The next wave of research should describe the details of the cognitive-behavioral programs provided to the clients, including the specifics of the treatment models and curricula being used, the training and credentials of the treatment staff, how frequent the treatment sessions are, information on supervision procedures to insure that the quality of the treatment provided is maintained, and the planned and actual time in the program for the clients. Research is also needed to investigate how effective cognitive-behavioral modules are in the context of other treatment modalities. How effective is cognitive-behavioral treatment in the context of a therapeutic community or in the context of a 12-Step program?

Before analyzing these studies, we expected not only the cognitive-behavioral programs but also the behavioral reinforcement/incentives programs to be verified effective. Based on the meta-analyses, we think that the behavioral reinforcement/incentives programs (i.e., standard behavior-modification programs, token economies, and contingency-contracting programs) should at this time still be viewed as neither confirmed nor disconfirmed. Some might find this surprising becauase the behavioral reinforcement approach has been shown in many good laboratory studies to control subhuman and human behavior patterns. In retrospect, we think that programs that focus on contingencies of reinforcement will be verified effective in establishing and increasing desirable behaviors by the offenders, while the

contingencies of reinforcement are operating (a straightforward principle of operant conditioning). However, the standard behavioral model includes the idea that when the contingencies of reinforcement are no longer kept operating, the targeted desirable behaviors are likely to decrease in frequency ("extinguish") and old patterns of behavior reestablish themselves. Thus, if the necessary contingencies of reinforcement are not in effect after the program, the clients are likely to resume committing undesirable behaviors. (It is common in these studies for the time at risk for recidivism to extend from 6 months to 2 years after the contingencies of reinforcement program has ceased.) Several of the studies allude to this behavioral reality:

> The question remains whether concentrated training in specific behaviors can be transferred to the wide variety of situations that confront children after discharge. (Handler, 1974, p. 15)
>
> The immediate efficacy of its behavior-modification techniques is verified. However, follow-up results . . . indicate the failure of the program to produce desirable social outcomes for discharged youth. (Davidson & Wolfred, 1977, p. 296)
>
> Post-treatment persistence of appropriate social behavior does not follow naturally from effective control over institutional behavior. The meaningfulness of the predictive value of institutional adaptation is questionable. (Ross & McKay, 1976, p. 171)
>
> The results showed difference [sic] during treatment favoring the Teaching-Family programs on rate of alleged criminal offenses. . . . In the post-treatment year, none of the differences between the groups was significant on any of the outcome measures. (Kirigin, Braukmann, Atwater, & Wolf, 1982, p. 1)

As noted in the presentation of findings above, we regard the behavioral reinforcement/incentive programs to be in a borderline area of verification. The three independent comparisons from one study (Davidson, Redner, & Amdur, 1990) all used experimental methods we rated as excellent. All three treatments consisted of setting up behavioral contracts for juvenile offenders and engaging in advocacy on behalf of those youth. In two of the conditions, undergraduates were trained in the techniques and implemented them. These two treatment conditions showed the highest effect sizes ($r = .23$ and $r = .31$). The third treatment condition involved teaching the juvenile's family to do behavioral contracting with, and advocacy for, the juvenile ($r = .06$). The only other contingency contracting study using excellent research methods used as its treatment setting up behavioral contracts between the youth and teachers and between the youth and parents ($r = .09$). In our opinion, most volunteers from outside the client's family, peer group, and school (or job) environments will probably greatly reduce or cease their involvement with the

offender when the 6-week or 6-month program interval ends, so the effective contingencies of reinforcement will end at that time. Our opinion is that behavioral reinforcement programs will be verified effective if and only if they can develop and maintain strong contingencies of reinforcement in the natural environment of the clients, for example, maintained by parents (or spouses) and teachers (or employers).[16]

A different focus—one that cognitive-behavioral programs tend to adopt—is to use behavioral learning techniques to change the general adaptive behaviors of the clients, that is, to have the clients return to their natural environment with new repertoires of skills so they can obtain reinforcement in socially acceptable ways instead of illegal ways. This may be part of the reason why the cognitive-behavioral programs discussed above are effective.

Reducing recidivism has been notoriously difficult. It is a relief to know that some correctional programs can indeed work to reduce recidivism by significant amounts. The policy implication is that directors of rehabilitation programs should consider having cognitive-behavioral programming as a primary or secondary component of their treatment programming. Two examples of the kinds of programs that policy makers should review for possible adoption are the cognitive skills training program developed by Ross, Fabiano, Ewles, and colleagues (Kownacki, 1995; Ross, Fabiano, & Diemer-Ewles, 1988) and the aggression replacement training program developed by Goldstein and Glick (1994). It is up to the treatment policy makers and program directors to review whether the (purportedly) cognitive-behavioral program under consideration is evidence based, employing principles and procedures corresponding to those present in the research reviewed here, and to assess how suitable it is for their particular clients and program environment. The broader programming challenge now is to help promote the "technology transfer" so the effective program models (and related staff recruitment, training, and quality-control processes) diffuse throughout the correctional community and become well implemented. The research challenge is to expand and develop the existing body of research evidence, so the effective elements of the behavioral/cognitive-behavioral models can be specified, and then used to improve the program models still further.

NOTES

1. The other three general categories were psychodynamic, life skills, and other.
2. The reported coefficient was phi, which is equivalent to the correlation coefficient.
3. Behaviorists refer to reinforcing stimuli and aversive stimuli, both of which have very specific technical definitions.

4. This category included contingency contracting and token economy programs but excluded aversive conditioning programs.

5. Here cognitive-behavioral programs are broadly defined to encompass a variety of approaches including social learning approaches. Such non-cognitive-behavioral programs as token economies are excluded, however.

6. Lists of the computerized databases searched and the specific journals searched are available on request.

7. 1968 was chosen because the principal investigator's earlier study (Lipton, Martinson, & Wilks, 1975) had covered 1945-1967.

8. This approach was initially developed for substance abuse, but it has been adapted to deal with other behaviors (such as sex offenses). This code was applied to those non-substance-abuse programs as well.

9. The listing of studies is available on request from the lead author.

10. Andrews et al. (1990) and Lipsey (1992) did not use specific criteria like these because the focus of their meta-analyses was not on estimating the effectiveness of specific types of programs (such as behavioral/cognitive-behavioral programs) but rather on relating characteristics of treatment programs (a four-level appropriateness-of-treatment variable, treatment dosage, sociological orientation) to their relative success. Whitehead and Lab (1989) used the much higher *r* (or phi) criterion of +.20. Izzo and Ross (1990) and Gottshalk, Davidson, Mayer, and Gensheimer (1987) just used the .05 level of statistical significance or a 95% confidence interval as the criterion. Garrett (1985) just used a narrative discussion, and did not use any specific quantitative criteria.

11. In evaluation research of correctional treatment programs there are a few "outlier" studies with a total *N* greater than 600. For example, 500 parolees receive an experimental program and the 5,000 persons who experienced ordinary parole in the state form the comparison group. To keep these few (typically lower quality) studies from overwhelming the other studies in the meta-analyses, CDATE's policy was to Windsorize at 600: any study with a total *N* greater than 600 was recoded to a total *N* equal to 600.

12. The actual analyses were conducted on the correlation coefficients transformed using Fisher's *z*. The values in the summary tables designate transformations back from Fisher's *z* to the correlation coefficient values, *r*.

13. Because the aversive conditioning study did not use reinforcement or incentives, it was excluded. Before analyzing the data, this hypothesis also included simple behavioral reinforcement programs, but it turned out none of the studies had that as the most important treatment code.

14. If we had not excluded the outlier (Davidson & Robinson, 1975), the weighted mean *r* would have been +.121, and the exact probability of the significance test would have been .037.

15. Cognitive skills programs are cognitive behavioral, too. These clearly identified programs were given a separate code as cognitive skills, rather than put in with the other cognitive-behavioral programs (which would have made that aggregated group more heterogeneous).

16. Of course, any correctional treatment program should conform to high ethical standards, such as respecting the dignity of the clients who are to receive the treatment.

REFERENCES

Akers, R. L. (1977). *Deviant behavior: A social learning approach*. Belmont, CA: Wadsworth.

Andrews, D. A., & Bonta, J. (1998). *The psychology of criminal conduct*. Cincinnati, OH: Anderson.

Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P., & Cullen, F. T. (1990). Does correctional treatment work? A clinically-relevant and psychologically-informed meta-analysis. *Criminology*, *28*(3), 369-404.

Bandura, A. (1969). *Principles of behavior modification*. New York: Holt, Rinehart & Winston.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Bryk, A. S., Raudenbush, S. W., & Congdon Jr., R. T. (1996). *Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs*. Lincolnwood, IL: Scientific Software International.

D'Zurilla, T. J., & Goldfried, M. R. (1971). Problem-solving and behavior modification. *Journal of Abnormal Psychology*, *78*, 107-126.

Davidson, W. S., & Robinson, M. J. (1975). Community psychology and behavior modification: A community based program for the prevention of delinquency. *Journal of Corrective Psychiatry and Behavior Therapy*, *21*, 1-12.

Davidson, W. S., Redner, R., & Amdur, R. (1990). *Alternative treatments for troubled youth: The case of diversion from the justice system*. New York: Plenum.

Davidson, W. S., & Wolfred, T. R. (1977). Evaluation of a community-based behavior modification program for prevention of delinquency: The failure of success. *Community Mental Health Journal*, *13*, 296-306.

Garrett, C. J. (1985). Effects of residential treatment of adjudicated delinquents. A meta-analysis. *Journal of Research in Crime and Delinquency*, *22*, 287-308.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Goldstein, A. P., & Glick, B. (1994). *The Prosocial Gang: Implementing aggression replacement training*. Thousand Oaks, CA: Sage.

Gottschalk, R., Davidson, W. S., II, Mayer, J. P., & Gensheimer, L. K. (1987). Behavioral approaches with juvenile offenders: A meta-analysis of long-term treatment efficacy. In E. K. Morris & C. J. Braukman (Eds.), *Behavioral approaches to crime and delinquency: A handbook of application, research and concepts* (pp. 399-420). New York: Plenum.

Handler, E. (1974). *The unspectacular results of three local residential treatment programs*. Urbana, IL: Jane Adams School of Social Work.

Hedges, L. V., & Olkin, I. (1983). Regression models in research synthesis. *American Statistician*, *37*, 137-140.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Izzo, R. L., & Ross, R. R. (1990). Meta-analysis of rehabilitation programs for juvenile delinquents. *Criminal Justice and Behavior*, *17*, 134-142.

Kirigin, K. A., Braukmann, C. J., Atwater, J. D., & Wolf, M. M. (1982). An evaluation of Teaching-Family (Achievement Place) group homes for juvenile offenders. *Journal of Applied Behavior Analysis*, *15*, 1-16.

Kownacki, R. J. (1995). The effectiveness of a brief cognitive-behavioral program on the reduction of antisocial behaviour in high-risk adult probationers in a Texas community. In R. R. Ross & R. D. Ross (Eds.), *Thinking straight: The reasoning and rehabilitation program for delinquency prevention and offender rehabilitation* (pp. 249-257). Ottawa, Canada: Air Training and Publications.

Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Corday, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook* (pp. 83-127). New York: Russell Sage.

Lipton, D. S., Martinson, R., & Wilks, J. (1975). *The effectiveness of correctional treatment*. New York: Praeger.

Maltz, M. D. (1984). *Recidivism*. Academic Press.

Marlatt, G. A., & Gordon, J. R. (1985). Relapse prevention: Theoretical rationale and overview of the model. In G. A. Marlatt, & J. R. Gordon (Eds.), *Relapse prevention: Maintenance strategies in the treatment of addictive behaviors* (pp. 3-70). New York: Guilford.

McGuire, J. (1996). *Cognitive behavioral approaches: An introductory course on theory and research*. Liverpool, UK: Department of Clinical Psychology, University of Liverpool.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321). New York: Russell Sage.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231-244). New York: Russell Sage.

Ross, R. R., Fabiano, E. A., & Diemer-Ewles, C. (1988). Reasoning and rehabilitation. *International Journal of Offender Therapy and Comparative Criminology*, *32*, 29-35.

Ross, R. R., & McKay, H. B. (1976). A study of institutional treatment programs. *International Journal of Offender Therapy and Comparative Criminology*, *20*, 165-173.

Schwarzer, R. (1989). *Meta-analysis programs*. Berlin, Germany: Institut für Psychologie, Freie Universität Berlin.

Skinner, B. F. (1953). *Science and human behavior.* New York: Free Press.

Whitehead, J. T., & Lab, S. P. (1989). A meta-analysis of juvenile correctional treatment. *Journal of Research in Crime and Delinquency*, *26*, 276-295.

Yochelson, S., & Samenow, S. E. (1977). *The criminal personality: The change process*. New York: Aronson.