The Outcome and Cost of Alcohol and Drug Treatment in an HMO: Day Hospital Versus Traditional Outpatient Regimens

Constance Weisner, Jennifer Mertens, Sujaya Parthasarathy, Charles Moore, Enid M. Hunkeler, Teh-wei Hu, and Joe V. Selby

Objective. To compare outcome and cost-effectiveness of the two primary addiction treatment options, day hospitals (DH) and traditional outpatient programs (OP) in a managed care organization, in a population large enough to examine patient subgroups.

Data Sources. Interviews with new admissions to a large HMO's chemical dependency program in Sacramento, California between April 1994 and April 1996, with follow-up interviews eight months later. Computerized utilization and cost data were collected from 1993 to 1997.

Study Design. Design was a randomized control trial of adult patients entering the HMO's alcohol and drug treatment program ($N = 668$). To examine the generalizability of findings as well as self-selection factors, we also studied patients presenting during the same period who were unable or unwilling to be randomized ($N = 405$). Baseline interviews characterized type of substance use, addiction severity, psychiatric status, and motivation. Follow-up interviews were conducted at eight months following intake. Breathanalysis and urinalysis were conducted. Program costs were calculated.

Data Collection. Interview data were merged with computerized utilization and cost data.

Principal Findings. Among randomized subjects, both study arms showed significant improvement in all drug and alcohol measures. There were no differences overall in outcomes between DH and OP, but DH subjects with midlevel psychiatric severity had significantly better outcomes, particularly in regard to alcohol abstinence (OR = 2.4; 95% CI = 1.2, 4.9). The average treatment costs were $1,640 and $895 for DH and OP programs, respectively. In the midlevel psychiatric severity group, the cost of obtaining an additional person abstinent from alcohol in the DH cohort was approximately $5,464. Among the 405 self-selected subjects, DH was related to abstinence (OR = 2.1; 95% CI = 1.3, 3.5).

Conclusions. Although significant benefits of the DH program were not found in the randomized study, DH treatment was associated with better outcomes in the self-selected group. However, for subjects with mid-level psychiatric severity in both the randomized and self-selected samples, the DH program produced higher rates

of abstention and was more cost-effective. Self-selection in studies that randomize patients to services requiring very different levels of commitment may be important in interpreting findings for clinical practice.

Key Words. Substance abuse, HMO, outcome, cost-effectiveness, generalizability

---

Studies of alcohol and drug treatment employing random assignment to inpatient and outpatient modalities suggest comparable effectiveness, despite the difference in costs (Alterman, O'Brien, McLellan, et al. 1994; Finney and Monahan 1996; McLellan, Belding, McKay, et al. 1997). Outpatient care is now the predominant treatment setting (Schmidt and Weisner 1993), but questions remain regarding the level of intensity necessary for cost-effective outcomes. The two most common models are day hospital (DH), comparable to inpatient settings in intensity, and traditional outpatient (OP) treatment, in which patients meet for two to eight hours per week. Based on evidence that higher abstinence rates are associated with more intensive treatment (Monahan and Finney 1996), we hypothesized that abstinence rates at six months would be higher among people treated in a DH program, and we conducted a randomized study of these two options in a large health maintenance organization (HMO).

A wide range of psychiatric problems and severity are found among addiction treatment populations, with the effectiveness of treatment varying by severity (Mattson and Allen 1991; McLellan, Belding, McKay, et al. 1997; Ross, Glasser, and Germanson 1988); this suggests that psychiatric symptomatology may determine in part the need for higher-intensity treatment (Project

---

This article was supported by grants from the National Institute on Alcohol Abuse and Alcoholism (R37 AA10359) and the National Institute on Drug Abuse (R01 DA08728).

Address correspondence to Constance Weisner, Dr.P.H., Senior Research Scientist, Kaiser Division of Research, 3505 Broadway, 12th Floor, Oakland CA 94611. Dr. Weisner is a Professor in the Dept. of Psychiatry, University of California, San Francisco. Jennifer Mertens, M.A. and Sujaya Parthasarathy, Ph.D. are Senior Analysts, Division of Research, Kaiser Permanente Medical Care Program, Oakland, CA; Charles Moore, M.D. is Medical Director, Kaiser Permanente Chemical Dependency Recovery Program, Sacramento, CA; Enid M. Hunkeler, M.A. is Senior Research Scientist, Division of Research, Kaiser Permanente Medical Care Program, Oakland, CA; Tehwei Hu, Ph.D. is Professor, School of Public Health, University of California at Berkeley; and Joe V. Selby, M.D., M.P.H. is Director, Division of Research, Kaiser Permanente Medical Care Program, Oakland, CA. This article, submitted to *Health Services Research* on December 16, 1998, was revised and accepted for publication on October 26, 1999.

MATCH Research Group 1993; Rounsaville, Dolinsky, Babor, et al. 1987). Based on these studies, we replicated research that had examined outcome by psychiatric status (McLellan, Luborsky, Woody, et al. 1983).

Changes in alcohol and drug treatment policy have primarily emphasized a decrease in treatment intensity, and many of the changes have been justified by findings from randomized clinical trials (Fuller 1990; Institute of Medicine 1990; Gerstein and Harwood 1990). However, it is difficult to recruit for randomized trials that assign subjects to services requiring different levels of commitment, given the wide range in addiction severity and motivation among patients and the conviction of many practitioners and clients that more intensive treatment is superior. Trials unable to recruit a large representative group may not produce results generalizable to the population needing treatment. For instance, demographic characteristics, addiction severity, motivation, and rates of starting treatment may differ between randomized subjects and patients self-selecting their treatment (Strohmetz, Alterman, and Walter 1990; U.S. National Institute on Alcohol Abuse and Alcoholism 1998). Thus, in order to assess generalizability, we included subjects in the study who were not randomly assigned to a treatment.

This study is the first to examine treatment intensity within a managed care population, to examine costs relative to effectiveness, to use sample sizes adequate for the examination of effects among patient subgroups, and to examine randomized and self-selected samples.

METHODS

Treatment Programs

The study site was the Kaiser Permanente Chemical Dependency Treatment Facility in Sacramento, California, a day hospital rehabilitation (DH) program in operation for the past two years, and a traditional outpatient (OP) program. Staff for both programs included a psychiatrist, primary care physicians, licensed masters-level social workers, and psychologists, registered nurses, and certified addiction counselors. Regular attendance and abstinence were mandatory. The content of services was the same in both programs, but the DH was structured to provide four times the intensity of the OP.

OP patients attended a 1 1/2-hour session three days per week for eight weeks for a total of 24 sessions; DH patients attended the program daily for six hours during the first three weeks. During the next five weeks, they attended the program four days per week for 1 1/2 hours per day, for a total

of 104 treatment sessions. Sessions in both programs included supportive group therapy, education, relapse prevention, and family-oriented therapy. Individual counseling, physician appointments, and pharmacotherapy were available as needed for both programs; patients were expected to attend 12-step meetings off-site, and all patients received random breathalyzer and urine screens weekly during the first four weeks and monthly thereafter. Aftercare for both programs consisted of one 1 1/2-hour group therapy meeting each week for ten months, and all patients were encouraged to attend.

Sample

Research subjects were men and women ages 18 and over, who met the criteria for drug or alcohol abuse or dependence and who, between April 1994 and April 1996, requested treatment at the program. Patients with dementia, mental retardation, or active psychosis, were not eligible. Independent research staff conducted the baseline interview, explained the two treatment options, asked subjects to accept random assignment, and obtained informed consent. The randomization rate was 62 percent ($N = 668$). Reasons for non-randomization included patient preference, time availability, workplace requirements, and clinical judgment. This non-randomized, self-selected sample represents a "treatment as usual" group. Eighty-five percent of the non-randomized group ($N = 405$) agreed to participate in all other aspects of the study.

Procedures

Research staff assessed treatment outcome six months after the end of the eight-week rehabilitation phase of treatment (referred to as the six-month follow-up). The follow-up interviews were administered by telephone from offices at the Kaiser Permanente Division of Research.

Measures

Addiction Severity. To assess substance problem severity at admission and follow-up, we used an abbreviated form of the Addiction Severity Index (ASI). The ASI is a valid and reliable instrument (McLellan, Kushner, Metzger, et al. 1992) that examines the type and severity of substance use; the severity of employment, medical, psychiatric, family/social, and legal problems; and demographic information. It measures the number, frequency, and duration of problem symptoms over the patient's lifetime and within the past 30 days, and provides a score from 0 to 1.0, with higher scores designating higher severity.

For a second psychiatric status assessment, we used the Symptom Distress Checklist, short form (SCL-66) (Lipman, Covi, and Shapiro 1979) to measure subjective symptoms of emotional distress in seven areas: somatization, obsession/compulsion, interpersonal sensitivity, depression, anxiety, hostility, and paranoid ideation.

Outcome. Consistent with the program goal, we used alcohol and drug abstinence during the past 30 days as the outcome measure; all ASI questions on alcohol and drug use during this period had to be reported negative. We also asked for the longest period of abstinence from alcohol and from drugs since subjects' treatment admission date, and we replicated the analysis with this outcome measure.

Dependence. We used questions from the Diagnostic Interview Schedule for Psychoactive Substance Dependence to provide a diagnosis for alcohol and drug (nine substance types) dependence and abuse. For each substance, we established whether the symptom was present or absent during the previous 30 days. The measurement of DSM-IV (American Psychiatric Association 1994) is conservative because one criterion, that is, continued use despite a "persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance," was not included.

Motivation. The baseline questionnaire included questions on ultimatums to enter treatment and on goals with regard to alcohol use and drug use (i.e., to stop completely versus cut down, stop for awhile, or stay at the same level).

Utilization. We used outpatient visit data from Kaiser's automated registration databases (Selby 1997) to identify utilization of subjects in the two programs (both those followed and not followed) for treatment costs. These cover all visits, including intake evaluations, medical visits, psychiatric visits, and individual counseling visits. The Treatment Services Review (TSR) (McLellan et al. 1992) was administered at two, four, six, and eight weeks and at the six-month follow-up to measure services received and out-of-plan health care utilization. Inpatient services (including stays at hospitals, trauma centers, nursing home, residential or other care facilities, including alcohol and drug treatment) and outpatient services (including visits to non-Kaiser medical doctors, psychiatrists, psychologists or other trained professionals) were measured.

Statistical Analyses. For both the randomized and self-selected samples, we employed an intent-to-treat design such that all subjects who were followed were included in the analyses, regardless of whether they returned after the intake interview to start treatment or the number of sessions they attended.

We used chi-square tests for categorical variables and *t*-tests for continuous variables to identify significant differences in baseline demographic, diagnostic, and severity characteristics by treatment group, by six-month follow-up status, and by randomization status.

We used logistic regression to examine the effect of treatment modality on abstention outcomes in both samples. These models included treatment modality as well as variables on which the two groups differed within each sample, and baseline ASI alcohol and/or drug severity. We also controlled for external and internal motivational factors by adding the employer mandate and "goal to stop using" measures (shown further on, in Table 1). Multiple linear regression was used to examine the effect of treatment modality on the longest period of abstention since admission, controlling for the measures discussed.

Finally, we examined the effect of treatment modality on outcome within stratified subgroups of psychiatric severity using logistic regression, controlling for baseline ASI alcohol and/or drug severity and type of dependence (i.e., alcohol, drug, or both). We stratified patients on level of psychiatric severity by collapsing the baseline psychiatric ASI scores into four categories: (1) zero; (2) lowest one-third of non-zero distribution (range: 0.001 to 0.406); (3) middle one-third (range: 0.407 to 0.591); and (4) highest one-third (range: 0.592 to 1). This was based on prior studies that have found differential relation to outcome by such categories, and similar distributions of the ASI by severity levels in different populations, public and private (McLellan, Luborsky, Woody, et al. 1983; Weisner, McLellan, and Hunkeler in press).

Treatment Costs. Treatment costs represent those costs applicable to the two programs by the study participants in a given time period. Program costs were determined using the program's automated general ledger. All department activities were classified into intakes, program component (individual psychotherapy and group therapy), psychiatric and medical services based on appointment information obtained from Kaiser's regional automated registration (REG+) system (Selby 1997). Relative values were assigned to each activity pool based on staff time within each pool. Costs were distributed across pools and services provided; the resulting average unit cost for each program activity multiplied by the patient's actual units of service during a given encounter provides the direct cost of the encounter. Overhead costs (facility and some administrative costs) were allocated to each unit of service via a step-down method. These methods have been used in other cost studies (French and McGeary 1997; Levin, Schmittiel, Kunz, et al. 1997).

Cost-Effectiveness Analysis. A cost-effectiveness ratio (C/E) is defined as the ratio of the cost of an intervention to the benefit (some standard measure

of clinical outcome) accrued from it. The outcomes under consideration were total abstinence, abstinence from alcohol, and abstinence from other substances. Given two mutually exclusive treatment modalities, we defined the incremental C/E ratio as the ratio of the difference in predicted average cost to the difference in the predicted probability of abstinence. We used multivariate regression models in estimating costs and effectiveness to account for differences in patient characteristics that were observed between the DH and OP cohort. Logistic regression models that included the same patient characteristics as those in the treatment outcome models were used to compute the predicted probability of abstinence for each subgroup (DH versus OP for the randomized and self-selected samples). Predicted costs were determined from ordinary least squares regression of treatment costs on the same set of predictors used in the logistic regressions for outcome. In this case, the estimated coefficient on DH directly gives the incremental cost. For total abstinence in the full model, all outcomes and costs were calculated at the mean severity level for non-white, employed subjects for standardization. We also analyzed separately the middle one-third psychiatric severity group and included dependence type among the set of predictors.

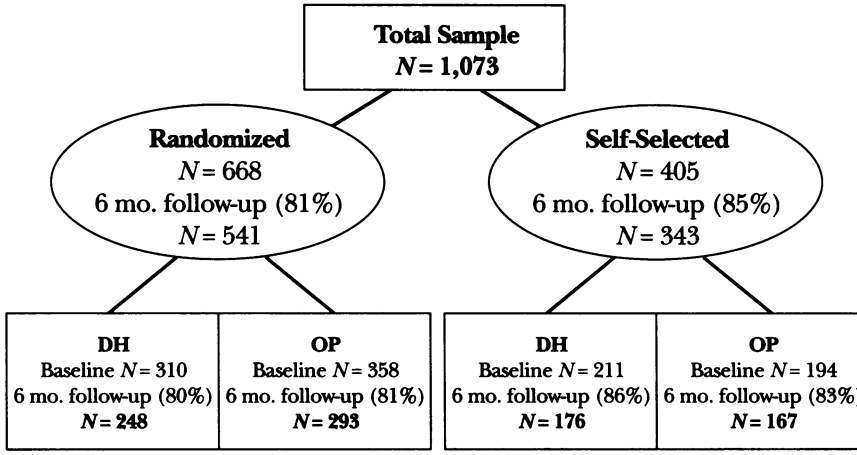
RESULTS

Subject Characteristics

Figure 1 shows the distribution of the 1,073 eligible subjects by randomization status, six-month follow-up status, and treatment group. Of these subjects, 668 (62 percent) were randomized, and 405 (38 percent) were unwilling or unable to be randomized but were willing to participate in all other study protocols (the self-selected group). A total of 884 subjects (541 randomized and 343 [82 percent] self-selected) were able to be followed for the six-month interview.

Table 1 presents baseline characteristics by treatment group and randomization status for subjects who were followed at six months and for the overall sample. We compared the two groups on 24 characteristics. Despite randomization procedures, a larger proportion of those randomized to DH than OP were unemployed and drug dependent (rather than solely alcohol dependent) and had higher drug, legal, and family ASI-measured severity problems. Differences in size of the two randomized groups approached significance ($p = .07$); we examined differences in subject characteristics and the number randomized to each program by the six-month cohort of subject recruitment and by the two interviewers who recruited 93 percent of the subjects. Subject characteristics and the proportion randomized to each

Figure 1: Randomized and Non-randomized Sample Enumeration



program differed in the same way across the four cohorts. One interviewer had a larger proportion overall of subjects randomized to the OP group (N = 184 OP versus N = 129 DH, p = .001), and the other to the DH group (N = 201 DH versus N = 186 OP, p = .430), and the proportion by interviewer to DH versus OP was not consistent across the six-month cohorts of recruitment. Interviewers were carefully trained and supervised; no methods of stratifying randomization were used. Randomization occurred at the end of the baseline interview. We control for all variables on which we found differences between groups in our multivariate analyses.

In the self-selected sample, DH subjects were more likely than OP subjects to be African American, to be unemployed, to have received employer mandates to treatment, and to have abstinence as a goal; they also had higher levels of employment, legal, and psychiatric ASI severity. A higher proportion of the self-selected DH group started treatment. Analyses of differences between those followed up and those lost to follow-up (described below) suggest that the differences by treatment group are not due to follow-up patterns.

Overall Differences Between the Randomized and Self-Selected Subjects

The differences in each program between randomized and self-selected subjects (p-values not shown) were in the same direction. Randomized subjects were younger, had higher ASI drug, employment, and family/social severity

Table 1: Baseline Characteristics by Treatment Group and Randomization Arm for Those Followed Up ($N = 884$) and for Full Sample ($N = 1{,}073$)

	Randomized				*Self-Selected*			
	Day Hospital		*Outpatient*		*Day Hospital*		*Outpatient*	
Characteristics	*Followed*	*Full*	*Followed*	*Full*	*Followed*	*Full*	*Followed*	*Full*
N	248	310	293	358	176	211	167	194
Gender (%)								
Women	35	34	35	34	35	34	37	35
Men	65	66	65	66	65	66	63	65
Ethnicity (%)								
White	75	72	80	76	70	67	83*	80
African American	11	13	8	9	16	19	8	9
Hispanic/Latino(a)	9	10	8	10	9	9	7	8
Other	4	5	4	3	5	4	3	3
Mean Age	37	37	37	37	39	39	40	39
Employment (%)								
Full/Self-employed	41	37	50*	50	39	41	68**	69
Part-time/Casual	8	8	11	11	6	6	11	9
Unemployed/Suspended	51	55	39	40	55	53	22	22
Employer-mandated treatment (%)	9	10	7	7	21	20	5**	6
Education (%)								
Less than high school	14	17	13	12	16	15	10	13
High school graduate	65	65	58	60	57	58	61	60
Some college	21	18	29	28	27	26	29	28
% with income $40K+	47	44	45	42	48	46	48	47
% Married/Living as married	46	46	47	46	45	45	50	49
DSM-IV dependence type (%)								
None	8	7	10	11	18	18	15	16
Alcohol only	38	36	47	43	45	42	52	50
Drug only	32	36	27	29	20	23	22	22
Alcohol and drug	22	21	16	18	17	17	11	12
% with goal to stop any substance used at intake	72	72	74	72	81	80	72	70
ASI means								
Alcohol	0.459	0.433	0.477	0.458	0.476	0.454	0.461	0.456
Drug	0.138	0.144	0.108**	0.118	0.110	0.119	0.090	0.091
Employment	0.396	0.421	0.385	0.384	0.391	0.408	0.313**	0.335
Medical	0.393	0.403	0.377	0.389	0.360	0.352	0.307	0.299
Legal	0.118	0.121	0.084*	0.089	0.115	0.121	0.072*	0.081
Family/Social	0.450	0.462	0.384**	0.398	0.354	0.361	0.337	0.348
Psychiatric	0.418	0.427	0.409	0.420	0.415	0.419	0.355*	0.358

Note: *P*-values compare DH with OP among those followed up: *$p < .05$; **$p < .01$.

scores, and were more likely to meet criteria for drug dependence; they were less likely to start treatment.

Loss at Follow-up

We found no bias introduced by differential follow-up rates between randomized and self-selected subjects, or between DH and OP programs in follow-up rates (see Figure 1). For both programs, those lost to follow-up were more likely to be younger, African American, of lower income, less educated, and drug-only dependent, with higher ASI severity scores in all domains except alcohol (on which they were less severe) and medical.

For both randomized and self-selected subjects, the utilization and cost patterns observed among those lost to follow-up (by treatment modality) mirrored those who were followed up. However, the non-responders (both DH and OP patients) were lower users of services when compared to those followed, and the decline in utilization over time was more rapid among those who were not followed up. For DH and OP subjects (regardless of randomization status), those who were followed had longer lengths of stay compared to those not followed, although in both samples (randomized and self-selected), the gap was larger for DH subjects and the followed DH subjects were more likely to start treatment than those not followed. The treatment costs presented here are thus higher than if there had been no loss to follow-up.

Early attrition from the treatment programs was similar to that found in prior studies of outpatient treatment (Carroll 1995). The use of conservative criteria showed, based on registration data, that among the randomized sample, 50 percent of DH patients and 59 percent of OP patients dropped out of treatment within three weeks; another 23 percent of DH patients and 17 percent of OP patients dropped out between four and eight weeks; 27 percent of DH and 24 percent of OP patients dropped out after nine weeks. Since the DH group had higher rates of retention during the most intense phase of treatment, but lower during the next five weeks, it is not clear that attrition would have influenced the effectiveness results.

Out-of-Plan Utilization

Only 27 patients (3 percent) reported out-of-plan medical services utilization; most of these were doctor office visits. Seven percent of the sample reported receiving some out-of-plan alcohol and drug treatment; 4 percent reported receiving other non-medical health care services (e.g., social worker, psychologist visits). We found no significant differences between DH and OP in out-of-plan utilization rates reported for the randomized and self-selected samples.

Validity of Self-Report Data

We conducted a validity test of the 30-day self-report data by using urinalysis for subjects who had one or more urine tests independent of the interview before the eight-week or six-month follow-ups. Urinalysis tested for alcohol, barbiturates, benzodiazopines, cocaine, cannabis/THC, opiates, phencyclidine, and amphetamines. Initial screening was performed with the Hitachi microparticle immunoassay technique; confirmation was done using GCMS (gas chromatography mass spectrometry) with the exception of confirmation for THC and benzodiazopines, which was performed with thin-layer chromatography. If benzodiazopines were detected in the initial screen but undetected in the second, then Remedi (high-pressure liquid chromatography) was employed. Comparisons yielded agreement rates of 93.4 percent. Only 2.5 percent tested positive for a substance but reported no use; moreover, only 4.1 percent tested negative for all substances but reported having used one or more substances. This agreement rate is in the range of or above agreement rates in published validation studies (Harrison 1995; Zanis, McLellan, and Randall 1994).

RANDOMIZED SAMPLE

Six-Month Abstention Rates

Subjects in both DH and OP programs showed significant improvement at the six-month follow-up on both ASI alcohol and drug severity composite scores ($p < .0001$ for each composite score in each treatment group, data not shown). No significant differences were found between DH and OP subjects for total abstention (56 percent versus 52 percent, respectively), alcohol abstention (66 percent versus 63 percent, respectively), and drug abstention (78 percent for both groups). A higher proportion of DH than OP subjects initiated treatment (79 percent versus 71 percent, respectively).

Effects of Treatment Modality

In multivariate models controlling for baseline differences between DH and OP (Table 2), DH was associated with slightly, but not significantly greater likelihood of each abstention measure. Alcohol severity scores were inversely related to the likelihood of alcohol abstention. Drug severity scores were also inversely related to alcohol, total, and drug abstention. Abstinence goals were predictive of total and alcohol abstention, and to a lesser extent, of drug abstention. A model that examined the effects of treatment modality on the

Table 2: Logistic Regression Models Predicting Abstention at Six Months: Randomized Sample

	Total Abstention			*Alcohol Abstention*			*Drug Abstention*		
	Odds Ratio	*95% CI*		*Odds Ratio*	*95% CI*		*Odds Ratio*	*95% CI*	
		Lower	*Upper*		*Lower*	*Upper*		*Lower*	*Upper*
Full randomized sample (N)	(515)*			(520)*			(516)*		
Intercept coefficient	−0.2682			0.4618			1.2660		
Day hospital	1.26	0.87	1.81	1.24	0.85	1.82	1.09	0.70	1.70
Alcohol ASI severity	0.68	0.37	1.28	0.33	0.17	0.63	—	—	—
Drug ASI severity	0.09	0.02	0.53	0.13	0.02	0.76	0.01	0.00	0.10
Employed status	1.32	0.91	1.90	1.17	0.80	1.71	1.30	0.84	2.02
Legal ASI severity	1.98	0.78	5.05	1.96	0.73	5.27	0.91	0.31	2.70
Family/Social ASI severity	1.09	0.57	2.11	0.93	0.47	1.85	1.28	0.58	2.82
Employer mandate	1.97	0.97	4.00	1.38	0.67	2.85	2.17	0.81	5.82
Goal to stop substance use	1.95	1.30	2.94	2.55	1.68	3.89	1.53	0.96	2.42
Midlevel psychiatric severity† *(N)*	(152)*			(153)*			(154)*		
Day hospital	1.91	0.98	3.74	2.36	1.15	4.85	0.93	0.44	1.97
Highest-level psychiatric severity† *(N)*	(150)*			(151)*			(150)*		
Day hospital	0.98	0.50	1.91	0.89	0.45	1.78	0.97	0.45	2.10

*Sample sizes vary because logistic regression deletes cases with missing values on any variable in the model.

†Models within psychiatric severity groups control for baseline alcohol ASI severity (for models of total and alcohol abstention), drug ASI severity (for models of total and drug abstention), and dependence type (no dependence, alcohol only, drug only; alcohol and drug is the reference group).

longest abstention from alcohol and drugs (in number of days between intake and follow-up) yielded similar results (not shown).

Effects of Treatment Modality by Psychiatric Severity

Before examining the effects of treatment modality on abstention among the four psychiatric subgroups, we examined variation in their psychiatric symptoms (from the ASI psychiatric items). The most prevalent symptoms among the two most severe subgroups (mid- and high-severity scores) were depression (81 percent and 98 percent, respectively) and anxiety (92 percent and 100 percent, respectively). The most severe subgroup was characterized by a high prevalence of violent behavior (58 percent), suicidal thoughts (67 percent), and suicide attempts (19 percent) in the 30 days prior to treatment intake. In contrast, the "scored zero" group reported none of these symptoms, and prevalence rates among the lowest one-third severity group were only 23 percent for depression, 50 percent for anxiety, 9 percent for violent behavior, 9 percent for suicidal thoughts, and 1 percent for suicide attempts. Similar and significant differences were found by psychiatric severity subgroup for SCL scores as well (somatization, obsession-compulsion, interpersonal sensitivity, depression, anxiety, hostility, and paranoid ideation). Correlations between the baseline psychiatric ASI composite and seven SCL scales that were also measured ranged from .43 (paranoid ideation) to .64 (depression).

We found no differences between DH and OP in baseline characteristics for those in the mid-range psychiatric severity subgroup. DH approached significance ($p = .06$) as a predictor of total abstention, even when controlling for type of dependence; DH was a significant predictor of alcohol abstention ($p = .019$) but not of drug abstention in this subgroup. DH was not associated with alcohol or drug abstention in any of the other psychiatric severity subgroups. TSR data showed that the two highest groups received the same amount of psychiatric services.

SELF-SELECTED SAMPLE

As with the randomized sample, both DH and OP showed significant improvement in alcohol and drug severity at the six-month follow-up. However, in contrast to the randomized sample, differences were significant between DH and OP for total (64 percent versus 47 percent, respectively; $p = .002$), alcohol (72 percent versus 61 percent, respectively; $p = .030$), and drug abstention (81 percent versus 72 percent, respectively; $p = .054$), as well as for treatment initiation (90 percent versus 73 percent, respectively; $p =$

.001). In addition, Table 3 shows that DH was a significant independent predictor of total and drug abstention and that it approached significance for alcohol abstention, even after controlling for motivation (i.e., a goal of abstinence), employer mandates, and other variables in the model on which the two samples differed. Goals to stop using any substance the patient had reported use of at intake were an important predictor of total (OR = 2.79, $p = .0003$) and alcohol abstention (OR = 4.04, $p = .0001$). We also examined an interaction of day treatment and the abstention goal measure; the interaction was not significant for any of the criteria and was dropped from the model. This suggests that abstention-motivated individuals did not have better outcomes in DH than in OP. This pattern of results was also replicated in the self-selected sample when we used "longest abstinence" from alcohol and drugs as the outcome criteria.

The self-selected sample was generally similar to the randomized sample regarding the prevalence of psychiatric symptoms by psychiatric severity group, as well as on the effects of DH by psychiatric severity group. Assignment to DH predicted total abstention among those in the middle one-third of the non-zero distribution. We found no significant effects for the other psychiatric status subgroups (not shown). The small sample size resulted in low power to detect significance for all but large effects. In the mid-level psychiatric severity group, 62 percent of DH subjects compared with 39 percent of OP subjects abstained from all substances ($p = .032$).

COSTS AND COST-EFFECTIVENESS

Treatment Costs

Table 4 shows that, for randomized subjects during the first three weeks subsequent to intake (excluding the intake visit), the average cost per DH subject was three times that of OP subjects ($747.57 versus $257.21). The greater intensity of DH during this phase is evident from the average number of group visits for DH subjects (34.5) compared to those in the OP program (6.5). In the five weeks following the intensive part of the program, the DH subjects' costs remained higher, but the intensity of group visits dropped significantly. DH subjects continued to receive a higher number of individual visits (1.33 compared to 0.87 for OP) and group visits (9.65 versus 5.32). For the entire eight-week treatment period, the average total cost was $1,144.53 for the DH group and $492.74 for the OP group. In the aftercare period, the average number of all types of visits and costs for the two groups was similar.

Table 3: Logistic Regression Models Predicting Six-Month Abstention: Self-Selected Sample

	Total Abstention			*Alcohol Abstention*			*Drug Abstention*		
	Odds Ratio	*95% CI*		*Odds Ratio*	*95% CI*		*Odds Ratio*	*95% CI*	
		Lower	*Upper*		*Lower*	*Upper*		*Lower*	*Upper*
Full self-selected sample† *(N)*	(335)*			(339)*			(335)*		
Day hospital	2.10	1.25	3.51	1.58	0.92	2.71	2.14	1.17	3.91
Midlevel psychiatric severity‡ *(N)*	(84)*			(86)*			(84)*		
Day hospital	2.93	1.16	7.41	1.54	0.57	4.14	2.31	0.84	6.34

*Sample sizes vary because logistic regression deletes cases with missing values on any variable in the model.

†Models for the full-sample control for baseline alcohol (for models of total and alcohol abstention) and drug (for models of total and drug abstention), ASI severity, psychiatric and legal ASI severity, ethnicity (white = 1, others = 0), employed status, employer mandates for treatment, and goals to stop substance use.

‡Models for the midlevel psychiatric severity group control for baseline alcohol ASI severity (for models of total and alcohol abstention) and drug ASI severity (for models of total and drug abstention), and dependence type (no dependence, alcohol only, drug only; alcohol and drug is the reference group).

Table 4: Utilization and Cost of DH and Traditional OP Programs, Randomized Sample ($N = 541$)

Cohort	*Intake to 3 Weeks*		*3 to 8 Weeks After Intake*		*8 Weeks to 6 Months, 8 Weeks After Intake*	
	% Patients with > 0 Visits	*Average Visits/Patient*	*% Patients with > 0 Visits*	*Average Visits/Patient*	*% Patients with > 0 Visits*	*Average Visits/Patient*
Day Treatment (N = *248)*						
Treatment Modality						
Psychiatric services	7.66%	0.12	8.87%	0.13	13.31%	0.32
Individual and medical (Detox) services	56.85%	1.24	53.23%	1.33	42.74%	1.92
Day treatment groups	65.73%	33.13	54.84%	3.71	9.27%	0.93
Other outpatient groups	75.81%	1.40	51.21%	5.94	42.74%	7.78
Any visit	87.10%	35.88	61.29%	11.11	47.98%	10.95
Average cost/patient		$747.57		$396.96		$495.50
Traditional Outpatient Treatment (N = *293)*						
Treatment Modality						
Psychiatric services	5.46%	0.06	6.48%	0.08	11.95%	0.30
Individual and medical (Detox) services	51.19%	0.91	41.64%	0.87	38.57%	1.61
Day treatment groups	10.24%	2.36	10.24%	0.58	7.51%	0.69
Other outpatient groups	78.50%	4.10	50.85%	4.74	39.25%	6.59
Any visit	84.30%		54.27%	6.27	43.69%	9.18
Average cost/patient		$257.21		$235.53		$403.91

In the self-selected sample (not shown), OP visit patterns and treatment costs were similar to those of the randomized OP group. However, subjects selecting DH had a higher number of visits and costs than subjects randomized to DH. Costs by treatment phase were $1,195, $578, and $737 for the initial three weeks, three to eight weeks, and eight-week to six-month periods, respectively. During all phases of the treatment, the self-selected DH subjects were higher users of services than were all other OP and DH groups.

Cost-Effectiveness

We examined the C/E ratios for those outcome measures where DH had a significant effect. These were (a) abstention from alcohol for the randomized, mid-level psychiatric severity group; (b) total abstinence for the self-selected sample; (c) abstention from drugs for the self-selected sample; and (d) total abstinence for the self-selected mid-level psychiatric severity group. The incremental C/E ratios were $5,464, $9,576, $23,721, and $4,629, respectively.

DISCUSSION

Differences existed between the randomized DH and OP patients, particularly with regard to drug severity and employment. On controlling for these variables, no main-effect differences were found in abstention outcomes for DH versus OP programs in the randomized sample. However, treatment effect appeared to differ by level of psychiatric severity. We found no differences in the highest or lowest severity groups. But, consistent with research that found a threshold effect of psychiatric severity associated with outcome (Gottheil, McLellan, and Druley 1992; McLellan, Luborsky, Woody, et al. 1983), randomized subjects with midlevel psychiatric severity (persons who scored roughly between .4 and .6 on the ASI psychiatric composite score) benefited more for alcohol (and to a lesser degree, total) abstention from DH, even when we controlled for type of dependence.

It is likely that the highest severity group, which had higher levels of violent and suicidal behavior, required other psychiatric services not provided by either program. The mid-level group benefited most from the intensity offered in DH compared to that in OP. This finding, if replicated in other populations, suggests that there may be clinical benefit to assigning mid-level psychiatric severity patients to more intensive treatments. It may also mean that the DH program provided the appropriate match of psychiatric services for that group.

We found strikingly different outcomes between the programs in the self-selected sample. Subjects in the DH program had higher abstinence rates, even when we controlled both for characteristics on which the DH and OP samples differed and for possible selection factors such as patient motivation and employer-mandated treatment. Importantly, analysis by psychiatric status among the self-selected sample again found effects only among the mid-level of psychiatric severity, where DH subjects had almost three times the likelihood of total abstention as those in OP.

Issues of Randomized Versus Self-Selected Samples

The DH program yielded better outcomes than did the OP program among subjects who self-selected treatment, but not among subjects who were randomly assigned to treatment. This is an important finding, and its examination was made possible by the rare opportunity to compare the two samples in detail. We do not know all of the factors behind the unwillingness of some subjects to be randomized, but they included clinical judgment, employer mandates, and unknown patient selection factors. Thus, to account for the differences in outcome, we examined differences between the DH self-selected subjects and the other three study groups. A higher proportion of the DH self-selected subjects than of the other three groups actually started treatment. However, the OP self-selected group had more characteristics consistently associated in the literature with a good outcome (Monahan and Finney 1996). They had higher levels of employment and lower levels of severity of drug, employment, legal, and psychiatric problems. At the same time, higher proportions of self-selected DH than OP subjects had received employer mandates—a factor related only to starting treatment, not to outcome. DH subjects in both samples also reported more motivation to abstain completely from all substance use and thus had a better fit with program philosophy when they began treatment. Although we could hypothesize that the patients who self-selected OP were less convinced of the severity of their addiction problem, or of the need to quit using, when we included these measures as well as employment status in the model, the DH effect was still apparent. This suggests that the better outcome of the self-selected DH patients is not completely explained by differences in patient factors, such as motivation. The significant effect that remains may be a result of unobserved differences and not a true benefit. This finding is important in our efforts to statistically control for differences in naturalistic studies.

It seems clear that in the case of this study, the randomized sample alone does not allow for a clear understanding of the relationship between patient

effects—for instance, the effects of motivation or employer pressure—and treatment effects as they affect outcome. Randomization responds to those issues when there is no bias in the willingness of the patient to be randomized and when the treatment effects are relatively uniform over broad classes of individuals. When treatment effects depend on a person's commitment to behavioral change and compliance, and on a fit between services and needs—and when many of the most motivated subjects are unwilling to be randomized—this may lead to an underestimation of the program's benefit among those randomized. Willingness to participate is most crucial in trials where substantially different levels of patient compliance are required. Nonparticipation in the trial reported here may have self-selected out those individuals for whom the clinical applicability of intensive treatment might have been strongest. We found motivation to be predictive of outcome in both the randomized and the self-selected samples. Clinicians' judgment may also have reduced the numbers of clients with good fit in the randomized sample. These findings argue for ensuring high recruitment rates to randomized studies and for examining both randomized and self-selected patients in the same studies, particularly when differences in compliance and requirements between programs are large. Historically, most studies that have examined the topic of substance abuse intensity have been clinical trials, and they often have not provided data on randomization rates. Therefore, we cannot compare our rate to prior studies. We designed our study to include persons who self-selected (the usual form of treatment selection); this reflects a health services research perspective concerned with external as well as internal generalizability. Further, prior studies have not provided comparisons of their treatment groups on such a broad range of characteristics, particularly on those variables that reflect motivation, where we found differences between the samples. The findings suggest that we cannot assume perfect randomization in health services trials set in the "real world," and that more data on demographics, severity levels, and motivation should be published along with outcome results.

Cost-Effectiveness

DH subjects used more program services than did those in OP, both in the randomized sample (4.8 times the number of services) and the self-selected sample (6.0 times) during the initial treatment phase. However, the cost differences were smaller because the visits were mostly group visits, which are less costly. DH patients also utilized more psychiatric and medical services in the program. Intensity differences lessened over time (a reflection of program

structure); however, the self-selected sample continued to use more services even up to six months after the initial intensive phase of the DH program. We found, for the randomized, mid-level psychiatric severity subgroup, an expenditure of $5,464 would result in alcohol abstention by one additional person in the DH program compared to those in the OP program. (Our analysis is limited to direct treatment costs of the provider. Opportunity costs of treatment such as wages lost from time off work, or other direct non-medical costs such as transportation and child care, are not included. These are likely to be different between patients in the two programs.)

Although the study's different results for randomized and self-selected patients are inconclusive in terms of the effectiveness of treatment intensity for the overall treatment population, they add evidence to the important role of psychiatric severity and treatment intensity. They also suggest the importance of replicating earlier clinical trials on intensity through the use of studies that include self-selected patients or that have population-based samples with very high randomization rates. Because many health services studies do not have the option of randomization, it was important that we contrast a randomized design with a treatment selection-as-usual design to demonstrate the differences that might be hidden in naturalistic studies. This study cannot provide an answer regarding whether the most motivated patients would have done equally well in the lower-intensity program or whether they did better in the day hospital program because they found the appropriate service match. It is important to conduct methodological work in this area.

ACKNOWLEDGMENTS

We wish to thank A. Thomas McLellan, Ph.D., Gary Friedman, M.D., James Wiley, Ph.D., John Finney, Ph.D., Bruce Fireman, M. A., and James McKay, Ph.D. for their valuable comments, and Agatha Hinman for editorial assistance on the final manuscript.

REFERENCES

Alterman, A. I., C. P. O'Brien, A. J. McLellan, D. S. August, E. C. Snider, M. Droba, J. W. Cornish, C. P. Hall, A. H. Raphaelson, and F. X. Schrade. 1994. "Effectiveness and Costs of Inpatient Versus Day Hospital Cocaine Rehabilitation." *The Journal of Nervous and Mental Disease* 182: 157–63.

American Psychiatric Association. 1994. *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. 4th ed.* Washington, DC: American Psychiatric Association.

Carroll, K. M. 1995. "Enhancing Retention in Clinical Trials of Psychosocial Treatments: Practical Strategies." In NIDA Research Monograph, no. 13. Center for Drug Treatment and Research. Washington, DC: Koba Institute, Inc.

Finney, J. W., and S. E. Monahan. 1996. "The Cost-Effectiveness of Treatment for Alcoholism: A Second Approximation." *Journal of Studies on Alcohol* 57 (3): 229–43.

French, M. T., and K. A. McGeary. 1997. "Estimating the Economic Cost of Substance Abuse Treatment." *Health Economics* 6 (5): 539–44.

Fuller, R. K. 1990. "Controlled Clinical Trials." *Alcohol, Health, and Research World* 14 (3): 239–44.

Gerstein, D. A., and H. J. Harwood, eds. 1990. *A Study of the Evolution, Effectiveness, and Financing of Public and Private Drug Treatment Systems.* Vol. 1 of *Treating Drug Problems.* Washington, DC: National Academy Press.

Gottheil, E., A. T. McLellan, and K. A. Druley. 1992. "Length of Stay, Patient Severity and Treatment Outcome: Sample Data from the Field of Alcoholism." *Journal of Studies on Alcohol* 53 (1): 69–75.

Harrison, L. D. 1995. "The Validity of Self-Reported Data on Drug Use." *The Journal of Drug Issues* 25 (1): 91–111.

Institute of Medicine. 1990. *Broadening the Base of Treatment for Alcohol Problems.* Washington, DC: National Academy Press.

Levin, T. R., J. A. Schmittdiel, K. Kunz, J. M. Henning, C. J. Henke, C. J. Colby, and J. V. Selby. 1997. "Costs of Acid-related Disorders to a Health Maintenance Organization." *American Journal of Medicine* 103 (6): 520–28.

Lipman, R. S., L. Covi, and A. K. Shapiro. 1979. "The Hopkins Symptom Checklist: Factors Derived from the HSCL-90." *Journal of Affective Disorders* 1 (1): 9–24.

Mattson, M. E., and J. P. Allen. 1991. "Research on Matching Alcoholic Patients to Treatments: Findings, Issues and Implications." *Addictive Diseases* 2 (2): 33–49.

McLellan, A. T., A. I. Alterman, G. E. Woody, and D. Metzger. 1992. "A Quantitative Measure of Substance Abuse Treatments: The Treatment Services Review." *Journal of Nervous and Mental Disease* 180 (1): 100–109.

McLellan, A. T., M. Belding, J. R. McKay, D. Zanis, and A. I. Alterman. 1997. "Can the Outcomes Research Literature Inform the Search for Quality Indicators in Substance Abuse Treatment?" Appendix B in *Managing Managed Care: Quality Improvement in Behavioral Health.* Washington, DC: National Academy Press.

McLellan, A. T., L. Luborsky, G. E. Woody, K. A. Druley, and C. P. O'Brien. 1983. "Predicting Response to Alcohol and Drug Abuse Treatments: Role of Psychiatric Severity." *Archives of General Psychiatry* 40 (6): 620–25.

McLellan, T. A., H. Kushner, D. Metzger, R. Peters, I. Smith, G. Grissom, H. Pettinati, and M. Argeriou. 1992. "The Fifth Edition of the Addiction Severity Index." *Journal of Substance Abuse Treatment* 9 (3): 199–213.

Monahan, S. C., and J. W. Finney. 1996. "Explaining Abstinence Rates Following Treatment for Alcohol Abuse: A Quantitative Synthesis of Patient, Research Design and Treatment Effects." *Addiction* 91 (6): 787–805.

Project MATCH Research Group. 1993. "Project MATCH: Rationale and Methods

for a Multisite Clinical Trial Matching Patients to Alcoholism Treatment." *Alcoholism: Clinical and Experimental Research* 17 (6): 1130–45.

Ross, H. E., F. B. Glasser, and T. Germanson. 1988. "The Prevalence of Psychiatric Disorders in Patients with Alcohol and Other Drug Problems." *Archives of General Psychiatry* 45 (November): 1023–31.

Rounsaville, B. J., Z. S. Dolinsky, T. F. Babor, and R. E. Meyer. 1987. "Psychopathology as a Predictor of Treatment Outcome in Alcoholics." *Archives of General Psychiatry* 44 (June): 505–13.

Schmidt, L., and C. Weisner. 1993. "Developments in Alcoholism Treatment: A Ten-Year Review." In *Recent Developments in Alcoholism,* edited by M. Galanter, ch. 11, pp. 369–96. New York: Plenum.

Selby, J. V. 1997. "Linking Automated Databases for Research in Managed Care Settings." *Annals of Internal Medicine* 127 (8): 719–24.

Strohmetz, D. B., A. I. Alterman, and D. Walter. 1990. "Subject Selection Bias in Alcoholics Volunteering for a Treatment Study." *Alcoholism: Clinical and Experimental Research* 14 (5): 736–38.

U.S. National Institute on Alcohol Abuse and Alcoholism. 1998. *Improving the Delivery of Alcohol Treatment and Prevention Services: A National Plan for Alcohol Health Services Research.* Pub. No. 97-4223. Rockville, MD: National Advisory Council on Alcohol Abuse and Alcoholism, Subcommittee on Health Services Research.

Weisner, C., A. McLellan, and E. Hunkeler. In press. "Addiction Severity Index Data from the General Membership and Treatment Samples of HMO Members: One Case of Norming the ASI." *Journal of Substance Abuse Treatment.*

Zanis, D. A., A. T. McLellan, and M. Randall. 1994. "Can You Trust Patient Self-Reports of Drug Use During Treatment?" *Drug and Alcohol Dependence* 35 (2): 127–32.