



## Drug Policy Research Center

A JOINT ENDEAVOR WITHIN RAND HEALTH AND  
RAND INFRASTRUCTURE, SAFETY, AND ENVIRONMENT

THE ARTS  
CHILD POLICY  
CIVIL JUSTICE  
EDUCATION  
ENERGY AND ENVIRONMENT  
HEALTH AND HEALTH CARE  
INTERNATIONAL AFFAIRS  
NATIONAL SECURITY  
POPULATION AND AGING  
PUBLIC SAFETY  
SCIENCE AND TECHNOLOGY  
SUBSTANCE ABUSE  
TERRORISM AND  
HOMELAND SECURITY  
TRANSPORTATION AND  
INFRASTRUCTURE  
WORKFORCE AND WORKPLACE

This PDF document was made available from [www.rand.org](http://www.rand.org) as a public service of the RAND Corporation.

[Jump down to document](#) ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

### Support RAND

[Purchase this document](#)

[Browse Books & Publications](#)

[Make a charitable contribution](#)

### For More Information

Visit RAND at [www.rand.org](http://www.rand.org)

Explore [RAND Drug Policy Research Center](#)

View [document details](#)

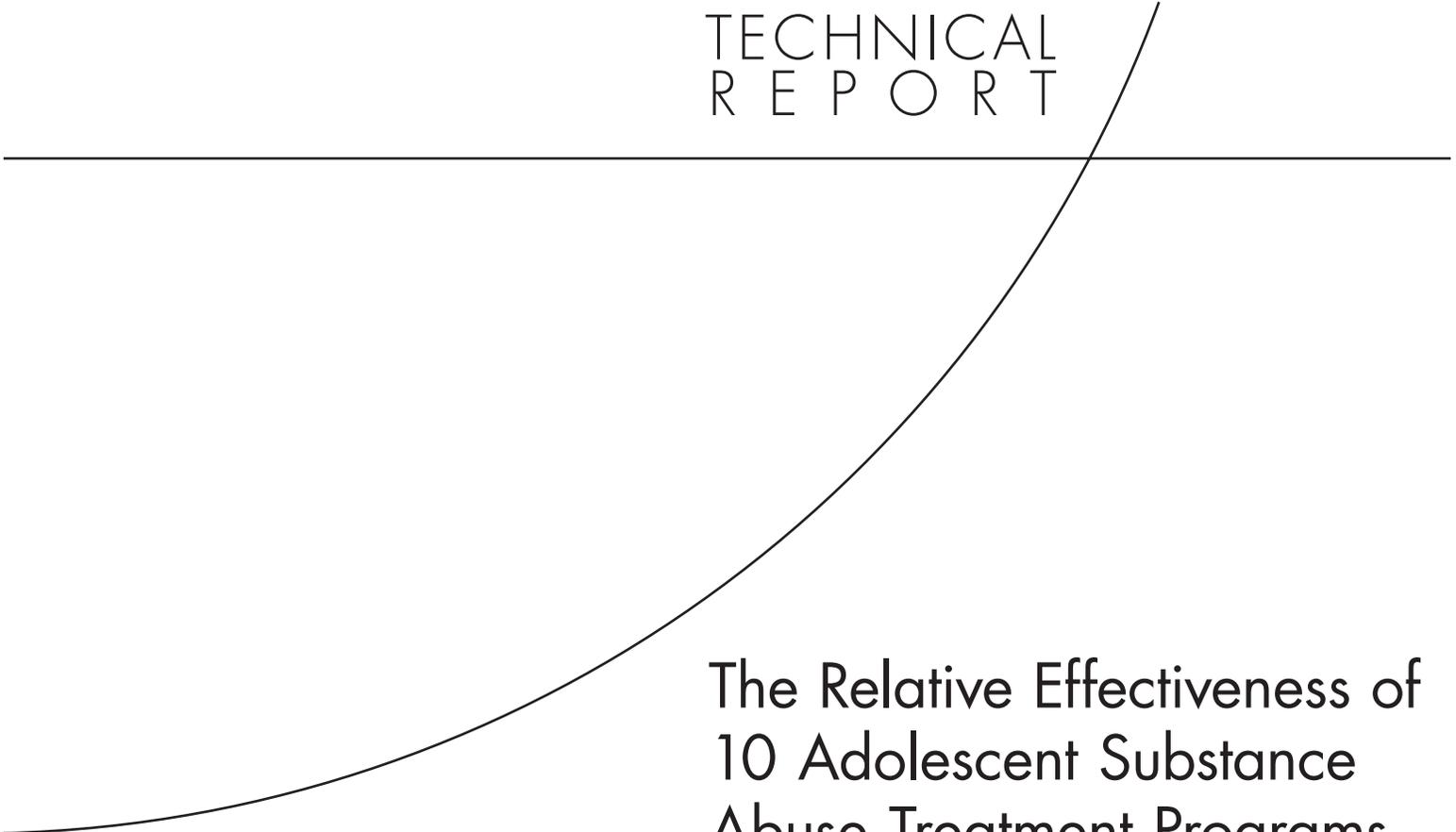
### Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use.

This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL  
R E P O R T

---



# The Relative Effectiveness of 10 Adolescent Substance Abuse Treatment Programs in the United States

Andrew R. Morral, Daniel F. McCaffrey,  
Greg Ridgeway, Arnab Mukherji,  
Christopher Beighley

Prepared for the Substance Abuse and Mental Health Services Administration,  
Center for Substance Abuse Treatment



Drug Policy Research Center

A JOINT ENDEAVOR WITHIN RAND HEALTH  
AND RAND INFRASTRUCTURE, SAFETY, AND ENVIRONMENT

The research described in this report was carried out under the auspices of the Drug Policy Research Center, a joint endeavor of RAND Infrastructure, Safety, and Environment (ISE) and RAND Health.

**Library of Congress Cataloging-in-Publication Data**

The relative effectiveness of 10 adolescent substance abuse treatment programs in the U.S. /

Andrew R. Morral ... [et al.].

p. cm.—(TR ; 346)

Includes bibliographical references.

ISBN 0-8330-3916-4 (pbk. : alk. paper)

1. Youth—Substance use—United States. 2. Substance abuse—Treatment—United States. 3. Substance abuse—United States—Prevention. I. Morral, Andrew R. II. Title: Relative effectiveness of eleven adolescent substance abuse treatment programs in the U. S. III. Series: Technical report (Rand Corporation) ; 346.

HV4999.Y68R45 2006

616.86'0608350973—dc22

2006007643

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

**RAND**® is a registered trademark.

© Copyright 2006 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2006 by the RAND Corporation  
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138  
1200 South Hayes Street, Arlington, VA 22202-5050  
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213  
RAND URL: <http://www.rand.org/>  
To order RAND documents or to obtain additional information, contact  
Distribution Services: Telephone: (310) 451-7002;  
Fax: (310) 451-6915; Email: [order@rand.org](mailto:order@rand.org)

## Summary

---

Each year, substance abuse treatment programs in the United States record approximately 150,000 admissions of youths under the age of 18. Nevertheless, little is known about the effectiveness of the types of community-based services typically available to youths and their families. Recognizing the need for better information on the effectiveness of adolescent treatment services, the Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment (SAMHSA/CSAT) established the Adolescent Treatment Models (ATM) program. ATM grants were available to established adolescent treatment programs with at least suggestive evidence of effectiveness. These grants supported an independent, longitudinal evaluation of the client outcomes. Eleven such studies were funded in 1998 and 1999, each of which used parallel data collection instruments and study designs (see Stevens & Morral, 2003, for discussion of the 11 programs and each site's study design).

One of CSAT's program goals for the ATM projects was to establish whether any of the selected programs produce especially good client outcomes. However, to draw conclusions about the relative effectiveness of the ATM programs, it is probably not sufficient to simply compare the outcomes of youths at each program. Instead, some allowance must be made for the possibility that the youths treated at each program may not be identical. Specifically, programs may treat youths that differ with respect to how hard they are to treat or how great a risk of poor outcomes they present, in which case a simple comparison of outcomes or change in problems over time would not properly establish the relative effectiveness of the programs.

Because there are different approaches to case-mix adjustment that might yield different conclusions on the effectiveness of the ATM programs, CSAT decided to have the longitudinal data collected at the ATM programs subjected to several case-mix adjusted analyses of the relative effectiveness of the studied programs, each using a different approach, but examining a similar set of outcomes.

RAND's approach to examining the relative effectiveness of the 11 programs is designed to answer the question, "Will youths representative of those entering Program X have better outcomes if they receive treatment at Program X, rather than at one of the other studied programs?" This approach considers the possibility that programs may be most effective for the types of youths they typically serve. That is, programs may have adapted their interventions to optimally address the service needs of their treatment populations.

Data for the analyses presented in this report come from the 11 ATM studies. In each, all adolescent admissions were invited to participate in a longitudinal survey consisting of a

baseline interview, and periodic follow-up interviews, including a 12-month post-admission assessment. All cases successfully interviewed at 12 months, as of March 2002, are included in the present analyses. This represents over 90 percent of the original baseline sample.

The substance abuse treatment programs for adolescents evaluated under ATM include three long-term residential (LTR) facilities (Dynamic Youth Community, Inc., in New York; Phoenix Academy of Los Angeles; and Thunder Road in Oakland), four short-term residential (STR) facilities (La Cañada in Tucson; Mountain Manor in Baltimore; Our Youth Our Future in Shiprock, New Mexico; and Thunder Road in Oakland), and four outpatient (OP) programs (Chestnut Health Systems in Bloomington, Illinois; Epoch Counseling Center in Baltimore County; Teen Substance Abuse Treatment in Maricopa County, Arizona; and The Village in Miami). Because of the small number of cases available for analysis at one ATM outpatient program, treatment effect estimates were not calculated for it.

In addition to the 263 LTR, 594 STR, and 404 OP cases followed under ATM at these 11 programs, an additional 128 LTR cases and 156 OP cases were also included in the ATM study and were available for inclusion in the present study as comparison cases. These cases came from a range of different community-based programs, which are not evaluated in this report.

Because this report is designed to examine whether there is any evidence of relative treatment effects, rather than to suggest that one or another treatment program is superior to others, we have obscured the identity of programs in the results section by referring to them as Programs A through J, designations assigned randomly within level of care.

### **Case-Mix Adjustment Approach**

Our estimates of the causal effect of going to a particular facility as opposed to any other facility within the same level of care uses a case-mix adjustment approach in which weights are used to upweight eligible comparison cases most comparable to the treatment group cases, and downweight those least comparable. The weights were chosen to maximize the similarity of the distributions for the target and comparison group cases on 86 control variables collected at treatment admission. Specifically we use propensity score weights estimated using the 86 control variables and a generalized boosting model (GBM). McCaffrey, Ridgeway, and Morral (2004) provide additional details on GBM propensity score weighting to estimate causal effects.

The pretreatment control variables used to perform the case-mix adjustment were drawn from the data collection instrument used in each ATM evaluation, the Global Appraisal of Individual Needs (GAIN; Dennis, 1999). The 86 GAIN items selected as control variables were identified by the GAIN's developer as operationalizing the American Society of Addiction Medicine (ASAM) Patient Placement Criteria (Mee-Lee, Shulman, Fishman, Gastfriend, and Griffith, 2001). ASAM patient placement criteria represent the best available empirical evidence and clinical judgment about the factors most important for determining patient placement and treatment needs. These variables concern substance use problems and severity; types,

frequencies and histories of substances used; substance abuse, mental health, and physical health treatment histories and need; and many other potential risk factors (see Tables A.1 through A.4 in the appendix for a complete list of control variables).

### Case-Mix Adjustment Results

Before adjusting comparison groups using propensity score weights, every treatment program was found to have cases that differed significantly from its respective comparison cases on many pretreatment risk factors. This shows that unadjusted comparisons of program outcomes, even within level of care, are likely to conflate true treatment effects with outcome differences attributable to outcome differences expected for cohorts exposed to different risks.

Propensity score weights greatly reduced the differences between control variable distributions. Indeed, satisfactory equivalence between treatment and comparison groups on the 86 pretreatment control variables was achieved for 4 of the 10 programs (B, G, H, and I), meaning that between-group differences on these pretreatment variables were no larger than would be expected had the same cases been randomly assigned to groups of equivalent size. Weighting was not sufficient to balance all pretreatment variables for Programs A, C, D, E, F, and J, meaning that at least one of the 86 variables showed group differences after weighting that were larger than would be expected had groups been formed through random assignment. For these programs, we took additional steps to ensure that outcome comparisons adjust for the remaining pretreatment group differences. Specifically, we included all such unbalanced variables as covariates in a weighted analysis of covariance (ANCOVA) model of treatment effects associated with these programs.

Using the derived propensity score weights, we estimated each program's relative treatment effects on six outcome measures assessed 12 months after treatment entry: (1) recovery, a dichotomous measure indicating the youth is free in the community and not using drugs, (2) substance problems, (3) substance use frequency, (4) illegal activities, (5) emotional problems, and (6) days in a controlled environment (like detention or a residential treatment program).

Results of the treatment effect estimates, by program, showed that across the 60 program and outcome pairs, only 11 significant relative treatment effects were observed. Moreover, many of the estimated effects were small. For instance, more than half of the estimated effects for substance abuse problems and substance abuse frequency were between  $-0.1$  and  $0.1$ .

Only five of the significant differences indicated treatment effects superior to those found in comparison sites. This scarcity of detectable positive relative treatment effects suggests that after controlling for pretreatment differences, the 12-month outcomes among the ATM programs differed by an amount too small to detect given the statistical power of our analyses.

The presence of six significant and negative effect sizes is also surprising but requires some caveats. First, half of these effects are for a single outcome variable, emotional problems. The relationship of this outcome to other traditional substance abuse treatment outcomes is not entirely clear. It could be argued, for instance, that some emotional distress might be an expected correlate of abstinence from the substances previously used to medicate those problems. Similarly, a fourth apparently negative effect size is the finding that Program C is asso-

ciated with increased days in a controlled environment. However, in the context of an LTR program designed to treat youths for a year or more, this apparently inferior outcome might actually be due to the program's differential success retaining youth in the controlled environment of treatment, a seemingly beneficial effect. More difficult to explain is the negative effect size of Program E on recovery and of Program H on illegal activities, the latter of which also benefited from a particularly good case-mix adjusted comparison group.

## Implications

When considering these 86 important pretreatment measures, the various programs appear to be serving distinctly different populations. Before weighting there were large and significant differences between every target program and its corresponding comparison programs. These differences were often for clearly important variables such as pretreatment substance use and the use of needles and opiates. Weighting by the propensity scores removes many of these differences but after weighting, several target programs continued to differ from their comparison group on potentially important risk factors.

The large differences between program clients resulted in small effective sample sizes for the weighted comparison cases. Combined with the small sample sizes in several programs the result was limited power to detect what are typically considered small effects. The use of ANCOVA in the current context potentially reduced power even more. Moreover, while treatment on the treated effect estimated in this report produce more substantively interesting results, it also comes at the cost of less precision for the estimated comparison group mean and potentially less power. Hence, the price of aggressive case-mix adjustment to avoid biased estimates may be a reduction in power for detecting true differences. Interestingly, in many cases described in this report, weighting and the consequent reduction in effective sample size did not appear to be the limiting factor for power. Instead, the reductions in effective sample size caused by weighting were compensated for by improvements in estimate precision. Future analyses should seek to evaluate treatment effects using larger samples, or by improving power in other ways.

Not surprisingly given the limitations of our analyses, the evidence favoring a view that one or more of the programs is especially effective with its treatment population is not particularly strong. However, what is somewhat surprising is that the most consistently small effects are for direct measures of substance use, the substance use problems and frequency indices. Given that the case-mix adjustment perspective we adopted for these analyses was designed to detect treatment effects resulting from adaptations programs have made to the specific characteristics of the clients they typically serve, or to detect more general types of superiority, we expected to find greater differences in the substance use measure. The findings of this report are similar to those reported from the Cannabis Youth Treatment project (Dennis et al., 2004) and Project MATCH (Matching Alcoholism Treatment to Client Heterogeneity) (Project Match Research Group, 1997). Specifically, differential treatment effects are not found, even though programs are compared, in some cases, to interventions involving fairly minimal treatment.

The modest evidence of differential effectiveness found in this study also may result from examining relative, rather than absolute treatment effects. Relative treatment effects yield

ambiguous results. For instance, our finding that almost all estimated effects for the substance-use measures were small and that no STR or OP program appears to be more effective than its comparison programs at reducing substance use or substance problems admits a number of different interpretations.

- If all programs had poor absolute treatment effects on these outcomes, we might expect no relative treatment effects.
- If all programs had very large, but equivalent, absolute treatment effects on these outcomes, we would similarly expect no relative treatment effects.
- If the comparison sample is drawn from some programs that are more effective and some that are less effective than the program being evaluated, we might expect to find no relative treatment effects.

Moreover, even if relative differences do exist, we might expect them to be smaller than the effects we would find when comparing treatment to a no-treatment alternative and our analyses might have less power to detect relative differences than they would to detect difference against a control condition that received no treatment.

The possibility that all programs were effective, thereby reducing the chances of noting differential or relative treatment effects is worth considering. The 11 programs evaluated under ATM were not a random sample of community-based substance abuse treatment programs for adolescents. In fact, they were selected by CSAT for evaluation because they were able to demonstrate some evidence of effectiveness. Thus, the outcomes these programs produced may well have been unusually good across all programs, thereby reducing the appearance of differential treatment effects.

Finally, our failure to find strong and persuasive evidence of treatment effectiveness could indicate that we were looking for that evidence in the wrong place. Large and significant treatment effects might exist for each evaluated treatment program, but these might be no longer detectable 12 months after treatment admission. McLellan et al. (2005) have argued, for instance, that for chronic, relapsing conditions like substance abuse, the treatment effects of interest are chiefly those observed during treatment. By analogy, they note that treatments for other chronic conditions, like diabetes or schizophrenia, are typically evaluated in terms of symptom management during treatment, not by whether or not such benefits remain detectable long after treatments are withdrawn.

Whereas this argument has considerable appeal for treatment of adult chronic substance abusers, the applicability to adolescent substance abusers, many of whom appear to be in the early stages of substance abuse, is not yet well established. Clearly parents and other stakeholders would like to identify treatments that have long-lasting effects, making the evaluation of 12-month outcomes for adolescent treatments a valid, if nevertheless unpromising, enterprise.

In the absence of widely accepted and suitably specific quality-of-care indicators for adolescent substance abuse treatment, SAMHSA and many states are now exploring the use of outcomes-based performance measurement systems. The results from this study highlight a

number of challenges to this plan and suggest strategies that should be considered in these and other efforts to understand treatment effects among individual programs, types of programs, or geographic regions.

### **Case-Mix Adjustment Is Important**

The risk profiles of clients at different programs are quite variable. As a result, especially good or bad client outcomes may not reflect the performance of the treatment program, so much as the risk profiles of the clients it serves. Until the outcomes of comparable youths are compared, the relative effectiveness of different programs or groups of programs cannot be established. Similarly, if a program's client population changes over time, then comparisons of outcomes over time may not correspond to changes in that program's effectiveness. Finally, if risk profiles of clients differ by state, then comparisons of state treatment outcomes that fail to properly adjust for these differences are likely to be misleading.

### **Case-Mix Adjustment Is Not Enough**

Even if the risk profiles of youths entering different programs were identical, this report highlights several ways in which valid conclusions about the effectiveness of different treatment programs may be difficult to draw from outcomes data. For instance, problems for interpretation arise when the proportion of cases providing follow-up data differs across programs. Thus, among the ATM programs, one had substantially lower rates of follow-up than the others, but the outcomes that were observed for this program appeared to be especially good. If those who are not followed up differed systematically from those who are, the outcomes for this program may be substantially more biased than those for the other ATM programs. Moreover, these are biases that may well be uncorrectable through, for instance, attrition weights, since the factors leading to study attrition may concern influences that occur long after the baseline assessment, such as nonresponsiveness to treatment. This suggests the need to ensure uniformly high rates of outcomes collection across programs.

Another challenge to the interpretation of outcomes concerns the high rates of institutionalization at follow-up observed in the ATM sample. When large proportions of cases are in controlled environments, many of the most important substance abuse treatment outcome measures are difficult to interpret. Thus, for instance, low levels of crime, drug use, and drug problems would ordinarily be viewed as a good outcome, but not if it reflects high rates of incarceration for a program's clients. Moreover, if programs from different geographic regions are compared, then differences in institutionalization rates cannot necessarily be attributed to treatment program effects. Instead, they might reflect differences in regional laws or law enforcement, the availability of inpatient services, or other community differences.

Indeed, community differences in resources, available support, opportunities, drug problems, and other characteristics may all influence the outcomes of youths in those communities, creating a significant confound for trying to establish effects specific to treatment programs on the basis of client outcomes, when those programs exist in different communities.

Each of these considerations suggest that even if case-mix adjustment is planned for outcomes-based performance measurement, valid conclusions about program performance may not be derivable from large-scale outcomes-monitoring systems.

**Program Differences May Be Small**

In this report, no program was found to consistently outperform any other, across outcomes. Indeed, few program differences were noted on any outcome, even though 70 percent of these analyses had sufficient power to detect effect sizes of 0.40 or greater. If differences in program effectiveness are generally small, this means that in order to detect performance differences, the outcomes of larger numbers of cases must be assessed before sufficient data is available to draw desired inferences. Since some programs see only small numbers of cases, and since the collection of outcomes data is often expensive, this finding has important implications for the feasibility of using outcomes to assess differences in the performance of individual programs.

**Outcomes-Based Performance Measurement May Be Impractical**

Many of the challenges described above raise important questions about the feasibility of producing valid treatment performance information from large-scale outcomes-based performance measurement efforts. This suggests that a more fruitful approach to performance measurement might be to invest more effort into identifying quality of care indicators for adolescent substance abuse treatment programs.