Understanding Results from Randomized Trials: Use of Program- and Client-Level Data to Study Medical and Nonmedical Treatment Programs*

LEE ANN KASKUTAS, DR.PH.,† LYNDSAY AMMON, M.P.H.,† JANE WITBRODT, M.P.H., KAREN GRAVES, PH.D.,† SARAH ZEMORE, PH.D., THOMASINA BORKMAN, PH.D.,† AND STEPHANIE WEISNER, B.A.

Alcohol Research Group, 2000 Hearst Avenue, Suite 300, Berkeley, California 94709

ABSTRACT. Objective: To address unanticipated results from randomized trials, researchers often focus on client-level data about services received during treatment. Program-level observations can also be helpful, especially in understanding treatment delivered in groups. Using both approaches, this article strives to better understand inconsistent results from a trial comparing medical and nonmedical group-format day treatment. **Method:** Dependent treatment seekers were randomized to a hospital-based medical day-treatment program or to one of two community-based nonmedical day treatment programs. Services received during treatment were captured from clients using the Treatment Services Review ($N = 230$ subjects; 78 women), and group sessions were observed to measure therapeutic style using an Event Form ($N = 48$ observations). **Results:** The trial had found better medical outcomes at the hospital than at either nonmedical program, but most other tests had demonstrated similar outcomes at day hospital and one of the nonmedical sites and worse outcomes (psychiatric, family/social and employment) at the other nonmedical study site. Our analysis of services reported by study participants found a pattern of fewer substance misuse-oriented groups and less serious discussions about medical, psychiatric and family/social problems at the same nonmedical site that had worse outcomes. The way that services were delivered at that site further helped to explain the poorer outcomes there: Groups tended to be more didactic, classroom-like and less discussion-oriented. **Conclusions:** Although services received are helpful in explaining treatment outcome, treatment observation adds explanatory value. Without increasing the cost of service provision, programs that rely heavily on didactic approaches might improve their outcomes simply by encouraging more interactive discussions that engage the clients. (*J. Stud. Alcohol* **66:** 682-687, 2005)

IN A RANDOMIZED TRIAL comparing outcomes and costs of day hospital treatment versus nonmedical day treatment for chemical dependency (Kaskutas et al., 2004), patients at the day hospital study site were less likely to report medical problems at follow-up than those at either of the two nonmedical study sites (as hypothesized). However, most other tests demonstrated similar outcomes at day hospital and one of the nonmedical study sites and worse outcomes at the other nonmedical study site (which also was less costly). This article attempts to explain this pattern of effects using client-level data about services received during treatment as well as observational data about therapeutic styles at the study sites.

Understanding the mechanisms of action that explain superior treatment outcomes in clinical trials has recently received increased research attention. For example, in attempting to explain the lack of support for hypothesized differences in Project MATCH (Matching Alcoholism Treatments to Client Heterogeneity), causal chain analysis was used to test theoretically based models of the change processes expected to underlie differences in treatment interventions (Longabaugh and Wirtz, 2001). That quantitatively oriented, client-level analysis was augmented by observational data about counselor style (Carroll et al., 1998). Although our study of day hospital versus nonmedical day treatment was of a much smaller scale than Project MATCH, a similar post hoc analysis of trial results is possible here as well. We study whether clients at the nonmedical program with worse outcomes received fewer services during their treatment there and whether there is evidence of a lapse in program fidelity at that program.

Received: September 27, 2004. Revision: April 26, 2005.

*This study was funded by grant R01 AA11279 from the National Institute on Alcohol Abuse and Alcoholism.

†Correspondence may be sent to Lee Ann Kaskutas at the above address, or via email at: lkaskutas@arg.org. Lyndsay Ammon is also with the University of California, Berkeley, School of Public Health, Berkeley, CA. Karen Graves is a research consultant in Ashland, VA. Thomasina Borkman is with the Department of Sociology and Anthropology, George Mason University, Fairfax, VA.

Method

Study design

Between May 1998 and December 2000, 271 adult alcohol- and/or drug-dependent individuals from northern California Bay Area communities were recruited for this study. Eighty-five percent ($n = 230$) of the participants were successfully re-interviewed at each of the three follow-ups (end of the treatment, 6 months and 12 months) and are the focus of the current study. Fieldwork details are reported in

Kaskutas et al. (2004). Briefly, the study randomized participants to either a day hospital treatment program or a community-based nonmedical day treatment program. Initially, the trial included one medical day hospital program and one nonmedical community-based program. Halfway through the trial, serious concerns about treatment quality at the nonmedical study program led us to stop randomizing study participants to it. For example, groups routinely started late, clients had to wait while copies of handouts were made, and staff often did not know clients' names. Another nearby nonmedical day treatment program was substituted to represent the community-based arm for the second half of the study period (Kaskutas et al., 2004).

Each study site offered group-oriented day treatment in a mixed-gender setting, and all schedules included daily check-in, educational groups about addiction, 12-step-oriented groups, relapse prevention and group therapy related to psychological issues. The hospital program had several medically oriented groups led by nurses and physicians. One community program offered an employment/job-readiness group. All held sober recreational events. The study sites are described below; reported costs are based on the DATCAP (French, 2001; French et al., 1997, 2002).

Day hospital (DH) program (n = 121). This was the only study site that was hospital-based or had medical staff involved in the program. Average daily census at DH was 15 patients. Treatment staff at this 3-week program included 24-hour nursing coverage, a consulting neuropsychologist, a consulting psychiatrist, an intake coordinator, two counselors in recovery (one certified chemical dependency counselor, one with a master's degree), two others not in recovery (with master's degrees), a program director and the medical director (a medical doctor). The average cost per episode per study subject at DH was $1,274.

Community program 1 (CP1, n = 52). The initial community-based day treatment study site was part of a recovery center that also offered outpatient and family services. Its 6-week day program was held in a converted garage on the recovery center property behind the main building. Consistent with its social model roots (Borkman et al., 1998), the day treatment staff at CP1 consisted of two recovering counselors (one state certified in California), augmented by two other recovering staff that sometimes led groups. Average daily census was 16 clients, most of whom (other than study subjects) were criminal-justice referred as a condition of parole for chemical dependence-related offenses. The per-episode cost at CP1 was $526.

Community program 2 (CP2, n = 57). This 4-week community-based program was situated in a three-story house next door to its parent organization's longer-term residential Therapeutic Community (De Leon, 2000). Average daily census at CP2 was 30 clients, with about 20% criminal justice referrals. CP2 was staffed by two Ph.D. psychologists, two master's-level therapists, two California state-certified addiction counselors and three other nondegreed counselors. One psychologist, one therapist and all five counselors were in recovery. CP2 cost an average of $1,163 per episode.

Measures

Service delivery was captured at the client level using a modified Treatment Services Review (TSR; Alterman and McLellan, 1993; McLellan et al., 1992a) and at the program level using an Event Form developed by the senior author to differentiate medical from nonmedical, social model treatment (Kaskutas, 1996). We modified the TSR to additionally assess sober recreational events. Outcomes were based on the Addiction Severity Index (ASI; McLellan et al., 1985, 1992b). Because of the parallel domains captured by the TSR and ASI, we are able here to study whether services received by clients at a particular study site (based on the TSR) reflect superior outcomes at that site (based on the ASI).

The TSR and ASI have been used in a number of published studies and have demonstrated strong reliability and validity; thus for this brief report, we do not describe those measures here. The Event Form, available from the senior author, consists of a 19-item checklist that captures the therapeutic style evident in the delivery of group sessions. It discriminates staff versus peer leadership in group discussions, interactive versus didactic group format and use of formal counseling techniques versus speaking from one's own experiential knowledge in recovery. Specifically, five therapeutic styles are captured: *experiential learning* (using the counselor's direct experience in recovery to respond to issues arising during a group session); *didactic learning* (traditional classroom style, using a lecture format, handouts, blackboards, etc.); *professional counseling style* (using psychological theory and approaches to interpret client experiences and problems); *peer helping* (clients helping one another); and *interactive approach* (discussions, with counselor and clients both driving the group discussions). The form was completed immediately following (not during) the session in which the observations were made. Observers were trained by the Event Form developer using vignettes reflective of the respective five therapeutic styles. Inter-rater reliability (e.g., kappa statistics, etc.) was not studied because only one observer was allowed to be present at the group sessions. However, post hoc discussions among the ethnographers were held regularly to review recent observations and coding conventions used to interpret a given group session. There was no evidence of observer effects and the within-site ratings were consistent.

Data analysis

Statistical analyses were conducted using SPSS for Windows, Release 12.0 (SPSS Inc., Chicago, IL). To isolate

TABLE 1. Services received per week, based on participant report in the TSR: Number of serious discussions, number of groups and percentage receiving medical services[a]

Services per week	DH (n = 121)	CP1 (n = 52)	CP2 (n = 57)	Program effects	Pairwise signif.
No. of serious discussions, mean (SD)[b]					
Alcohol	2.5 (1.8)	1.2 (1.6)	2.7 (2.3)	CP2 > CP1	p = .001
				DH > CP1	p = .001
Drug	2.1 (1.9)	1.2 (1.5)	2.3 (2.3)	CP2 > CP1	p = .001
				DH > CP1	p = .001
Medical	0.5 (1.0)	0.2 (0.4)	0.5 (1.0)	DH > CP1	p = .008
Psychiatric	1.0 (1.4)	0.3 (0.7)	1.1 (1.7)	CP2 > CP1	p = .002
				DH > CP1	p = .001
Family/social	1.2 (1.4)	0.5 (0.9)	1.8 (1.7)	CP2 > CP1	p < .001
				CP2 > DH	p = .003
				DH > CP1	p = .001
Employment	1.1 (1.4)	1.0 (1.5)	1.4 (2.1)	–	NS
Legal	0.2 (0.7)	0.2 (0.7)	0.3 (0.9)	–	NS
No. of group sessions, mean (SD)[b]					
Alcohol/drug education	2.7 (1.6)	1.2 (1.6)	3.0 (2.1)	CP2 > CP1	p < .001
				DH > CP1	p < .001
Relapse prevention	1.7 (1.4)	0.6 (0.7)	2.1 (1.8)	CP2 > CP1	p < .001
				DH > CP1	p < .001
Relaxation/feedback	0.7 (1.3)	0.7 (0.9)	1.0 (1.1)	–	NS
Sober recreational events	0.6 (1.1)	0.6 (0.9)	1.8 (1.7)	CP2 > CP1	p < .001
				CP2 > DH	p < .001
AA/NA/CA	2.3 (5.2)	1.3 (1.8)	0.3 (1.6)	DH > CP2	p = .001
Medical services received (%)[c]					
Any medical visits	7	4	12	–	NS
Any prescription drug use	13	6	2	DH > CP2	p = .037

Notes: TSR = Treatments Services Review; DH = day hospital; CP = community program; signif. = significance; AA/NA/CA = Alcoholics Anonymous/Narcotics Anonymous/Cocaine Anonymous. [a]Analysis only includes those study participants who completed all three follow-up interviews for the study; [b]analysis of covariance controlling for length of stay; [c]tests of proportion.

differences in services received at the three study sites, a two-stage analysis plan was implemented. Omnibus tests were first used to test respectively for differences of means and rates across the three programs; for these, the standard significance level of $p < .05$ was used. When overall program differences emerged, pairwise tests were then undertaken, in which case Bonferroni adjustment was applied. Because there were three such sets of pairwise program comparisons (DH to CP1, DH to CP2 and CP1 to CP2), $p = .0167$ (.05 / 3) is our threshold when comparing mean services received by clients at the study sites (analysis of covariance [ANCOVA] tests; Table 1), differences in therapeutic style observed at the study sites (analysis of variance [ANOVA] tests; Table 2) and the multivariate predictors of outcomes (which included program in a final step; not shown).

Results

Sample characteristics

The study sample was demographically diverse. For example, there was strong representation for women (34%, n = 78) and Black and Hispanic ethnicity (35% Black, 21% Hispanic) and a wide spread for age (mean age was 42 years, ranging from 19 to 75 years). About two fifths were married or living with a significant other. Eighty-five percent completed a high school education or greater. Thirty percent of the sample reported an annual income below $10,000, with 35% making above $35,000 yearly. There were no significant differences in ASI problem severity between subjects randomized to the three study sites.

Treatment services

Our analysis of the TSR questions focuses on three types of services (Table 1): serious discussions, group sessions and specific medical services. These data are used to study whether differences in services received at the study sites were consistent with the pattern of key main effects found in the randomized trial, as reported in Kaskutas et al. (2004) and summarized below.

Better medical outcomes at day hospital program

Medical services received only partially disentangle our finding of superior medical outcomes at the day hospital program compared with the two nonmedical programs. For

TABLE 2. Mean number of times therapeutic styles were observed at study sites based on Event Forms

	DH Mean (SD) (n = 22)	CP1 Mean (SD) (n = 15)	CP2 Mean (SD) (n = 23)	Program effects[a]	Pairwise signif.
Therapeutic styles[b]					
Experiential learning	1.1 (1.0)	1.3 (1.2)	2.1 (1.2)	*CP2 > CP1*	*p = .045*
				CP2 > DH	p = **.008**
Didactic learning	1.3 (0.9)	2.1 (1.4)	0.87 (1.0)	**CP1 > CP2**	p = **.002**
				CP1 > DH	*p = .037*
Counseling approach	1.0 (0.8)	0.6 (0.6)	0.7 (0.5)	–	NS
Peer helping	1.6 (1.3)	1.2 (1.1)	2.4 (1.0)	**CP2 > CP1**	p = **.002**
				CP2 > DH	*p = .025*
Interactive approach	1.1 (0.8)	0.6 (0.9)	1.8 (0.7)	**CP2 > CP1**	p < **.001**
				CP2 > DH	p = **.006**

Notes: Signif. = significance. [a]Significant results bolded, marginal results italicized; [b]analysis of variance.

example, participants at DH reported more serious discussions about their medical problems than those at CP1 (p = .008), but not more than those at CP2 (Table 1). No differences emerged in other services pertaining to medical problems.

Worse psychiatric, family/social and employment outcomes at CP1

Participants at CP1 were less likely than those at CP2 or DH to be problem free at follow-ups in two ASI outcome domains (psychiatric problems and family/social problems) and their employment and family/social problems were more severe as well (Kaskutas et al., 2004). TSR data presented a fairly clear pattern of significantly fewer services addressing psychiatric and family/social problems at CP1 than at the other two study sites. However, the mean number of employment services was not different among the three programs (Table 1).

Better alcohol outcomes at CP2 than CP1

In terms of alcohol and drug outcomes, the main effects from the clinical trial found more alcohol abstinence and reduced odds of reporting alcohol problems at follow-up for those randomized to CP2 than CP1 (no differences had been found between DH and either community program; Kaskutas et al., 2004). The TSR data showed that participants at CP1 received significantly fewer alcohol and drug education groups and fewer relapse prevention groups than those at CP2 or at DH (Table 1). In addition, subjects randomized to CP2 reported significantly more sober recreational events than subjects at either DH or CP1 but fewer Alcoholics Anonymous/Narcotics Anonymous/Cocaine Anonymous meetings than DH (Table 1). Thus the TSR data that most closely reflected the better alcohol outcomes at CP2 versus CP1 was sober recreational events, as that was the only substance misuse group that showed similar rates for CP1 and DH clients but better rates for CP2 than CP1 clients.

Therapeutic style

Differences in therapeutic styles observed at the three study sites using the Event Form are shown in Table 2. There were significantly more observations of experiential learning style at CP2 than DH (p < .0167), more peer helping and interactive discussions at CP2 than at CP1 (p < .0167), more interactive discussions at CP2 than DH (p < .0167) and more instances of a didactic approach to the group sessions at CP1 than at CP2 (p < .0167). In addition, there were marginal results suggesting more experiential learning at CP2 than CP1 (p = .045), more didactic education at CP1 than DH (p = .037) and more peer helping at CP2 than DH (p = .025).

Reconsidering the main effects (results not shown)

Given the mixed pattern of evidence for the trial's main effects based on analyses of services received by study participants at the study sites, we next considered problem status at both follow-ups from a multivariate perspective that took into account baseline problem severity (entered in Step 1), services received at the treatment program that addressed the problem domain (Step 2) and the program effect (Step 3). Should program effects emerge after adjusting for initial problem severity and services received during treatment, this would suggest that other aspects of treatment (not captured by the TSR) account for the main (program) effects. For each of the seven ASI outcome domains, more severe problems at baseline were predictive of having problems at both follow-ups. However, in no cases did services received at the study program significantly predict problem status at follow-up. Instead, program effects

again emerged as highly predictive, with superior psychiatric (odds ratio [OR]: 3.2, 95% CI: 1.31-7.60) and family/social (OR: 4.6, 95% CI: 1.80-12.23) outcomes at DH than at CP1. These results further suggest that there are differences in the treatment programs that are not reflected by client-level services tracked by the TSR.

Discussion

Our analysis of client-level data on services received during treatment was useful in disentangling several of the differences reported from our randomized trial comparing medical and nonmedical chemical dependency programs (Kaskutas et al., 2004). That study had found better medical outcomes at the hospital program than at either nonmedical community program; worse psychiatric, family/social and employment problems at CP1 than at DH or CP2; and worse alcohol outcomes at CP1 than at CP2. Studying differences in services received among subjects in that study, we found here that more medical services had been reported at DH than at CP1 only; and significantly more alcohol, family/social and psychiatric services had been reported at both DH and CP2 than at CP1. Thus the only outcome domain for which the TSR data was not informative in disentangling outcomes was employment.

To further understand the program-level differences, data from the Event Form observations were used to study differences in therapeutic styles at the study sites. For these analyses, a mismatch between observed and expected therapeutic styles would signal a lapse in treatment fidelity, which in turn could help to explain reduced therapeutic effectiveness at a study site. Here we assumed that a good medically oriented chemical dependency program could be expected to use professional counseling techniques and didactic educational approaches that draw on medical and psychiatric training; and a good nonmedical community program should be expected to exploit the counselors' (and clients') experiential knowledge of recovery, encourage peer helping and avoid didactic lectures in favor of dynamic interactive groups. True to its orientation, CP2 employed experiential learning more than DH, but CP1 did not. CP1 was more didactic than CP2 and at least as didactic as DH and did less peer helping and held fewer interactive groups than CP2—demonstrating a lack of fidelity to what should have been its key therapeutic ingredients.

Reliance on didactic approaches coupled with lack of interaction during groups may be especially ineffective for programs whose staff expertise lies in their own personal knowledge of recovery. In this regard, the Event Form data were especially helpful in distinguishing treatment quality and fidelity between nonmedical community-based programs. As noted earlier, the decision to stop randomizing subjects to CP1 was made because of concerns about treatment quality that arose during our routine program observations. Although those observations employed the Event Form, the decision to stop using CP1 was not based on an analysis of the Event Form data. It is reinforcing to see our decision borne out by the empirical analysis of the Event Form data and suggests that providers and researchers interested in assessing treatment quality (in particular at nonmedical community-based programs) might consider therapeutic style—especially along the dimensions of didacticity, interaction, peer helping and experiential learning.

The combination of client- and program-level data helped to disentangle some (but not all) of the differences in the randomized trial comparing medical and nonmedical chemical dependency programs. The explanatory value of another program-level difference, cost, must also be considered. The average per-episode cost per subject at CP1 was significantly lower than those for CP2 and DH (Kaskutas et al., 2004). It is possible that below a certain threshold, quality services cannot be delivered. However, it is likely that pressure to reduce costs will continue to be exerted on treatment programs, and programs need to be aware of what they can do to improve outcomes without increasing costs. Our results suggest that treatment providers and payers might consider whether the programs they are supporting rely heavily on didactic approaches (i.e., rote lectures, written handouts, lengthy written exercises, etc.). Without increasing the cost of service provision, such programs might improve their outcomes simply by encouraging more peer helping and interactive discussions in treatment.

Limitations

In addition to lower costs, CP1 is distinguished from the other study sites in other ways that might be responsible for the worse outcomes at that program. For example, the group sessions were held in a converted garage, and there were no professionally trained staff. These differences may also have contributed to the generally poor outcomes among clients randomized to CP1, and it is possible that they may be more important (to outcomes) than the observed differences in therapeutic style or in quantity of services delivered. Although it is not possible to assess this from our data, the current study suggests some areas of program differences. These differences may be as important in influencing patient outcomes as whether a program is community- or hospital-based.

Acknowledgments

The authors thank Dr. Jason Bond for his statistical advice and Dr. Constance Weisner for her thoughtful comments on the manuscript.

References

Alterman, A.I. and McLellan, A.T. Inpatient and day hospital treatment services for cocaine and alcohol dependence. J. Subst. Abuse Treat. **10:** 269-275, 1993.

Borkman, T.J., Kaskutas, L.A., Room, J., Bryan, K. and Barrows, D. An historical and developmental analysis of social model programs. J. Subst. Abuse Treat. **15:** 7-17, 1998.

Carroll, K.M., Connors, G.J., Conney, N.L., Diclemente, C.C., Donovan, D.M., Kadden, R.R., Longabaugh, R.L., Rounsaville, B.J., Wirtz, P.W. and Zweben, A. Internal validity of Project MATCH treatments: Discriminability and integrity. J. Cons. Clin. Psychol. **66:** 290-303, 1998.

De Leon, G. The Therapeutic Community: Theory, Model, and Method, New York: Springer, 2000.

French, M.T. Drug Abuse Treatment Cost Analysis Program (DATCAP): Program Version, Seventh Version (available at: www.datcap.com), Coral Gables, FL: University of Miami, 2001.

French, M.T., Dunlap, L.J., Zarkin, G.A., McGeary, K.A. and McLellan, A.T. A structured instrument for estimating the economic cost of drug abuse treatment: The Drug Abuse Treatment Cost Analysis Program (DATCAP). J. Subst. Abuse Treat. **14:** 445-455, 1997.

French, M.T., Salomé, H.J., Sindelar, J.L. and McLellan, A.T. Benefit-cost analysis of addiction treatment: Methodological guidelines and empirical application using the DATCAP and ASI. Hlth Serv. Res. **37:** 433-455, 2002.

Kaskutas, L.A. An Observer Event Form for Categorizing Processes of Alcohol Treatment: Appendix A, Rockville, MD: Center for Substance Abuse Treatment, 1996.

Kaskutas, L.A., Witbrodt, J. and French, M.T. Outcomes and costs of day hospital treatment and nonmedical day treatment for chemical dependency. J. Stud. Alcohol **65:** 371-382, 2004.

Longabaugh, R. and Wirtz, P.W. Substantive review and critique. In: Longabaugh, R. and Wirtz, P.W. (Eds.) Project MATCH Hypotheses: Results and Causal Chain Analyses. NIAAA Project MATCH Monograph Series, Vol. 8, NIH Publication No. 01-4238, Washington: Government Printing Office, 2001, pp. 305-330.

McLellan, A.T., Alterman, A.I., Cacciola, J., Metzger, D. and O'Brien, C.P. A new measure of substance abuse treatment: Initial studies of the treatment services review. J. Nerv. Ment. Dis. **180:** 101-110, 1992a.

McLellan, A.T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., Pettinati, H. and Argeriou, M. The fifth edition of the Addiction Severity Index. J. Subst. Abuse Treat. **9:** 199-213, 1992b.

McLellan, A.T., Luborsky, L., Cacciola, J., Griffith, J., Evans, F., Barr, H.L. and O'Brien, C.P. New data from the Addiction Severity Index: Reliability and validity in three centers. J. Nerv. Ment. Dis. **173:** 412-423, 1985.