ARTICLE

Using Pay for Performance to Improve Treatment Implementation for Adolescent Substance Use Disorders

Results From a Cluster Randomized Trial

Bryan R. Garner, PhD; Susan H. Godley, RhD; Michael L. Dennis, PhD; Brooke D. Hunter, MS; Christin M. L. Bair, BS; Mark D. Godley, PhD

Objective: To test whether pay for performance (P4P) is an effective method to improve adolescent substance use disorder treatment implementation and efficacy.

Design: Cluster randomized trial.

Setting: Community-based treatment organizations.

Participants: Twenty-nine community-based treatment organizations, 105 therapists, and 986 adolescent patients (953 with complete data).

Intervention: Community-based treatment organizations were assigned to 1 of the following conditions: the implementation-as-usual (IAU) control condition or the P4P experimental condition. In addition to delivering the same evidence-based treatment (ie, using the Adolescent Community Reinforcement Approach [A-CRA]), each organization received standardized levels of funding, training, and coaching from the treatment developers. Therapists in the P4P condition received US $50 for each month that they demonstrated competence in treatment delivery (ie, A-CRA competence) and US $200 for each patient who received a specified number of treatment procedures and sessions (ie, target A-CRA) that has been found to be associated with significantly improved patient outcomes.

Main Outcome Measures: Outcomes included A-CRA competence (ie, a therapist-level implementation measure), target A-CRA (ie, a patient-level implementation measure), and remission status (ie, a patient-level treatment effectiveness measure).

Results: Relative to therapists in the IAU control condition, therapists in the P4P condition were significantly more likely to demonstrate A-CRA competence (24.0% vs 8.9%; event rate ratio, 2.24; 95% CI, 1.12-4.48; $P = .02$). Relative to patients in the IAU control condition, patients in the P4P condition were significantly more likely to receive target A-CRA (17.3% vs 2.5%; odds ratio, 5.19; 95% CI, 1.53-17.62; $P = .01$). However, no significant differences were found between conditions with regard to patients' end-of-treatment remission status.

Conclusion: Pay for performance can be an effective method of improving treatment implementation.

Trial Registration: clinicaltrials.gov Identifier: NCT01016704

Arch Pediatr Adolesc Med. 2012;166(10):938-944.
Published online August 13, 2012.
doi:10.1001/archpediatrics.2012.802

Author Affiliations: Lighthouse Institute, Chestnut Health Systems, Normal, Illinois.

In 2001, THE INSTITUTE OF MEDICINE published *Crossing the Quality Chasm: A New Health System for the 21st Century*, which called for the need to "align financial incentives with the implementation of care processes based on best practices and the achievement of better patient outcomes."[1(p184)] In the decade since this landmark report was published, pay for performance (P4P [ie, providing financial incentives for the achievement of predefined criteria]) has been a topic of considerable interest[2-23] and is a strategy specifically recommended by the Institute of Medicine[24] to help improve the delivery of high-quality care.

For editorial comment see page 964

The number of P4P programs in the United States has grown rapidly, with evidence from a study[20] suggesting that more than 150 such programs exist. However, this rapid diffusion of P4P programs has occurred largely in the absence of randomized controlled studies, despite repeated calls for experimental research to

evaluate P4P approaches.[2,3,8,9] It is ironic that the use of P4P has proliferated without experimental support at the same time when evidence-based treatments are not being diffused to practice settings.[25-28]

The present study presents the main effectiveness findings from the Reinforcing Therapist Performance experiment,[29] which is a cluster randomized trial designed to evaluate the efficacy of using P4P methods to improve treatment implementation and effectiveness. This design was used because the primary interest was to examine P4P as an organizational-level intervention and because validity threats are possible from the randomization of patients within therapists (eg, contamination) or of therapists within treatment organizations (eg, compensatory rivalry and resentful demoralization). In addition to adding to the limited knowledge about the effectiveness of P4P methods in general, our findings are significant given that the study was conducted within the context of a national initiative to improve treatment for adolescent substance use disorders, a problem identified as "America's #1 Public Health Problem" according to a 2011 publication by researchers at Columbia University (New York, New York).[30]

METHODS

STUDY POPULATION

Between October 1, 2006, and October 1, 2007, a total of 34 community-based treatment organizations across the United States received discretionary grant funding from the Substance Abuse and Mental Health Services Administration's Center for Substance Abuse Treatment to implement an evidence-based behavioral treatment called the Adolescent Community Reinforcement Approach (A-CRA).[31-33] Although the A-CRA consists of 19 different treatment procedures (designed to help increase adolescents' access to reinforcers through operant conditioning principles and skills training activities so that non–substance using behaviors are rewarded and can replace substance use behavior), more than 1 procedure may be provided in any single session, and any procedure can occur successively throughout treatment. A detailed description of this implementation initiative has been published.[34] Briefly, consistent with the implementation science research literature,[26] the approach was a complex multilevel process involving multiple "core implementation components." For example, therapists at each treatment organization received standardized A-CRA training that included reading the treatment manual, passing a knowledge test, and attending a 3½-day training workshop. To support quality implementation, therapists also received quantitative and qualitative feedback from trained raters and participated in biweekly calls with the developers of the A-CRA model. Each treatment organization also received approximately US $300 000 for each of 3 years to support the implementation. Although a convenience sample, this initiative provided an ideal setting to experimentally test the extent to which P4P methods can be used to improve treatment implementation given that each organization was delivering the same evidence-based treatment and was receiving the same training model and level of funding.

With institutional review board approval, organizations implementing A-CRA treatment in an outpatient setting were eligible and were invited to participate in this study. The criterion for the inclusion of therapists was employment at a participating organization as an A-CRA treatment therapist. Each participating organization signed a memorandum of understanding, and therapists were approached individually and were invited through an informed consent process to participate in the study. The recruitment of organizations was completed between November 17, 2008, and January 12, 2009.

INTERVENTION

In addition to the implementation-as-usual (IAU) procedures delivered by organizations and therapists in both treatment conditions, participating therapists working at organizations assigned to the P4P condition had the opportunity to earn monetary bonuses for the achievement of 2 predefined treatment implementation performance measures. Specifically, building on prior research that identified specified levels of A-CRA treatment associated with significantly better follow-up outcomes,[29,35] therapists could earn US $200 for each of their patients who received at least 10 of 12 specific A-CRA procedures delivered within the first 14 weeks of treatment and in no fewer than 7 sessions (target A-CRA). To reinforce the quality of treatment delivery, therapists also could earn US $50 for each month that they demonstrated competent delivery of all components of at least 1 A-CRA treatment procedure during the same treatment session (A-CRA competence). Notably, the achievement of both implementation measures was objectively determined based on expert review of session recordings using a detailed rating manual.[36] To demonstrate the delivery of target A-CRA, therapists were required to provide recorded evidence that they had delivered at least 10 of 12 specified A-CRA procedures and had delivered at least 7 treatment sessions. Similarly, to ensure a representative sample of treatment session recordings from which to randomly select, demonstration of A-CRA competence required therapists to submit a session recording from at least 80% of their treatment sessions conducted during the month.

Determining incentive sizes was a difficult aspect of designing this trial. Guided by prior related research,[11,37] we chose incentive amounts that we estimated would enable therapists in the P4P condition to, on average, earn incentive amounts that during a 12-month period would add up to approximately 4% to 7% of their mean annual base salary of US $35 000. We believed that such amounts were large enough to significantly improve therapist performance yet were small enough to be considered within a practical range for community-based treatment providers to implement.

During the second to third weeks of each month, all participants in the P4P condition received e-mail notifications documenting their achievement of target A-CRA and A-CRA competence during the prior calendar month. Payments were sent to participants the following week by direct deposit to the therapist's designated account or by a check made payable and mailed to the therapist.

PRIMARY OUTCOME MEASURES

Treatment Implementation

The treatment implementation measures of the study were therapist-level A-CRA competence and patient-level target A-CRA. The achievement (dichotomously coded as yes or no) of each of these outcome measures was determined by one of us (C.M.L.B.) via review of digital audio recordings of treatment sessions. To monitor coding accuracy, a trained rater who was blinded to study conditions independently rated randomly selected examples of target A-CRA and A-CRA competence each month. Across 21 ratings of A-CRA competence, the agreement between raters was 95%. Across 18 ratings of target A-CRA ratings, the agreement between raters was 100%.

Treatment Effectiveness

Although target A-CRA and A-CRA competence were the 2 treatment implementation measures addressed for change, we also evaluated the extent to which the P4P intervention influenced treatment effectiveness using patient-level remission status, which was a primary outcome measure in the Cannabis Youth Treatment study.[31] Patients were considered in remission when they reported no past-month substance use, abuse, or dependence problem, while living in the community (vs incarceration, inpatient treatment, or other controlled environment). Remission status was collected using the Global Appraisal of Individual Needs (GAIN).[38] Intake and 6-month follow-up GAIN assessments were completed by trained GAIN interviewers from each treatment organization.

RANDOMIZATION

After the recruitment of treatment organizations and the initial group of therapists from each participating organization, condition assignment for each organization (ie, cluster) was determined using an urn randomization program (gRand; Yale University).[39] Specifically, the program used organizational-level information (dichotomized according to median split) to balance conditions. Data used for the randomization included the following for each organization: mean therapist age, number of therapists, percentage of female therapists, percentage of therapists of white race/ethnicity, mean session recording rate, mean therapist-reported target A-CRA rate, percentage of female patients, percentage of patients of white race/ethnicity, percentage of patients of Hispanic race/ethnicity, mean patient-level remission status at the follow-up assessment, and A-CRA training staff ratings of the organization's expected study performance. If staff turnover occurred, replacement staff were approached about study participation. After the organizations had been randomly assigned to a condition, 12 of 14 (85.7%) IAU therapists and 11 of 15 (73.3%) P4P therapists were recruited and agreed to participate in the study.

BLINDING

It was impossible to blind organizations, therapists, or all research staff to condition assignment. This was necessary because of the nature of the intervention.

HYPOTHESES

The objectives of this study were to evaluate the efficacy of using P4P methods to improve treatment implementation (ie, therapist-level A-CRA competence and patient-level target A-CRA) and treatment effectiveness (ie, patient-level remission status). We hypothesized that relative to the IAU condition (1) therapists in the P4P condition would have a significantly higher likelihood of demonstrating A-CRA competence, (2) patients in the P4P condition would have a significantly higher likelihood of receiving target A-CRA, and (3) patients in the P4P condition would have a significantly higher likelihood of attaining remission status. Based on our initial power calculations, which had assumed an 80% patient follow-up rate, each of these hypotheses had 80% or higher power for a 2-tailed test with $P < .05$ to detect medium effect sizes (effect size guidelines are given in the "Statistical Analysis" subsection).

STATISTICAL ANALYSIS

The planned primary analyses for the study were adjusted results that took into account the multilevel nature of the data (ie, patients clustered within therapists and therapists clustered within treatment organizations) and included propensity score adjustment measures. The inclusion of propensity score adjustment measures is recommended as an efficient method of adjusting for biases that may be introduced due to using a cluster randomized design.[40]

Three adjusted intent-to-treat multilevel models were conducted using commercially available software (HLM version 6; Scientific Software International Inc).[41] The first adjusted model regressed therapist-level A-CRA competence (using Poisson distribution) on therapist propensity score adjustment and organization-level condition assignment. The second adjusted model regressed patient-level target A-CRA (using Bernoulli distribution) on patient-level propensity score, therapist-level propensity score, and organization-level condition assignment. The third adjusted model regressed patient-level remission status on patient-level propensity score, therapist-level propensity score, and organization-level condition assignment. In addition to reporting statistical significance (ie, 2-sided $P < .05$), we provide effect sizes (odds ratio [OR] or event rate ratio [ERR] with 95% CI) for all results. Consistent with data by Bedard et al,[42] effect sizes were defined as follows: small effect (OR of 1.3 or ERR of 0.8), medium effect (OR of 1.5 or ERR of 0.7), and large effect (OR of 2.0 or ERR of 0.5).

RESULTS

MAIN STUDY FINDINGS

The **Figure** shows the flow of organizations, therapists, and patients through each stage of the study. **Table 1** gives the results of the logistic regression analyses used to create the therapist propensity score adjustment and patient propensity score adjustment measures. The table summarizes characteristics of the therapists at study recruitment and of the patients at treatment intake. No adverse events were reported.

After controlling for therapist propensity to be assigned to the P4P condition, adjusted analysis results (**Table 2**) revealed that therapists assigned to the P4P condition had a significantly higher likelihood of demonstrating A-CRA competence relative to therapists assigned to the IAU condition (24.0% for P4P vs 8.9% for IAU; ERR, 2.24; 95% CI, 1.12-4.48; $P = .02$). After controlling for therapist and patient propensity to be assigned to the P4P condition, patients in the P4P condition had a significantly higher likelihood of receiving target A-CRA relative to patients assigned to the IAU condition (17.3% for P4P vs 2.5% for IAU; OR, 5.19; 95% CI, 1.53-17.62; $P = .01$). Finally, after controlling for therapist and patient propensity to be assigned to the P4P condition, no statistically significant difference in patient-level remission status was observed between the 2 conditions (41.8% for P4P vs 50.8% for IAU; OR, 0.68; 95% CI, 0.35-1.33; $P = .25$).

POST HOC ANALYSES

Given that the large effects of the P4P intervention on treatment implementation (ie, therapist-level A-CRA competence and patient-level target A-CRA) did not translate into a statistically significant difference in patient-level treatment effectiveness (ie, remission status), we

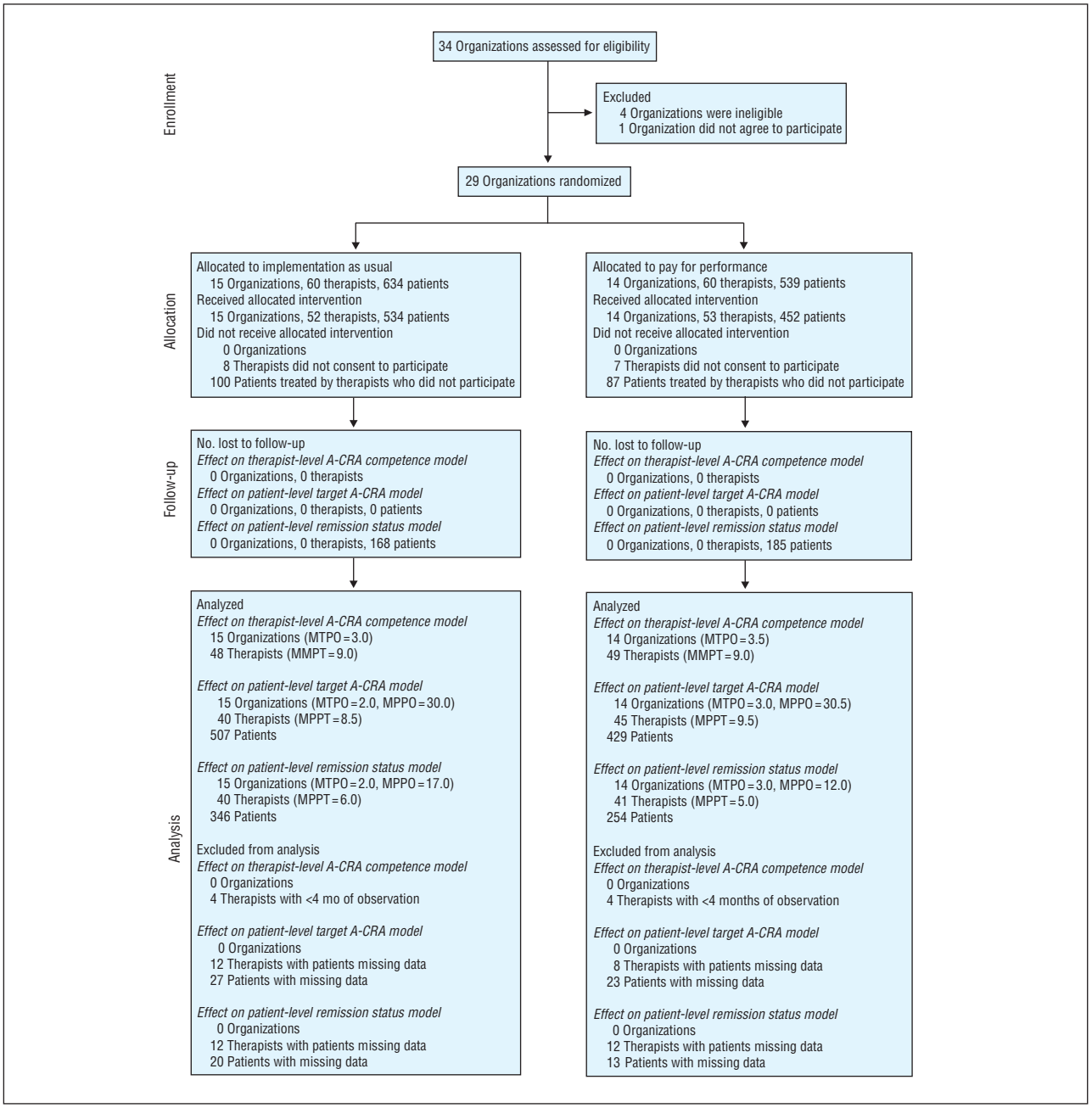

Figure. Flow of treatment organizations, therapists, and patients through the study. A-CRA indicates Adolescent Community Reinforcement Approach; MMPT, median months per therapist; MPPO, median patients per organization; MPPT, median patients per therapist; and MTPO, median therapists per organization.

conducted post hoc analyses to examine the extent to which A-CRA competence and target A-CRA were associated with remission status. Multilevel bivariate analyses indicated that (1) therapist-level A-CRA competence was not significantly associated with patient-level remission status (P = .82) and (2) patient-level target A-CRA was significantly associated with patient-level remission status (OR, 1.91; 95% CI, 1.02-3.58; P = .04). Given the significant positive association between target A-CRA and remission status, we then examined the extent to which the relationship between target A-CRA and remission status may have been moderated by condition assignment. However, moderator analyses did not reveal a significant interaction between condition assignment and target A-CRA with respect to patient remission status (P = .37). Finally, although the follow-up rates were similarly low for both conditions (64.8% for IAU and 56.2% for P4P), we examined the extent to which patients who were included as part of the treatment effectiveness analyses were significantly different from patients who were lost to follow-up analysis. For the IAU condition, no significant differences in baseline characteristics were observed between patients included in the treatment effectiveness analysis and those lost to follow-up analysis. For the P4P condition, patients included in the treatment effectiveness analysis reported significantly more severe substance-related problems at study intake than those lost to follow-up analysis (P = .03).

Table 1. Baseline Characteristics and Propensity Scores

Characteristic	Allocation		Results of Propensity Analysis		
	Implementation as Usual	Pay for Performance	Coefficient (SE)	Odds Ratio (95% CI)	*P* Value
Therapists (n = 105)					
Female sex, %	67	81	0.91 (0.54)	2.49 (0.87-7.13)	.09
White race/ethnicity, %	58	51	–0.13 (0.49)	0.88 (0.34-2.32)	.80
Age, mean (SD), y	37.2 (11.7)	36.0 (10.7)	0.00 (0.02)	1.00 (0.95-1.05)	.96
Master's degree or higher, %	60	49	–0.44 (0.48)	0.65 (0.25-1.65)	.36
Experience, mean (SD), mo	60.3 (80.2)	35.9 (37.5)[a]	–0.01 (0.01)	0.99 (0.98-1.00)	.07
Remission patients, %	6	6	0.58 (1.00)	1.78 (0.25-12.76)	.56
Mini-IPIP personality factor score, mean (SD)[b]					
Extraversion	3.6 (0.7)	3.5 (0.8)	–0.34 (0.33)	0.72 (0.37-1.37)	.31
Agreeableness	4.4 (0.5)	4.4 (0.5)	–0.12 (0.47)	0.88 (0.35-2.23)	.79
Conscientiousness	3.9 (0.7)	4.2 (0.6)[a]	0.78 (0.35)	2.17 (1.09-4.31)	.03
Neuroticism	2.5 (0.6)	2.5 (0.7)	–0.12 (0.36)	0.88 (0.44-1.79)	.74
Intellect and imagination	4.0 (0.7)	3.9 (0.7)	0.09 (0.32)	1.10 (0.58-2.06)	.78
Patients (n = 953)[c]					
Female sex, %	22	27[a]	0.43 (0.16)	1.54 (1.13-2.09)	.01
White race/ethnicity, %	36	28[a]	–0.35 (0.14)	0.70 (0.53-0.93)	.01
Age, mean (SD), y	15.8 (1.3)	15.9 (1.5)	0.07 (0.05)	1.08 (0.98-1.18)	.13
Criminal justice involvement, %	64	68	0.26 (0.14)	1.30 (0.98-1.72)	.07
Prior substance use treatment, %	37	28[a]	–0.44 (0.14)	0.65 (0.49-0.86)	<.001
% Days abstinent, mean (SD)	58.5 (36.6)	66.2 (35.7)[a]	0.60 (0.19)	1.82 (1.26-2.64)	<.001

[a] $P < .05$.
[b] Based on the 20-item short form of the International Personality Item Pool (IPIP) 5-factor model (score range, 1-5 [low to high]).[43]
[c] Number of patients with complete data who could be used for propensity analysis.

Table 2. Main Findings With Propensity Score Adjustments[a]

Outcome Measure Predictor	Event Rate Ratio or Odds Ratio (95% CI)[b]	*P* Value
Model 1		
Therapist-level demonstration of A-CRA competence		
Organization-level assignment to P4P condition	2.24 (1.12-4.48)	.02
Therapist-level propensity score adjustment measure	0.55 (0.14-2.14)	.38
Model 2		
Patient-level receipt of target A-CRA		
Organization-level assignment to P4P condition	5.19 (1.53-17.62)	.01
Therapist-level propensity score adjustment measure	0.27 (0.02-3.89)	.34
Patient-level propensity score adjustment measure	1.16 (0.31-4.62)	.83
Model 3		
Patient-level 6-mo remission status		
Organization-level assignment to P4P condition	0.68 (0.35-1.33)	.25
Therapist-level propensity score adjustment measure	1.72 (0.42-7.06)	.45
Patient-level propensity score adjustment measure	148.43 (27.49-801.34)	<.001

Abbreviations: A-CRA, Adolescent Community Reinforcement Approach; P4P, pay-for-performance.
[a] The intracluster correlation coefficient could not be computed for model 1 because it uses a Poisson distribution. The intracluster correlation coefficient could not be computed for model 2 and model 3 because they use a Bernoulli distribution.
[b] Event rate ratios are reported for model 1. Odds ratios are reported for model 2 and model 3.

COMMENT

Findings from this trial suggest that P4P can be an effective method of improving implementation of evidence-based treatment in practice settings. As hypothesized, we found that offering monetary bonuses directly to therapists had a large effect on increasing their demonstration of (1) monthly competency in implementing treatment procedures with patients and (2) the delivery of a predefined threshold level of treatment to adolescent patients. Given the numerous calls for research to experimentally test the effectiveness of using P4P methods,[2,3,8,9] these findings represent a significant addition to the existing P4P literature.

Despite the large treatment implementation effects observed between study conditions, the observed rates of A-CRA competence and target A-CRA had considerable room for improvement even within the P4P condition. However, it is important to understand that the introduction of monetary incentives necessitated that the implementation measures were based on objective criteria (ie, expert review of actual session recordings) as opposed to therapist self-report. Therefore, therapists were required to record and submit many of their treatment sessions (≥80% of their sessions each month for A-CRA competence and ≥7 sessions per patient for target A-CRA) as part of the criteria for demonstrating the 2 implementation measures that were reinforced in this P4P experiment. We believe that it is essential to have independent and objective measurement of treatment fidelity to achieve quality implementation of evidence-based treatments when using P4P approaches. However, our study findings suggest that compliance with documen-

tation is an area to be addressed as part of future P4P research with treatment providers. It also is important to clarify that target A-CRA represents a very high threshold of A-CRA treatment. Indeed, in prior randomized clinical trials of A-CRA, only 34% of patients received target A-CRA based on therapist-reported procedures delivered (not taped reviews).[35] If we had used therapist report in this study, the rates of target A-CRA would have been higher in both conditions (28.9% for the P4P condition and 14.4% for the IAU condition).

With regard to treatment effectiveness, the rates of remission observed in both conditions of this study were substantially higher than the 24% mean remission rate observed in the Cannabis Youth Treatment study.[31] However, these higher-than-expected rates of remission made it difficult for the P4P intervention to produce a significant incremental difference between the 2 study conditions. Therefore, we did not find support for our hypothesis that patients in the P4P condition would have significantly higher remission rates at the end of treatment, despite post hoc analyses that revealed a significant relationship between target A-CRA and remission status. Although the lack of a direct effect of P4P on patient remission status might be explained by the higher-than-expected remission rates for both groups (ie, ceiling effect), the poor overall patient follow-up rate of 60.9% makes it difficult to draw strong conclusions about the true effect of P4P on remission status.

In addition to having important strengths (eg, randomized design), this study has substantial limitations to be acknowledged. For example, because the study group for this trial was a convenience sample of 29 treatment organizations participating in a well-resourced national initiative to implement evidence-based treatment for adolescent substance use disorders, the extent to which these findings will generalize to other treatments, settings, or populations needs further testing. In addition, although therapist compliance with the submission of recorded sessions did not limit our ability to examine the effect of P4P on our 2 primary treatment implementation measures given that it was an explicit part of demonstrating achievement, this issue limited our ability to draw stronger conclusions about the relationship between these 2 implementation measures and the patient treatment effectiveness outcome (ie, remission status). Also, the generally low patient follow-up rate combined with the differential patient attrition between conditions made it difficult to reach conclusions about the true effect that the P4P intervention had on improving patient-level remission status. Finally, because biometric data (eg, breathalyzer and urine test results) were not collected, it was impossible to verify the accuracy of patients' self-reported remission status.

In conclusion, this study provides experimental support for the effectiveness of using P4P as a method to improve implementation of evidence-based treatments in practice settings. In addition to examining whether the P4P intervention might have had an effect on other treatment effectiveness measures (eg, days of abstinence and substance use–related problems), future research is needed to examine the extent to which the P4P approach used in this study was cost-effective. Although cost-effectiveness studies[7,8,20,21] have been a rare area of P4P research, they are critically important given that potential funders of P4P programs will need information about what to expect for a return on their investments in such endeavors.[20]

Accepted for Publication: April 4, 2012.
Published Online: August 13, 2012. doi:10.1001/archpediatrics.2012.802
Correspondence: Bryan R. Garner, PhD, Lighthouse Institute, Chestnut Health Systems, 448 Wylie Dr, Normal, IL 61761 (brgarner@chestnut.org).
Author Contributions: Dr Garner had full access to all the data in the study and takes responsibility for the integrity of the data and accuracy of the data analysis. *Study concept and design:* Garner, S. Godley, Dennis, Bair, and M. Godley. *Acquisition of data:* Garner, S. Godley, Dennis, and Bair. *Analysis and interpretation of data:* Garner and Hunter. *Drafting of the manuscript:* Garner and Hunter. *Critical revision of the manuscript for important intellectual content:* Garner, S. Godley, Dennis, Hunter, Bair, and M. Godley. *Statistical analysis:* Garner and Hunter. *Obtaining funding:* Garner, S. Godley, Dennis, and M. Godley. *Administrative, technical, or material support:* Garner, S. Godley, Dennis, and Bair. *Study supervision:* Garner, Dennis, Bair, and M. Godley.
Financial Disclosure: None reported.
Funding/Support: Financial assistance for this study was provided by grant R01-AA017625 from the National Institute on Alcohol Abuse and Alcoholism and by grants TI17589, TI17604, TI17605, TI17638,TI17646, TI17673, TI17702, TI17719, TI17724, TI17728, TI17742, TI17744, TI17751, TI17755, TI17761, TI17763, TI17765, TI17769, TI17775, TI17779, TI17786, TI17788, TI17812, TI17817, TI17830, TI17847, TI17864, TI19313, and TI19323 and contract 270-07-0191 from the Substance Abuse and Mental Health Services Administration's Center for Substance Abuse Treatment.
Role of the Sponsors: The study sponsors and funders had no involvement in the design or conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript.

REFERENCES

1. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
2. Dudley RA. Pay-for-performance research: how to learn what clinicians and policy makers need to know. *JAMA*. 2005;294(14):1821-1823.
3. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA*. 2005;294(14):1788-1793.
4. Doran T, Fullwood C, Gravelle H, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med*. 2006;355(4):375-384.
5. Garner BR, Godley SH, Bair CML. The impact of pay-for-performance on therapists' intentions to deliver high-quality treatment. *J Subst Abuse Treat*. 2011;41(1):97-103.
6. McLellan AT, Kemp J, Brooks A, Carise D. Improving public addiction treatment through performance contracting: the Delaware experiment. *Health Policy*. 2008;87(3):296-308.
7. Nahra TA, Reiter KL, Hirth RA, Shermer JE, Wheeler JRC. Cost-effectiveness of hospital pay-for-performance incentives. *Med Care Res Rev*. 2006;63(1)(suppl):49S-72S.
8. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med*. 2006;145(4):265-272.

9. Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev*. 2006;63(2):135-157.
10. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med*. 2006;355(18):1895-1902.
11. Shepard DS, Calabro JAB, Love CT, McKay JR, Tetreault J, Yeom HS. Counselor incentives to improve client retention in an outpatient substance abuse aftercare program. *Adm Policy Ment Health*. 2006;33(6):629-635.
12. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med*. 2007;357(2):181-190.
13. Epstein AM. Pay for performance at the tipping point. *N Engl J Med*. 2007;356(5): 515-517.
14. Glickman SW, Ou FS, DeLong ER, et al. Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*. 2007;297(21):2373-2380.
15. Lindenauer PK, Remus D, Roman S, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med*. 2007;356(5):486-496.
16. Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA*. 2007;297(7):740-744.
17. An LC, Bluhm JH, Foldes SS, et al. A randomized trial of a pay-for-performance program targeting clinician referral to a state tobacco quitline. *Arch Intern Med*. 2008;168(18):1993-1999.
18. Bremer RW, Scholle SH, Keyser D, Houtsinger JVK, Pincus HA. Pay for performance in behavioral health. *Psychiatr Serv*. 2008;59(12):1419-1429.
19. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med*. 2009; 361(4):368-378.
20. Greene SE, Nash DB. Pay for performance: an overview of the literature. *Am J Med Qual*. 2009;24(2):140-163.
21. Emmert M, Eijkenaar F, Kemter H, Esslinger AS, Schöffski O. Economic evaluation of pay-for-performance in health care: a systematic review [published online June 10 2011]. *Eur J Health Econ*. doi:10.1007/s10198-011-0329-8. 21660562.
22. Scott A, Sivey P, Ait Ouakrim D, et al The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochran Database Syst Rev*. 2011;(9):CD008451. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD008451.pub2/abstract;jsessionid=80CA53339EF0B599C69F8796E06C1458.d01t02. Accessed May 31, 2012.
23. Vandrey R, Stitzer ML, Acquavita SP, Quinn-Stabile P. Pay-for-performance in a community substance abuse clinic. *J Subst Abuse Treat*. 2011;41(2):193-200.
24. Institute of Medicine. *Rewarding Provider Performance: Aligning Incentives in Medicare*. Washington, DC: National Academy Press; 2007.
25. Institute of Medicine. *Bridging the Gap Between Practice and Research: Forging Partnerships With Community-Based Drug and Alcohol Treatment*. Washington, DC: National Academy Press; 1998.
26. Fixsen DL, Naoom SF, Blasé KA, Friedman RM, Wallace F. *Implementation Research: A Synthesis of the Literature*. Tampa, FL: National Implementation Research Network; 2005.
27. Rohrbach LA, Grana R, Sussman S, Valente TW. Type II translation: transporting prevention interventions from research to real-world settings. *Eval Health Prof*. 2006;29(3):302-333.
28. Garner BR. Research on the diffusion of evidence-based treatments within substance abuse treatment: a systematic review. *J Subst Abuse Treat*. 2009;36 (4):376-399.
29. Garner BR, Godley SH, Dennis ML, Godley MD, Shepard DS. The Reinforcing Therapist Performance (RTP) experiment: study protocol for a cluster randomized trial. *Implement Sci*. 2010;5:5 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2824685/?tool=pubmed. Accessed May 27, 2012.
30. National Center on Addiction and Substance Abuse (CASA) at Columbia University. *Adolescent Substance Use: America's #1 Public Health Problem*. New York, NY: CASA; 2011.
31. Dennis ML, Godley SH, Diamond G, et al. The Cannabis Youth Treatment (CYT) study: main findings from two randomized trials. *J Subst Abuse Treat*. 2004; 27(3):197-213.
32. Godley MD, Godley SH, Dennis ML, Funk RR, Passetti LL. The effect of assertive continuing care on continuing care linkage, adherence and abstinence following residential treatment for adolescents with substance use disorders. *Addiction*. 2007;102(1):81-93.
33. Godley SH, Garner BR, Passetti LL, Funk RR, Dennis ML, Godley MD. Adolescent outpatient treatment and continuing care: main findings from a randomized clinical trial. *Drug Alcohol Depend*. 2010;110(1-2):44-54.
34. Godley SH, Garner BR, Smith JE, Meyers RJ, Godley MD. A large-scale dissemination and implementation model for evidence-based treatment and continuing care. *Clin Psychol (New York)*. 2011;18(1):67-83.
35. Garner BR, Godley SH, Funk RR, Dennis ML, Smith JE, Godley MD. Exposure to Adolescent Community Reinforcement Approach treatment procedures as a mediator of the relationship between adolescent substance abuse treatment retention and outcome. *J Subst Abuse Treat*. 2009;36(3):252-264.
36. Smith JE, Lundy SL, Gianini L. *Community Reinforcement Approach (CRA) and Adolescent Community Reinforcement Approach (A-CRA) Coding Manual*. Albuquerque: University of New Mexico; 2007.
37. Bucklin BR, Dickinson AM. Individual monetary incentives: a review of different types of arrangements between performance and pay. *J Organ Behav Manag*. 2001;21(3):45-137.
38. Dennis ML, Titus JC, White M, Unsicker J, Hodgkins D. *Global Appraisal of Individual Needs (GAIN): Administration Guide for the GAIN and Related Measures*. Bloomington, IL: Chestnut Health Systems; 2003.
39. Charpentier PA. *Urn Randomization Program gRand* [computer program]. Version 1.10. New Haven, CT: Yale University; 2003.
40. Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med*. 2009;6(5):e1000065 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2668175/?tool=pubmed. Accessed May 27, 2012.
41. Raudenbush SW, Bryk AS, Congdon R. *HLM 6 for Windows* [computer program]. Lincolnwood, IL: Scientific Software International Inc; 2004.
42. Bedard PL, Krzyzanowska MK, Pintilie M, Tannock IF. Statistical power of negative randomized controlled trials presented at American Society for Clinical Oncology annual meetings. *J Clin Oncol*. 2007;25(23):3482-3487.
43. Donnellan MB, Oswald FL, Baird BM, Lucas RE. The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychol Assess*. 2006; 18(2):192-203.