Published in final edited form as:

Drug Alcohol Depend. 2009 September 1; 104(1-2): 56–64. doi:10.1016/j.drugalcdep.2009.03.018.

USING PROPENSITY SCORES TO ADJUST FOR SELECTION BIAS WHEN ASSESSING THE EFFECTIVENESS OF ALCOHOLICS ANONYMOUS IN OBSERVATIONAL STUDIES

Yu Ye¹ and Lee Ann Kaskutas^{1,2}

¹ Alcohol Research Group, Public Health Institute

² School of Public Health, University of California Berkeley

Abstract

Background—The effectiveness of Alcoholics Anonymous (AA) is difficult to establish. Observational studies consistently find strong dose-response relationships between AA meeting attendance and abstinence, and the only experimental studies favoring AA have been of 12-step facilitation treatment rather than of AA per se. Pending future randomized trials, this paper uses propensity score (PS) method to address the selection bias that potentially confounds the effect of AA in observational studies.

Method—The study followed a treatment sample 1 year to assess post-treatment AA attendance and abstinence (n=569). Propensity scores were constructed based on known confounders including motivation, problem severity, and prior help-seeking. AA attendance during the 12-month follow-up period was studied as a predictor of alcohol abstinence for the 30 days prior to the follow-up interview. PS stratification and PS matching techniques were used to adjust for the self-select bias associated with respondents' propensity to attend AA.

Results—The overall advantage in abstinence initially observed narrowed when adjusted. The odds ratio associated with AA attendance reduced from 3.6 to 3.0 after PS stratification and 2.6 after PS matching to AA attenders. Support for AA effectiveness was strengthened in the quintile with lower propensity scores and when AA non-attenders were matched as the target group, but was weakened among those in the higher PS quintiles and when matching to AA attenders.

Discussion—These results confirm the robustness of AA effectiveness overall, because the results for higher abstinence associated with AA attendance following propensity score adjustment remained significant, and the reduction in the magnitude of AA's effect was moderate. However, the effect modification by propensity scores in both PS stratification and PS matching approaches seems to suggest that AA may be most helpful, or matter more, for those with a lower propensity to attend AA. Conversely, for those with a high propensity to go to AA (operationalized as higher motivation, greater problem severity, more prior AA and treatment exposure, etc.), attending AA may not make as much of a difference. It will be important that future studies replicate our results, as this is the first paper to use propensity score adjustment in this context.

Corresponding author: Mr. Yu Ye, Alcohol Research Group, Public Health Institute, 6475 Christie Avenue, Suite 400, Emeryville, CA 94608, E-mail: yye@arg.org, (510)597-3440, Fax (510)985-6459.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Alcoholics Anonymous; Propensity score; Effectiveness; Selection bias

1 Introduction

Although Alcoholics Anonymous (AA) is the most widely-sought source of help for addressing severe alcohol problems (Room and Greenfield, 1993; Schmidt et al., 2007; Weisner et al., 1995), the effectiveness of AA is difficult to establish. Observational studies consistently find strong dose-response relationships between AA meeting attendance and abstinence, but AA involvement in such studies is subject to selection bias, calling into question the specificity of the AA effect (Kaskutas, 2007, in press). That is, by self-selecting to attend AA or not, drinkers sort themselves into groups which may differ on the underlying factors which drove their decision. For example, more motivated individuals may “self-select” to attend AA; and this higher motivation, rather than their AA attendance, could actually account for their subsequent abstinence.

To remove the threat to validity that arises from such self-selection, experimental studies are needed, but results from the few that have been conducted have been mixed. None of the studies that have randomized subjects to Alcoholics Anonymous (Brandsma et al., 1980; Ditman et al., 1967; Walsh et al., 1991) have found positive effects for the AA condition, whereas randomized studies of 12-step facilitation (TSF, a treatment intended to help patients become engaged in AA) have reported higher rates of abstinence in the TSF condition at 1 year (Project MATCH Research Group, 1997; Timko and Debenedetti, 2007) and 3 years (Project MATCH Research Group, 1998). These mixed results may be due to the very concept of randomization to AA, which is a voluntary organization, freely available to anyone with a desire to stop drinking; individuals can still “self-select” to attend or not attend AA, regardless of the condition to which they may have been randomized.

Although randomized control trials are often considered the gold standard research study design, for research on AA (and, some have argued, for evidence-based practice more broadly (Tucker and Roth, 2006)) they are limited. However, advanced statistical techniques such as instrumental variables, structural equation modeling, and propensity scores are available to confront some of the biases inherent in observational studies of AA. This paper uses propensity scores to assess AA effectiveness before and after adjusting for one’s propensity to go to AA.

1.1 Methods of Statistical Control

In the absence of randomization, control for self-selection can be addressed via statistical methods. While our study is the first to our knowledge to apply the propensity score technique to the study of AA, other statistical approaches have been applied to the question of AA effectiveness. One technique, from econometrics, uses instrumental variables as a way of adjusting for the likelihood of self-selecting into AA. Instrumental variables are measures that are highly associated with the exposure of interest, but are not related to the outcome. For example, in studies of AA’s effectiveness on abstinence, instrumental variables must be highly associated with AA attendance, but have no direct relationship with abstinence—a requirement often difficult to meet.

Results using instrumental variables have also been mixed. Fortney et. al. (1998) used two instrumental variables, ability to drive oneself to AA meetings and the presence of an AA meeting in the town of residence, and found that the effect of AA on abstinence was greatly reduced, from an Odds Ratio (OR) of 3.7 (using standard logistic regression) to an OR of 1.7 with correction for these two instrumental variables. Using different instrumental variables

(perceived seriousness of drinking and use of information-seeking as a coping method), Humphreys et. al. (1996) found that AA was significantly associated with less severe drinking, with a stronger effect observed post-adjustment.

The instrumental variable method is believed to adjust for two sources of selection bias, the first caused by omitted variables (the “omitted variable” problem) and the second arising from reverse causation (or simultaneity, see (Wooldridge, 2002, p.50). The omitted variable problem occurs because individuals *select* to seek the help or treatment being evaluated (in our case, AA) due to underlying but unobserved factors which are correlated with help seeking. Reverse causation happens when treatment (or help) is selected as the result of the outcome, rather than the outcome being due to the treatment, commonly caused by the problem of temporality (in that the outcome occurs before the treatment). By using instrumental measures correlated with treatment seeking/AA but not *independently* correlated with the outcome (abstinence, for instance), the instrumental variable method addresses both of these types of selection bias. The primary challenge with the technique is the choice of appropriate instrumental variables; frequently, the instrumental variables do not explain much variation in the explanatory variable (treatment/AA) (Bound et al., 1995).

A study using another statistical approach, structural equation modeling, which attempts to address the directionality of influences, showed that prior AA involvement predicted subsequent lower drinking problems whereas prior drinking problems were not predictive of subsequent AA participation (McKellar et al., 2003). While addressing the problem of reverse causation, structural equation modeling can not fully adjust for the selection bias caused by not having potential confounders (the omitted variable problem).

1.2 Propensity Scoring

The propensity score (PS) method is an alternative to the above approaches, by allowing investigators to consider all known and measured potential confounders without compromising statistical power. Proposed initially by Rosenbaum and Rubin, the PS method either statistically stratifies or matches individuals by their propensity to do something, in our case, to go to AA. This propensity is operationalized as a “propensity score.” In their 1983 seminal paper (Rosenbaum and Rubin, 1983), Rosenbaum and Rubin showed that the scalar propensity score is sufficient to remove bias from all observed covariates. They then demonstrated bias reduction and adjusted effect estimations, using approaches of both PS stratification (Rosenbaum and Rubin, 1984) and PS matching (Rosenbaum and Rubin, 1985). Following their work, many methodological papers and review articles have been published giving more detailed information on model construction, variable selection, and robustness of estimation (e.g. (Brookhart et al., 2006; D’Agostino Jr., 1998; Dehejia and Wahba, 2002)).

There are several advantages using the PS method. The degrees of freedom required when entering numerous predictor variables in a regression equation, for instance, can result in Type II error (failing to detect a true difference) unless the sample size is quite large. In contrast, a great number of potential confounders can be included in the construction of the propensity score, because the variables are not being used as individual independent variables; they are only used to construct the score. In addition, the PS approach allows one to examine whether the treatment group and the untreated group are fully balanced in terms of all observed potential confounders. For our purposes, the approach is aimed to balance the AA-attender and AA-nonattender groups as treatment conditions similar to what is accomplished by randomization in randomized controlled trials. Propensity scores take into account all known observed potential confounders.

Of course, PS analysis cannot do anything about unknown unobserved (unmeasured) confounders, which can only be balanced by randomization. However, when randomization

itself introduces bias, propensity scores represent a good option for addressing the omitted variable problem that arises from not controlling for all confounders when attempting to establish causation. As noted above, randomization to AA is problematic because AA is freely available in the community and it is not possible to forbid study subjects to attend AA; this introduces its own selection bias.

The capability of the propensity score method for addressing the selection bias problem depends vitally on the extent of the available covariate measures used to capture potential confounders. If key confounders are omitted in the construction of propensity scores, the selection bias arising from omitted variables fails to be adjusted. The confounders of concern are those that are strongly related both to AA attendance and to outcome.

Our analysis here is enhanced by a rich set of potential confounders between AA attendance and abstinence. These include self-motivation (operationalized here as readiness to change) (Isenhardt, 1997; Kaskutas et al., 2002) and coercion by others (Ammon et al., 2008; George and Tucker, 1996; Weisner et al., 2003; Weisner and Matzger, 2002), which are expected to affect study participants' decisions to go to AA and *independently* contribute to their becoming abstinence. A different type of example is problem severity, which is a strong predictor of AA attendance in almost any study of AA affiliation (for a review, see (Bogenschutz, 2008; Emrick et al., 1993); also see (Humphreys et al., 1991; Kaskutas et al., 2002; Morgenstern et al., 2003); and see (Tucker and Gladysjo, 1993) for an exception). Yet individuals with greater alcohol problem severity may be less likely to quit drinking (Kaskutas et al., 2002) or to cut-down significantly (Matzger et al., 2004), regardless of going to AA or not. Failing to adjust for motivation would lead to overestimation of AA effectiveness, while not adjusting for problem severity could underestimate the effect.

Other confounders, which to varying extents might affect AA-going and abstinence, include help-seeking experiences, such as having prior alcohol treatment (Humphreys et al., 1998; Kaskutas et al., 2008), the type of setting in which treatment was received (for example, clinically-oriented, hospital setting, community setting, etc.; see (Barrows, 1998; Borkman et al., 2007; Kaskutas et al., 2004)), and prior experiences with AA (Timko et al., 2006a), Social influences also could act as confounders, in either direction, as with positive social support (Kaskutas et al., 2002) or heavy drinking networks (Kaskutas et al., 2002).

In addition to these established confounders, several demographic variables have been found to predict AA involvement, outcome, or both. Although the direction of influence is not consistent, and their effect is not always significant, likely candidates include female gender and married status, as well as age, education, ethnicity, and SES (Ammon et al., 2008; Bogenschutz, 2008; Dawson et al., 2005; Del Boca and Mattson, 2001; Emrick et al., 1993; Kaskutas et al., 2008; Timko, 2008; Timko et al., 2006a; Timko et al., 2002; Tonigan et al., 2002; Witbrodt et al., 2007).

In the current study, we utilized a prospective cohort study which recruited participants seeking substance abuse treatment who were followed-up a year post-treatment entry. We employ two propensity score methods (stratification and matching) to evaluate AA's effectiveness on alcohol abstinence. The propensity scores will adjust for potential selection bias by including all potential, measured confounders of AA attendance and alcohol abstinence. AA exposure is evaluated by any AA attendance for the 12-month period preceding the 1-year follow-up interview, and the outcome is abstinence for the last 30 days prior to the 1-year follow-up. Although AA attendance could happen within the last 30 days, the short period of overlap minimizes the problem of temporality and helps adjust for potential selection bias caused by "reverse causation".

2 Method

2.1 Sample

Baseline data used here are drawn from clients in 10 large specialty alcohol treatment programs in a northern California county. Inpatient and outpatient programs with fewer than five new clients per week were excluded, as were DUI programs and programs that are primarily aftercare in nature. The study sites represent clients with diverse insurance and funding resources, including public, private and health maintenance organizations (HMOs), with approximately equal numbers of clients recruited from each funding pool. In-person interviews were conducted within the first 3 days of treatment at the inpatient sites, or within the first 3 sessions at the outpatient ones. The baseline data collection, performed in 1995 and 1996, resulted in 927 completed interviews, with a response rate of 80%. For more information on the sampling frame, please see (Kaskutas et al., 1999). One-year follow-up interviews were conducted by telephone (or in person when unreachable by telephone) with 78% of the baseline sample ($n=722$).

Of these, 153 individuals were excluded from the analysis (68 current abstainers who had not drank alcohol in the year prior to baseline, 20 individuals who were missing data for AA exposure, and 65 participants who were missing data on variables required to construct the propensity score), resulting in an available sample size of $n=569$. Sample demographics are shown in table 1 by AA attendance status.

2.2 Measures

Outcome—The outcome variable used here is a dichotomous variable indicating whether the respondent drank any alcohol in the 30 days before the 1-year follow-up interview.

Exposure—AA attendance at the 1-year follow-up based upon the previous 12-month meeting attendance.

Potential confounders—We were interested in six categories of 21 potential confounders: Self motivation and external coercion, alcohol problem severity, help-seeking, social influences, and demographic characteristics. All confounders were baseline measures.

Motivation to change was operationalized through the readiness to change summary index of 12 items from Prochaska's stage of change scale (range 12–60, Cronbach's $\alpha=0.87$) (Prochaska and DiClemente, 1984). The 12 items of the summary index represent all four subscales of the stage of change construct (precontemplation, contemplation, action and maintenance). These have been used in other publications from this study (Kaskutas et al., 2002) Five response categories were offered, ranging from “strongly agree” to “strongly disagree.”

Coercion was based on two indices respectively measuring the number of sources of pressure who suggested treatment (range 0–7), and who gave an ultimatum to seek treatment (range 0–7). Sources included family, friends, doctors, co-workers, clergy, legal and social system contacts.

Baseline problem severity was assessed using three summative measures: the Addiction Severity Index (ASI) composite score (range 0–1) (McLellan et al., 1985) representing past 30-day alcohol problem severity; the number of dependence symptoms (based on nine items, such as got drunk when should not, blacked-out, had eye-opener, had shakes; range 0–9, $\alpha=0.84$); and the number of alcohol-related consequences (based on eight items, e.g., being arrested when drinking, having an accident or close call when drinking; range 0–8, $\alpha=0.55$) experienced in the past 12 months.

Three measures of help-seeking were used to capture formal and informal service utilization. At baseline, we assessed the number of AA meetings in the past year, the number of specialty treatment episodes attended in the past year, and the type of program at which the study participant was seeking treatment when recruited for the study: a publicly-funded program in the community, a privately-funded hospital-based program, or an outpatient clinic in a Health Maintenance Organization (HMO). In contrast to the other program types, publicly-funded programs rely mainly on the 12-step community and 12-step philosophy as the key therapeutic ingredients (see (Borkman et al., 2007; Kaskutas, 1998)) and may be equally effective (Witbrodt et al., 2007), so this measure is an especially good candidate for potential confounding of AA effectiveness. At follow-up, the key variable, "AA attendance," was based on whether the respondent had attended an AA meeting in the past 12 months or not.

Two types of social influences were considered: the size of the social support network, and its drinking-related characteristics. To capture network size, we asked about the number of people you have available "to talk to when you are worried about personal problems," the number who have "helped you with practical things when you needed it," and the number of family members and friends who "you have regular contact with" (all range 0–30). The drinking influences in the social network focused on "problem drinkers" and individuals who "encourage you to drink or use drugs." These variables were entered both as counts and as the percentages of the group of regular contacts (counts range from 0–20 and 0–30 respectively)

Finally, five demographic variables were considered as potential predictors of AA attendance: gender, age, level of education (range 1–6), ethnicity (white, black, other), and marital status (married, separated/divorced/widowed, single).

2.3 Analysis

The propensity score analysis involved three steps: (1) constructing the propensity scores; (2) balancing the sample based on their propensity scores; and (3) estimating the effect of AA on abstinence with and without the propensity scores. In step 1, the propensity scores were estimated using multivariate logistic regression models predicting AA exposure during the 1-year follow-up period. First, 21 potential baseline predictor variables (described above, and listed in Table 1) were entered as fixed main effect independent variables. A forward selection approach was then used in the PS estimation, with quadratic terms for the continuous variables as well as all interaction terms between the 21 potential confounders entered into the model sequentially, and with those quadratic and interaction terms significant at $p < 0.05$ retained. The propensity scores, the predicted probabilities of attending AA, are calculated for each individual using the person's values for each predictor variable and the respective coefficient from the model. The final estimated propensity scores are the conditional probability of AA attendance estimated from a vector of the 21 observed independent predictor variables plus one quadratic and eight interaction terms.

Step 2 employed two techniques to balance the sample: PS stratification and PS matching. With PS stratification, the predicted probabilities from the fitted logistic regression were sorted for the whole sample by increasing magnitude, and divided into five equal size groups (quintiles), such that the predicted probability of AA attendance was lowest in the first quintile and highest in the 5th quintile. Individuals in each quintile thus are (after appropriate testing, described below) considered homogenous or balanced in terms of the potential confounders of AA effectiveness between the AA-attender and AA-nonattender groups.

Using the PS matching approach, the logit form of the estimated probability was used to match between the treated and untreated groups, following suggestions in the literature (D'Agostino, 1998; Rosenbaum and Rubin, 1985). The STATA user-written command PSMATCH2 (Leuven and Sianesi, 2003) was used to locate the nearest available match, with the maximum

distance (referred to as a caliper) limited to 0.1. This caliper width used here, equivalent to 7% of the standard deviation of the logit PS for the treated and untreated groups combined, is much less than the 20% recommended by Rosenbaum and Rubin (1985) and thus is expected to remove even more bias (because we are matching to individuals with a more similar propensity to attend AA).

Two different matching analyses were performed: (1) matching the AA-attender group using the AA-nonattender group, and (2) matching the AA-nonattender group using the AA-attender group. For the first matching, 282 AA-attender study participants were matched by 102 AA-nonattenders, with those in the AA-nonattender group used one or more times to achieve the best match. Similarly for the second matching, 209 subjects in AA-nonattender group were matched by 104 AA-attenders. All cases without matches were discarded from subsequent analysis of the matched PS sample.

Before the PS-stratified and the PS-matched samples were used to evaluate the effect of AA on abstinence, we assessed whether the covariates used in the PS construction were fully balanced under PS stratification and under PS matching. Two-way ANOVAs were fitted to evaluate the stratification approach, with the dichotomous AA-exposure variable (AA exposed vs. not) and quintile membership entered as predictors of each propensity score variable separately. The F statistics from the AA-attender variables were then used to assess if there were significant differences between the AA-attender and AA-nonattender groups. For the PS matching approach, the differences between the matched AA-attender and AA-nonattender groups for each propensity score variable were evaluated using both t-tests and the standardized difference. The standardized difference is the percentage of the absolute difference in sample means divided by an estimate of the pooled standard deviation. Since t-tests before and after PS matching cannot be compared directly due to change in sample size and inflated standard errors (because of multiple matchings), standardized differences have been recommended as a better measure in assessing balance (Austin, 2008).

In step 3, once the covariates were shown to be fully balanced between the AA-attender and AA-nonattender groups, a simple difference in means was used to estimate the AA effect (step 3). For the stratification approach, the differences were first derived for each quintile, and the average difference across the quintiles was considered the overall adjusted effect. For the matching approach, the differences in means were calculated with both the AA-attender matched and AA-nonattender matched scenarios, from STATA PSMATCH2 module (Leuven and Sianesi, 2003).

3 Results

3.1 Propensity Score Stratification/Matching

Table 1 lists all potential confounder variables used to construct the propensity score (step 1), and shows the effectiveness of PS stratification and PS matching in balancing the attender and nonattender groups (step 2). The first part of the table (left-hand side columns) presents comparisons between the AA-attender and AA-nonattender groups at the 1-year follow-up before considering the PS method. For example, the proportion of males in the AA-attender group was 59% and the proportion of the AA-nonattender group that was male was 56%. The mean ages were 38.8 years and 36.8 years respectively.

The fourth column reports the F-statistics comparing the proportions or means for each potential confounder variable between the AA-attender and AA-nonattender groups. As shown by the before-stratification F-statistics, several large differences were observed between the AA-attender and nonattender groups prior to PS stratification, particularly for baseline variables such as readiness to change, pressure from others to get treatment, problem severity,

prior AA attendance and type of treatment, and, in both the magnitude of the standardized differences as well as the F statistics and resultant significance tests from one-way ANOVAs.

To determine whether the PS stratification fully balanced the AA and non-AA groups, a two-way ANOVA was performed on each covariate with the PS quintile added as a control factor. This is the F-statistic reported in the fifth column (PS stratification F-statistics). In all cases, the F statistics were greatly reduced by PS stratification, with no significant differences observed post-stratification. For example, an important potential confounder of AA's effect on abstinence is motivation, with the average scores on the readiness to change index at baseline significantly different for the AA-attender and nonattender groups (50.0 and 46.6 respectively, $p < .001$). After PS stratification, motivation did not significantly vary between the AA-attender and nonattender groups ($F = 0.08$).

Similarly, those in the AA-attender group reported an average of 5.2 dependence symptoms at baseline, compared to only 3.69 symptoms for the AA-nonattender group. This difference was significant at $p < .001$, with an F-statistic of 42.5. However, when a two-way ANOVA was conducted that compared the number of dependence symptoms as a function of AA attendance status and of PS stratification quintile, the F-statistic was only 0.18 and was no longer significant. Likewise, consider the large difference in number of AA meetings attended in the 12-months prior to baseline by AA attendance at the 1-year follow-up: 36.6 meetings were attended during the year prior to baseline in the group that had attended AA during the 12-months prior to follow-up, versus 8.08 meetings in the group that had not been exposed to AA during the follow-up period. This difference in number of meetings was highly significant prior to stratification ($F = 42.3$) but stratification seems to have effectively controlled this potential confounder ($F = 0.32$, not significant). Another

Next, we consider the effect of PS matching on the potential confounders. When the AA-attender group at follow-up was matched using the AA-nonattender group (shown at the right-hand columns of table 1), PS matching performed well in balancing the potential confounders, with most standardized differences between the attender and non-attender groups reduced to below 10% and no significant differences observed. For example, the standardized difference in dependence symptoms between the AA attender and nonattender groups was reduced from 56% (before PS matching) to 9.8% (after PS matching), more than an 80% reduction in bias. Likewise, the difference in the mean number of AA meetings attended in the 12 months before baseline went from 59% between those who did and did not attend AA at follow-up, to 8.2% after matching. However, three covariates (age, being single, and readiness to change) still had standardized differences larger than 10% (but less than 15%) after matching, and one of the coercion measures (counts of sources suggesting treatment) had a standardized difference as large as 16.8% between the attender and non-attender groups after matching

When the AA-nonattender group was matched using the AA-attender group, most of the covariates had standardized differences less than 10% and no significant differences emerged (results not shown). However, five covariates still had standardized differences larger than 10% (but less than 15%) after matching, and two other covariates had standardized differences larger than 15% (ASI severity 16%, and readiness to change 17%).

While no fixed cutpoint has been defined for the achievement of fully balancing the groups, the relatively high standardized differences that remained after PS matching for some confounders suggest that the effect of these variables should be further adjusted when estimating the effect of AA on abstinence. Thus, we will present conditional effects as well as unconditional effects. Two conditional effects models were considered, using generalized estimating equations (GEE) that adjusted for the respective sets of covariates whose standardized differences remained above 15% and 10% after PS matching

3.2 AA Attender versus Non-attender

After examining the extent to which the PS stratification and PS matching approaches balanced the potential confounders between the AA-attender and AA-nonattender groups, the AA effect before and after the PS method was studied (step 3; Table 2). Before any PS adjustment, the difference in the rates of past 30-day abstinence prior at the 1-year follow-up between the AA-attender and AA-nonattender groups was 30.8% (i.e., 68.6% were abstinent in the AA-attender group and 37.7% were abstinent in the AA-nonattender group), representing an odds ratio (OR) of 3.6. Using the PS stratification method, the overall adjusted difference between AA-attenders and AA-nonattenders was 26.5% (i.e., .696 - .431, see “adjusted” row in the “Propensity Score Stratification” section of the table); this is equivalent to an OR of 3.0. For each quintile, the differences in the proportion who were abstinent in the AA-attender versus the AA-nonattender groups were 37.7%, 37.7%, 27.4%, 17.6% and 11.8%, respectively, from the lowest to highest propensity score quintile.

Using the PS matching approach (bottom section of Table 2), when the AA-attender group was matched using the AA-nonattender group, the difference in abstinence rates was 19% between the AA-attender and AA-nonattender groups (i.e., 67% minus 48%). In this scenario, the overall matched sample had a higher average PS than the original, unmatched sample, because the AA-attender group (the target group being matched to) had a higher PS on average than the overall sample before matching. When, instead, the AA-nonattender group was matched using the AA-attender group, the difference in abstinence rates between the groups was higher—32%. The equivalent OR's for 30-day alcohol abstinence at the 1-year follow-up were 2.2 when the AA-attender group was matched as the target group and 3.7 when the AA-nonattender group was matched. To further adjust for covariates not fully balanced in PS matching, two GEE models were fitted to derive the conditional AA effect after matching. When the AA-attender group was matched using AA-nonattenders, the ORs changed from 2.2 to 2.3 after controlling for the one covariate with standardized difference (SD) bigger than 15%, and to 2.6 after the four covariates with SD bigger than 10% were controlled. Similarly, the ORs changed from 3.7 to 3.4 (adjusting for SD>15% or for SD>10%), when the AA-nonattender group was matched using AA-attenders.

4 Discussion

Those who attended AA the year following treatment-seeking differed significantly from those who did not attend, in terms of age (the mean age for the AA-attenders was higher than for the nonattenders) and marital status (more married people were in the AA-nonattenders than attenders). AA-attenders at follow-up also tended to have had higher alcohol problem severity and greater readiness to change at baseline, and were more likely to have reported prior exposure to AA, to have received pressure to seek help, and to have been recruited at a public community program or a private hospital treatment program rather than at an HMO program. These variables represent a certain predisposition to attend AA, referred to in the literature as a self-selection bias. To the extent that these variables also independently relate to higher rates of abstinence, they are potential confounders of AA effectiveness. Randomized trials would balance the AA-attender and AA-nonattender groups, removing the self-selection bias by randomizing individuals to either attend AA or not attend AA. As discussed in the introduction, no randomized trials of AA per se have reported a positive effect for those randomized to the AA condition. In part, this was because many individuals who were randomly assigned to the non-AA conditions also attended AA, highlighting the difficulty with trying to establish AA effectiveness using randomized designs.

We used a statistical approach, propensity scores, to equalize the AA-attender and AA-nonattender groups prior to evaluating the relationship between AA attendance and alcohol abstinence in a sample of treatment seekers. Although this could not balance the groups on

unobserved confounders, it was effective at balancing the groups on the observed confounders. After stratifying the study participants based on their propensity to attend AA (operationalized as a propensity score), differences in age, marital status, problem severity, readiness to change, prior AA exposure, and pressure to seek help between the AA attenders and AA nonattenders disappeared. Both propensity score techniques, stratification and matching, yielded similar results. Although there is no single optimal approach to apply the PS method, the use of these two techniques suggests the robustness of the findings. We conclude from this exercise that propensity score stratification and propensity score matching can be used to partially address the selection bias associated with one's propensity to attend AA.

How much of a correction did the propensity score method introduce into the assessment of AA effectiveness? Using the PS stratification technique, the difference in abstinence rates between the AA-attenders and nonattenders was adjusted downwards, by about 14%. Prior to PS stratification or matching, the rates of abstinence for the 30 days prior to follow-up were 68.6% among those who had attended AA and 37.7% for those who had not (a difference of 30.8%). Using stratification, the rate of abstinence among those who had not attended AA was 43.1%, while it remained at 69.6% for those who had attended AA (a difference of 26.5%). Stratification thus narrowed the advantage for abstinence in the AA-attender group by 4.3 percentage points, or by 14% ($4.3\% \div 30.8\%$).

When the AA attender group was matched using the nonattender group, the AA advantage narrowed further still, yielding a difference of 19.1%, equivalent to an OR of 2.2. When the few covariates that were not fully balanced were further adjusted in the GEE models, the OR changed to 2.6, still less than the OR of 3.6 in the unadjusted sample. While this PS matching technique of matching the treatment group (here, the AA attenders) as the preferred target group is most commonly reported in the literature, the difference between the unadjusted and adjusted AA effect should be viewed with some caution: This matched sample had a higher average PS than the original, unmatched sample, so that the comparison between the unadjusted and adjusted effects was based on different distributions of the underlying factors related to AA attendance. Nonetheless, while there definitely was a selection bias that confounded the effect of AA in the sample, it was not great. This aspect of our results confirm the robustness of AA effectiveness overall, because the results for higher abstinence associated with AA attendance following either type of propensity score adjustment were not greatly different from the unadjusted effect for AA attendance.

The PS stratification technique proved especially illustrative in terms of the degree of confounding caused by self-selection in evaluating AA effectiveness in non-experimental studies. It seems to suggest that, in terms of abstinence, AA is more helpful, or matters more, for those with a lower propensity to attend AA. Conversely, for those with a high propensity to go to AA, attending AA does not seem to make much of a difference in their odds of abstinence. We say this because the difference in the rate of abstinence between the AA attenders and nonattenders was very small, and was not statistically significant, in the stratification subclass which had the *highest* propensity scores for AA attendance: only a 12% difference in abstinence between the AA-attender and AA-nonattender groups in stratification subclass 5 (OR=1.6).

In contrast, 70% of the AA attenders in the stratification subclass with the *lowest* propensity scores (subclass 1) were abstinent, but only 32% of the AA nonattenders were abstinent—a 38% difference in abstinence rates, with an odds ratio of 4.9 favoring abstinence for those who attended AA in the year following treatment. The effect modification was also evident in stratification subclass 2, with a 38% difference in abstinence rate observed between AA attenders and nonattenders; for individuals in that stratum, attending AA increased their odds of abstinence by 5.0 over those who did not attend AA following treatment.

The observed modification of the effect across different strata with varying propensity scores using the PS stratification technique was similarly detected in PS matching, but is less intuitive. To summarize those results, the Odds Ratios for abstinence were 3.7 when the AA-attender group was matched using the AA-nonattender group, and 2.2 when the AA-nonattender groups was matched using the AA-attender group. When the AA-attender group was matched as the target group, the average PS, the probability of going to AA, for the overall matched sample was 0.68, whereas the average PS for the whole matched sample was 0.43 when the AA-nonattender group was matched as the target group. As in PS stratification, the OR for abstinence associated with AA shrank from 3.7 (studying AA's effect among a sample with a lower propensity to attend AA when we matched the AA-non-attender group by the attenders) to 2.2 (studying a sample with a higher propensity to attend AA when we matched AA-attender group by the non-attenders). After controlling for covariates not fully balanced by PS matching, this variability in AA effectiveness was still observed, but was of a smaller magnitude: the adjusted OR for abstinence when AA attenders were matched using nonattenders was 2.6, in comparison to an adjusted OR of 3.4 when AA nonattenders were matched using attenders.

This pattern of effect modification suggests that when propensity scores are higher (and by extension, one's motivation, problem severity, and prior treatment and AA exposure levels are high), the AA effect is not as strong as when the propensity scores (and propensity to attend AA) are lower. This may help to explain, in part, why TSF-type interventions are effective: they help those who would otherwise not avail themselves of AA, to do so. Results from Timko's trial of intensive referral to AA found a stronger effect for those with less prior AA exposure (Timko et al., 2006b), which is consistent with our findings of a stronger AA effect among those in the low propensity quartiles (where prior AA exposure also was low). The propensity scores constructed here also considered motivation to change, and the pattern of findings with PS stratification lend partial support for concerns about AA effectiveness being confounded in naturalistic studies if only the more motivated individuals tend to attend AA. Since the propensity scores include multiple potential confounds, it is not possible using this technique to parse out the bias associated with motivation versus prior AA exposure or problem severity, etc.

Researchers have argued for the need to "extend the evidence hierarchy" beyond the randomized clinical trial in assessing the effectiveness of services for individuals with alcohol and drug problems (Tucker and Roth, 2006). Statistical techniques that address the selection bias in observational studies represent part of this call to enhance evidence-based practice. We encourage more research in this area, as there is a great need for better studies of AA effectiveness. We recruited individuals at treatment entry, so ours is a study of AA effectiveness among treatment seekers who received at least some treatment. Future work should apply this technique to samples of untreated problem drinkers; and, to individuals who complete a full course of treatment. Our results suggest that the propensity score approach can be very useful in understanding the underlying confounders of AA effectiveness, and it will be important that our results be replicated, as ours is the first to apply this approach to the study of AA effectiveness. Please note that the two techniques used here, PS stratification and PS matching, are not the only applications of the propensity score method; other approaches, such as PS weighting and PS covariance adjustment, have also been proposed in the literature. These alternative techniques were not explored in this study, because of the potential bias they might have introduced. In an analysis comparing five methods for adjusting a confounding bias, Kurth et al. (2006) showed that the "inverse-probability-of-treatment-weighted" approach (PS weighting) greatly biased the treatment effect, as a result of excessively influential weights being assigned to observations with low propensity scores. The other method, covariance adjustment, can increase the bias if the covariance matrices in the treated and control groups are unequal (Rubin, 1979), prompting D'Agostino Jr. (1998) to suggest that PS matching and stratification methods (the approach taken here) should be preferred.

Finally, some limitations of the study should be noted. First, our analysis is restricted to a sample with relatively small size. This is a limitation particularly for the PS matching method, in which a large control pool is normally required for one-on-one matching. While matching with replacement was applied to reach the best match, the results might be sensitive to those individuals used many times during the matching. To examine the sensitivity of our finding, and also to confirm that our findings hold for a stronger (and perhaps more meaningful) level of AA exposure than having attended perhaps a single meeting, we replicated our analysis using a different categorization to designate AA exposure during the follow-up period: 3 or more meetings, versus no more than 2 meetings. Using the PS stratification technique, the adjusted OR for abstinence associated with attending 3 or more AA meetings at follow-up (compared to 2 or fewer meetings) was 3.3, a moderate drop from the unadjusted OR of 3.8. Effect modification was also observed in the quintiles from lowest to highest PS, with respective ORs of 4.7, 5.3, 2.5, 2.8 and 2.3. Under PS matching, the OR for abstinence was 2.8 when the 3+AA attender group was matched by the <3 meetings group, and the OR was 3.3 when the <3-meetings group was matched by the 3+ group. Overall, the sensitivity analysis confirmed the main findings of our analysis. However, the effect modification comparing the two different matching target groups was not quite as strong as we had observed before, calling for replication of our study with a bigger sample (and, for considering various cut-points of AA exposure, guided by the literature).

Second, only observed confounders could be included in the construction of the propensity scores; thus “omitted variables” not observed remain unadjusted. Our study did not ask respondents about several important mediators of AA effectiveness that have been recognized since these data were gathered, such as self-efficacy (Kelly et al., 2002; Morgenstern et al., 1997) and coping skills (Timko et al., 2005) (Humphreys et al., 1999). We also do not have data on length of stay in treatment. These variables could well act as additional confounders, and would be appropriate to include in the Propensity Score calculation.

Third, while we addressed the potential temporality problem somewhat by minimizing the overlap of periods measuring AA attendance (last 12 months at the 1-year follow-up) and abstinence (last 30 days at the 1-year follow-up), our approach is not totally free from the “reverse causation” problem. For example, earlier AA attenders might keep going to AA after they quit drinking, while those who keep drinking are more likely to drop out of AA. In this case, the later AA attendance is caused (or sustained, one could argue) by abstinence, rather than the other way around. To at least partially address this problem, in our study the abstainers at baseline were dropped from the analysis.

Still other limitations arise from attrition from the follow-up interview and from cases on variables necessary for constructing the propensity scores. While the attrition analysis found no differences in alcohol problem severity, males and African Americans were under-represented at follow-up (Kaskutas et al., 2002). However, since AA attenders and non-attenders were matched either by stratification or by case-by-case matching, subjects in analyses using the propensity score approach thus are not considered randomly selected, and inference is not based on probability sampling. For this reason, the problem of non-random missing cases is somewhat less a concern here.

Lastly, these data were collected about 10 years ago. This is somewhat less of a concern, because of our emphasis on a statistical technique to address a substantive open question in the literature of high relevance to providers and policymakers. However, it is possible that other confounders of AA effectiveness (in addition to those just mentioned) may be at work today, which may have arisen due to economic and other changes in the last 10 years.

References

- Ammon LN, Bond J, Matzger H, Weisner C. Gender differences in the relationship of community services and informal support to seven-year drinking trajectories of alcohol-dependent and problem drinkers. *Journal of Studies on Alcohol and Drugs* 2008;69:140–150. [PubMed: 18080074]
- Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med* 2008;27:2037–2049. [PubMed: 18038446]
- Barrows D. The community orientation of social model and medical model recovery programs. *J Subst Abuse Treat* 1998;15:55–64. [PubMed: 9534127]
- Bogenschutz, MP. Individual and contextual factors that influence AA affiliation and outcomes. In: Galanter, M.; Kaskutas, LA., editors. *Research on Alcoholics Anonymous and Spirituality in Addiction Recovery*. Springer; New York: 2008. p. 413-433.
- Borkman TJ, Kaskutas LA, Owen P. Contrasting and converging philosophies of three models of alcohol/other drugs treatment: Minnesota Model, Social Model, and addiction Therapeutic Communities. *Alcohol Treat Quart* 2007;25:21–38.
- Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *J Am Stat Assoc* 1995;90:443–450.
- Brandsma, JM.; Maultsby, MC., Jr; Welsh, RJ. *Outpatient Treatment of Alcoholism: A review and comparative study*. University Park Press; Baltimore, MD: 1980.
- Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006;163:1149–1156. [PubMed: 16624967]
- D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265–2281. [PubMed: 9802183]
- D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17:2265–2281. [PubMed: 9802183]
- Dawson DA, Grant BF, Stinson FS, Chou PS, Huang B, Ruan WJ. Recovery from DSM-IV alcohol dependence: United States, 2001–2002. *Addiction* 2005;100:281–292. [PubMed: 15733237]
- Dehejia RH, Wahba S. Propensity score-matching methods for nonexperimental causal studies. *The Review of Economics and Statistics* 2002;84:151–161.
- Del Boca, FK.; Mattson, ME. The gender matching hypothesis. In: Longabaugh, R.; Wirtz, PW., editors. *Project MATCH Hypotheses: Results and causal chain analyses*. National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 2001. p. 186-203.
- Ditman KS, Crawford GG, Forgy EW, Moskowitz H, MacAndrew C. A controlled experiment on the use of court probation for drunk arrests. *Am J Psychiatry* 1967;124:160–163. [PubMed: 4951569]
- Emrick, CD.; Tonigan, JS.; Montgomery, HA.; Little, L. Alcoholics Anonymous: what is currently known?. In: McCrady, BS.; Miller, WR., editors. *Research on Alcoholics Anonymous: Opportunities and Alternatives*. Rutgers Center of Alcohol Studies; New Brunswick, NJ: 1993. p. 41-78.
- Fortney J, Booth B, Zhang M, Humphrey J, Wiseman E. Controlling for selection bias in the evaluation of Alcoholics Anonymous as aftercare treatment. *J Stud Alcohol* 1998;59:690–697. [PubMed: 9811090]
- George AA, Tucker JA. Help-seeking for alcohol-related problems: Social contexts surrounding entry into alcoholism treatment or Alcoholics Anonymous. *J Stud Alcohol* 1996;57
- Humphreys K, Kaskutas LA, Weisner C. The relationship of pre-treatment Alcoholics Anonymous affiliation with problem severity, social resources, and treatment history. *Drug Alcohol Depend* 1998;49:123–131. [PubMed: 9543649]
- Humphreys K, Mankowski ES, Moos RH, Finney JW. Do enhanced friendship networks and active coping mediate the effect of self-help groups on substance abuse? *Ann Behav Med* 1999;21:54–60. [PubMed: 18425655]
- Humphreys K, Mavis BE, Stöffelmayr BE. Factors predicting attendance at self-help groups after substance abuse treatment: preliminary findings. *J Consult Clin Psychol* 1991;59:591–593. [PubMed: 1918563]

- Humphreys K, Phibbs CS, Moos RH. Addressing self-selection effects in evaluations of mutual help groups and professional mental health services: an introduction to two-stage sample selection models. *Eval Prog Plan* 1996;19:301–308.
- Isenhardt CE. Pretreatment readiness for change in male alcohol dependent subjects: Predictors of one-year follow-up status. *J Stud Alcohol* 1997;58:351–357. [PubMed: 9203115]
- Kaskutas LA. Special issue: the social model approach to substance abuse recovery. *J Subst Abuse Treat* 1998;15
- Kaskutas, LA. AA Effectiveness: Faith Meets Science [Invited plenary]. Annual Medical Scientific Conference of the American Society of Addiction Medicine; April 27; Miami, FL. 2007.
- Kaskutas LA. Alcoholics Anonymous effectiveness: faith meets science. *J Addict Dis*. in press
- Kaskutas LA, Ammon LN, Weisner C. A naturalistic analysis comparing outcomes at substance abuse treatment programs with differing philosophies: social and clinical model perspectives. *International Journal of Self Help and Self Care* 2004;2:111–133.
- Kaskutas LA, Bond J, Humphreys K. Social networks as mediators of the effect of Alcoholics Anonymous. *Addiction* 2002;97:891–900. [PubMed: 12133128]
- Kaskutas LA, Weisner C, Lee M, Humphreys K. Alcoholics Anonymous affiliation at treatment intake among white and black Americans. *J Stud Alcohol* 1999;60:810–816. [PubMed: 10606493]
- Kaskutas, LA.; Ye, Y.; Greenfield, TK.; Witbrodt, J.; Bond, J. Epidemiology of Alcoholics Anonymous participation. In: Galanter, M.; Kaskutas, LA., editors. *Research on Alcoholics Anonymous and Spiritual Aspects of Addiction Recovery*. Springer; New York: 2008. p. 261-282.
- Kelly JF, Myers MG, Brown SA. Do adolescents affiliate with 12-step groups? A multivariate process model of effects. *J Stud Alcohol* 2002;63:293–304. [PubMed: 12086130]
- Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 2006;163:262–270. [PubMed: 16371515]
- Leuven, E.; Sianesi, B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. 2003. <http://ideas.repec.org/c/boc/bocode/s432001.html>
- Matzger H, Delucchi K, Weisner C, Ammon LN. Does marital status predict long-term drinking? Five-year observations of dependent and problem drinkers. *J Stud Alcohol* 2004;65:255–265. [PubMed: 15151358]
- McKellar J, Stewart E, Humphreys K. Alcoholics Anonymous involvement and positive alcohol-related outcomes: cause, consequence, or just a correlate? A prospective 2-year study of 2,319 alcohol-dependent men. *J Consult Clin Psychol* 2003;71:302–308. [PubMed: 12699024]
- McLellan AT, Luborsky L, Cacciola JS, Griffith J, Evans F, Barr HL, O'Brien CP. New data from the Addiction Severity Index: reliability and validity in three centers. *J Nerv Ment Dis* 1985;173:412–423. [PubMed: 4009158]
- Morgenstern J, Bux DA Jr, Labouvie E, Morgan T, Blanchard KA, Muench F. Examining mechanisms of action in 12-step community outpatient treatment. *Drug Alcohol Depend* 2003;72:237–247. [PubMed: 14643940]
- Morgenstern J, Labouvie E, McCrady BS, Kahler CW, Frey RM. Affiliation with Alcoholics Anonymous following treatment: a study of its therapeutic effects and mechanisms of action. *J Consult Clin Psychol* 1997;65:768–777. [PubMed: 9337496]
- Prochaska, JO.; DiClemente, CC. Toward a comprehensive model of change. In: Miller, W.; Heather, N., editors. *Treating Addictive Behaviors: Processes of change*. Plenum; New York, NY: 1984. p. 3-27.
- Project MATCH Research Group. Matching alcoholism treatment to client heterogeneity: Project MATCH posttreatment drinking outcomes. *J Stud Alcohol* 1997;58:7–29. [PubMed: 8979210]
- Project MATCH Research Group. Matching alcoholism treatments to client heterogeneity: project MATCH three-year drinking outcomes. *Alcohol Clin Exp Res* 1998;22:1300–1311. [PubMed: 9756046]
- Room R, Greenfield TK. Alcoholics Anonymous, other 12-step movements and psychotherapy in the U.S. population, 1990. *Addiction* 1993;88:555–562. [PubMed: 8485433]

- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41–55.
- Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984;79:516–524.
- Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 1985;39:33–38.
- Rubin DB. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *J Am Stat Assoc* 1979;74:318–324.
- Schmidt LA, Ye Y, Greenfield TK, Bond J. Ethnic disparities in clinical severity and services for alcohol problems: results from the National Alcohol Survey. *Alcohol Clin Exp Res* 2007;31:48–56. [PubMed: 17207101]
- Timko, C. Outcomes of AA for special populations. In: Galanter, M.; Kaskutas, LA., editors. *Research on Alcoholics Anonymous and Spirituality in Addiction Recovery*. Springer; New York: 2008. p. 373-392.
- Timko C, Billow R, DeBenedetti A. Determinants of 12-step group affiliation and moderators of the affiliation-abstinence relationship. *Drug Alcohol Depend* 2006a;83:111–121. [PubMed: 16338102]
- Timko C, DeBenedetti A. A randomized controlled trial of intensive referral to 12-step self-help groups: one-year outcomes. *Drug Alcohol Depend* 2007;90:270–279. [PubMed: 17524574]
- Timko C, DeBenedetti A, Billow R. Intensive referral to 12-step self-help groups and 6-month substance use disorder outcomes. *Addiction* 2006b;101:678–688. [PubMed: 16669901]
- Timko C, Finney JW, Moos RH. The 8-year course of alcohol abuse: gender differences in social context and coping. *Alcohol Clin Exp Res* 2005;29:612–621. [PubMed: 15834227]
- Timko C, Moos RH, Finney JW, Connell EG. Gender differences in help-utilization and the 8-year course of alcohol abuse. *Addiction* 2002;97:877–889. [PubMed: 12133127]
- Tonigan JS, Miller WR, Juarez P, Villanueva M. Utilization of AA by Hispanic and non-Hispanic white clients receiving outpatient alcohol treatment. *J Stud Alcohol* 2002;63:215–218. [PubMed: 12033698]
- Tucker JA, Gladsjo JA. Help-seeking and recovery by problem drinkers: characteristics of drinkers who attended Alcoholics Anonymous or formal treatment or who recovered without assistance. *Addict Behav* 1993;18:529–542. [PubMed: 8310872]
- Tucker JA, Roth DL. Extending the evidence hierarchy to enhance evidence-based practice for substance use disorders. *Addiction* 2006;101:918–932. [PubMed: 16771885]
- Walsh DC, Hingson RW, Merrigan DM, Levenson SM, Cupples LA, Heeren T, Coffman GA, Becker CA, Barker TA, Hamilton SK, McGuire TG, Kelly CA. A randomized trial of treatment options for alcohol-abusing workers. *N Eng J Med* 1991;325:775–782.
- Weisner C, Delucchi K, Matzger H, Schmidt L. The role of community services and informal support on five-year drinking trajectories of alcohol dependent and problem drinkers. *J Stud Alcohol* 2003;64:862–873. [PubMed: 14743951]
- Weisner C, Greenfield TK, Room R. Trends in the treatment of alcohol problems in the U.S. general population, 1979 through 1990. *Am J Public Health* 1995;85:55–60. [PubMed: 7832262]
- Weisner C, Matzger H. A prospective study of the factors influencing entry to alcohol and drug treatment. *J Behav Health Serv Res* 2002;29:126–137. [PubMed: 12032970]
- Witbrodt J, Bond J, Kaskutas LA, Weisner C, Jaeger G, Pating D, Moore C. Day hospital and residential addiction treatment: randomized and nonrandomized managed care clients. *J Consult Clin Psychol* 2007;75:947–959. [PubMed: 18085911]
- Wooldridge, JM. *Econometric analysis of cross section and panel data*. The MIT Press; Cambridge, MA: 2002.

Table 1

Comparison of potential baseline confounders in the AA-attender and AA-nonattender groups, before and after propensity score (PS) stratification and PS matching

	Comparisons prior to stratification or matching			PS stratification		PS Matching ¹	
	AA-attender (N=336) Mean (sd)	AA- nonattender (N=233) Mean (sd)	Standardized Difference ² in %	F-statistics	F-statistics	Standardized Difference ² in %	t-statistics
Demographics							
Male	0.59 (0.49)	0.56 (0.50)	6.93	0.66	0.08	0.00	0.00
Mean Age	38.8 (10.1)	36.8 (11.5)	18.4	4.76*	0.00	-11.1	-0.69
Ethnicity: White	0.63 (0.48)	0.58 (0.49)	9.67	1.29	0.00	4.32	0.30
Black	0.26 (0.44)	0.26 (0.44)	-0.36	0.00	0.01	-4.02	-0.27
Others	0.11 (0.32)	0.16 (0.37)	-13.3	2.51	0.00	-1.03	-0.07
Marital: Married	0.34 (0.48)	0.45 (0.50)	-21.4	6.34*	0.01	-0.72	-0.05
Sep/Div/Widow	0.36 (0.48)	0.28 (0.45)	16.8	3.84	0.01	-9.09	-0.60
Single	0.30 (0.46)	0.27 (0.45)	5.72	0.45	0.00	-10.1	-0.70
# of years of Education	3.34 (1.02)	3.18 (0.98)	15.7	3.36	0.00	-1.40	-0.09
Motivation							
Readiness to change index	50.0 (6.7)	46.6 (7.5)	48.3	32.8***	0.08	-11.5	-0.85
Coercion							
# who pressure you to get treatment	1.85 (1.31)	1.58 (1.16)	22.6	6.88**	0.00	16.8	1.23
# who give you ultimatum	0.58 (0.80)	0.53 (0.72)	5.96	0.48	0.03	0.93	0.07
Problem severity							
ASI composite alcohol score	0.43 (0.32)	0.34 (0.30)	30.8	13.0***	0.00	6.03	0.43
# of Dependence symptoms	5.20 (2.77)	3.69 (2.67)	55.8	42.5***	0.18	9.84	0.65
# of Alcohol-rel. consequences	1.42 (1.42)	0.99 (1.15)	33.4	14.8***	0.08	-3.81	-0.26
Help-seeking							
# of AA meetings last year	36.6 (62.9)	8.08 (27.2)	58.8	42.3***	0.32	8.15	0.69
# of treatment episodes last year	4.24 (18.2)	1.96 (14.6)	13.9	2.55	0.02	4.00	0.85
Type of treatment							
Private	0.29 (0.46)	0.18 (0.38)	28.2	10.60**	0.05	2.51	0.16

	Comparisons prior to stratification or matching			PS stratification		PS Matching ¹	
	AA-attender (N=336) Mean (sd)	AA- nonattender (N=233) Mean (sd)	Standardized Difference ² in %	F-statistics	F-statistics	Standardized Difference ² in %	t-statistics
HMO	0.33 (0.47)	0.67 (0.47)	-72.0	71.3***	0.00	-5.98	-0.39
Public	0.38 (0.48)	0.15 (0.36)	51.5	34.7***	0.07	4.11	0.27
<i>Social Influences</i>							
Size of support network:							
# to talk to	4.25 (5.25)	3.62 (4.32)	13.1	2.29	0.13	-8.34	-0.77
# can get help from	4.15 (4.86)	4.15 (4.74)	0.01	0.00	0.00	-8.95	0.69
# in regular contact with	5.61 (5.59)	6.20 (5.39)	-10.8	1.60	0.06	1.92	0.14
Drinking of network:							
# of heavy or problem drinkers	0.85 (2.17)	0.86 (2.03)	-0.20	0.00	0.06	6.71	0.65
Prop. heavy/problem drinkers	0.13 (0.25)	0.15 (0.25)	-6.53	0.59	0.00	7.46	0.53
# who encourage you to drink	0.39 (2.37)	0.25 (1.20)	7.27	0.66	0.76	9.36	1.33
Prop. who encourage you to drink	0.038 (0.15)	0.046 (0.17)	-4.72	0.31	0.00	8.55	0.74

¹ Standardized differences and t-statistics shown in the table are for PS matching on the AA-attender group

² Standardized difference (in %) = $100 * (\bar{x}_I - \bar{x}_C) / [(s_I^2 + s_C^2)/2]^{1/2}$, where $\bar{x}_I$ and $\bar{x}_C$ are means, and s_I^2 and s_C^2 are variances, in the "treated" (AA attender) and "control" (AA non-attender or matched) groups.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

Table 2

Rates of 30-day abstinence at 1-year follow-up based on AA attendance status during the follow-up period, before and after propensity score stratification and propensity score matching

	AA-attender at follow-up	N	Rate of abstinence at follow-up mean (se)	Differences in rate of abstinence Mean (se)	Odds Ratios associated with AA attendance
<i>Before Propensity Score Adjustment</i>					
	No	228	0.377 (0.032)		
	Yes	334	0.686 (0.025)	-0.308 (0.041)***	3.60
<i>Propensity Score Stratification</i>					
Subclass (from lowest AA propensity to highest AA propensity)					
1	No	93	0.323 (0.049)		
	Yes	20	0.700 (0.105)	-0.377 (0.116)**	4.90
2	No	68	0.368 (0.059)		
	Yes	43	0.744 (0.067)	-0.377 (0.091)***	5.00
3	No	37	0.432 (0.083)		
	Yes	75	0.707 (0.053)	-0.274 (0.095)**	3.16
4	No	21	0.476 (0.112)		
	Yes	92	0.652 (0.050)	-0.176 (0.117)	2.06
5	No	9	0.556 (0.176)		
	Yes	104	0.673 (0.046)	-0.118 (0.165)	1.65
Adjusted	No	228	0.431 (0.106)		
	Yes	334	0.696 (0.068)	-0.265 (0.126)*	3.02
<i>Propensity Score Matching (nearest distance with replacement)</i>					
Matching AA-attender group	No	102	0.479 (0.067)		
	Yes	282	0.670 (0.028)	-0.191 (0.073)**	2.21
Matching AA-nonattender group	No	209	0.383 (0.034)		
	Yes	104	0.699 (0.059)	-0.316 (0.068)***	3.74

*
p<0.05

100

 $p<0.001$
**