

Published in final edited form as:

Drug Alcohol Depend. 2013 April 1; 129(1-2): 54–59. doi:10.1016/j.drugalcdep.2012.09.011.

A novel application of propensity score matching to estimate Alcoholics Anonymous' effect on drinking outcomes

Stephen Magura^{a,**}, Joseph McKean^b, Scott Kosten^b, and J. Scott Tonigan^c

^aThe Evaluation Center, Western Michigan University, Kalamazoo MI 49008 USA

^bStatistics Department, Western Michigan University, Kalamazoo MI 49008 USA

^cCenter on Alcoholism, Substance Abuse and Addictions, University of New Mexico, Albuquerque, New Mexico 87106 USA

Abstract

Background—Randomized controlled trials (RCTs) of mutual aid, including Alcoholics Anonymous (AA), are notoriously difficult to conduct and correlational studies are problematic to interpret due to potential confounds.

Methods—A secondary analysis was conducted of Project MATCH, a RCT of alcoholism treatments. Although MATCH did not randomly assign subjects to AA vs. no AA, the 12 Step Facilitation (TSF) condition did result in a higher proportion of subjects attending community AA meetings than in the other two treatment conditions. The key inference is that there exists a latent subgroup in MATCH who attended AA *only* because its constituents received TSF, not because of the “normal” factors leading to self-selection of AA. A novel application of propensity score matching (PSM) allowed four latent AA-related subgroups to be identified to estimate an unconfounded effect of AA on drinking outcomes.

Results—The study hypotheses were supported: subjects who consistently attended AA solely due to their exposure to TSF (the “Added AA” subgroup) had better drinking outcomes than equivalent subjects who did not consistently attend AA, but would have so attended, had they been exposed to TSF (the “Potential AA” subgroup); this indicates an AA effect on drinking.

Conclusions—The analysis presents evidence that consistent AA attendance improves drinking outcomes, independent of “normal” confounding factors that make correlations between AA attendance and outcomes difficult to interpret.

© 2012 Elsevier Ireland Ltd. All rights reserved.

**Corresponding author: The Evaluation Center, Western Michigan University, 1903 W. Michigan Ave., Kalamazoo MI 49008 USA, Tel: 269-387-5895 Fax: 269-387-5923 stephen.magura@wmich.edu.

Contributors

Magura conceptualized the study, provided scientific oversight, contributed to all phases of the work, and was the primary author of the manuscript. McKean and Kosten developed the program code, conducted the statistical analysis, and contributed to the interpretation of results and the manuscript. Tonigan helped develop and provided the study datasets and contributed to the interpretation of results and the manuscript.

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.

Conflict of Interest

All authors declare they have no conflict of interest.

Keywords

propensity score; Alcoholics Anonymous; Project MATCH; self-help groups; alcoholism; bootstrapping

1. INTRODUCTION

Alcoholics Anonymous (AA) was founded in 1933 as a mutual aid fellowship for alcoholics and is currently the largest addiction-related self-help organization in the world (Humphreys, 2004). AA reports that in 2011, there were 57,905 groups and 1,279,664 members in the U.S. and 107,967 groups and 2,057,672 members worldwide (Alcoholics Anonymous, 2011). The latest U.S. National Survey on Drug Use and Health (Substance Abuse and Mental Health Services Administration, 2011) reports that 4.1 million people received help for a substance abuse problem in 2010, including 2.3 million who used self-help groups exclusively or in conjunction with treatment. For many people with alcohol problems, AA is the only source of help they ever use (Hasin and Grant, 1995).

There is a considerable research literature that suggests associations between AA participation and less drinking or abstinence (Tonigan et al., 1996). These studies are almost universally correlational in nature, however, including those with longitudinal data. Moreover, the methodological quality of the studies was most frequently rated as “poor” by the article’s authors, often due to lack of representative selection of subjects and lack of random assignment designs. In fact, Tonigan et al. (1996) present evidence that poor methodological study quality may lead to overestimating the impact of AA on drinking outcomes. Several randomized controlled trials (RCTs) of AA have been conducted (e.g., Ditman et al. 1967; Brandsma et al. 1980; Walsh et al., 1991), but they also suffer from significant limitations in their methods or interpretability; incidentally, none reported a positive AA effect. Attempted RCTs of AA face practical obstacles; in particular, participation in AA groups cannot ethically be denied to study “controls” and some alcoholism study subjects always go to AA on their own, even if not specifically assigned to go (Humphreys, 2004; Brandsma et al., 1980; Tonigan, 2003).

Additional correlational studies indicating positive effects of AA have been published subsequent to Tonigan et al.’s (1996) review (e.g., Kaskutas et al., 2002, 2005; Gossop et al., 2003; Chappel and DuPont, 1999; Schuckit et al., 1997), including prior analyses for Project MATCH (Tonigan et al., 2003; Kelly et al., 2011a). Three studies that randomly assigned subjects to 12-step Facilitation (TSF) found that the TSF subjects reported higher rates of alcohol abstinence at follow-up (Timko and DeBenedetti, 2007; Walitzer et al., 2009; Litt et al., 2009).

A major problem in trying to determine the existence of an effect of AA participation on drinking is the tendency for people with alcohol problems to self-select for AA. Differences between AA participants and non-participants have been well documented in the literature (e.g., Bogenschutz, 2008; Emrick, 1993; Morgenstern et al, 2003). Factors that could affect both AA participation and drinking, and thus could act as confounders when relating AA participation to drinking, are alcohol problem severity, psychological problem severity, motivation for change, prior AA experience, prior treatment experience, social support for recovery and sociodemographic characteristics, among many others. Several studies have used special statistical techniques (instrumental variables, propensity analysis) to control for subject self-selection into AA, reporting positive AA effects (Fortney et al., 1998; Humphreys et al. 1996; Ye and Kaskutas, 2009). While certainly advancing the field, even these techniques are limited in non-experimental studies because of the difficulty of

identifying plausible instrumental variables or the restriction to observed predictors of AA participation in propensity analysis, which may be incomplete.

The purpose of the current study is to present a novel application of propensity score matching (PSM) to estimate the unbiased effect of AA participation on drinking outcomes that capitalizes directly on the treatment randomization element in a U.S. alcoholism treatment dataset, Project MATCH (Project MATCH Research Group, 1993; Babor and Del Boca, 2003). The key observation inspiring the study is that subjects who received 12-step Facilitation (TSF) in Project MATCH attended AA at a higher rate than subjects receiving the other two therapies. Although alcoholics generally self-select themselves for AA, this observation implies that there is a subset of subjects in Project MATCH who attended AA *only* because they received TSF - not because of the “normal” factors leading to self-selection. For example, it may be that the “additional” AA participants in TSF are less committed to abstinence or perceive lower problem severity than the “expected” AA participants; elevation on these factors is found to be associated with self-selection into AA (Morgenstern et al., 1997; Humphreys et al., 1991). However, explicit “covariate control” for such factors, even when a study attempts to measure them, inevitably results in serious problems of interpretation (Christenfeld et al., 2004).

Conversely, we can infer the existence of a subset of subjects among those receiving the other two therapies that did not consistently attend AA, but would have done so had they received TSF. Thus, the primary hypothesis is that subjects who attended AA consistently, but did so only because TSF induced them to do so, will have better drinking outcomes than subjects who did not attend AA consistently, but would have done so had they received TSF. It is proposed that this subgroup comparison, if it can be constructed, essentially eliminates the influence of AA self-selection factors from the estimate of AA’s effect on drinking outcomes, thus yielding a putatively unbiased estimate of AA’s effect on drinking.

2. METHODS

2.1. Subjects

Project MATCH was a U.S. alcoholism treatment trial conducted between 1989 and 1994. Study subjects were recruited at outpatient sites (n = 952 at 5 sites) and aftercare sites (n = 774 at 5 sites), the latter after discharge from inpatient or day hospital treatment. More than 90% were diagnosed as dependent on alcohol and none were dependent on illicit drugs with the exception of marijuana. Selected sociodemographic characteristics of the subjects are in Project MATCH Research Group (1997: table 2).

2.2. Procedures

Volunteers underwent informed consent, completed a baseline assessment battery and were randomly assigned to one of three treatments: Twelve Step Facilitation (TSF), Cognitive-Behavioral Therapy (CBT), or Motivational Enhancement Therapy (MET). The primary purpose of TSF was to educate subjects about 12-step recovery and maximize their participation in community AA groups. In contrast, the primary focus of CBT was to increase ability to deal with high-risk situations that commonly lead to relapse of drinking and the primary focus of MET was to mobilize the subject’s own resources to bring about the changes needed to achieve sobriety. TSF and CBT had 12 scheduled weekly sessions and MET had 4 sessions, all over a 3-month period. Follow-up interviews were conducted at the scheduled end of treatment (after 3 months) and after 6, 9, 12, and 15 months. Follow-up rates were more than 90% in both subsamples at each follow-up. Project MATCH is described in detail in Project MATCH Research Group (1993) and Babor and Del Boca

(2003). The current secondary analysis of the dataset was approved by the Institutional Review Board of Western Michigan University.

2.3. Measures

The primary drinking outcome for the present analysis is alcohol abstinence, as measured by “percent days abstinent” for the 90 day period prior to the 15 month follow-up (PDA15), which was arcsine transformed in accord with all analyses previously reported for Project MATCH. The secondary drinking outcome is a previously validated outcome composite classified as: Abstinent from alcohol (1); moderate drinking without problems related to drinking (2); heavy drinking or problems related to drinking (3); heavy drinking and problems related to drinking (4) (Cisler and Zweben, 1999; Zweben and Cisler, 2003). This outcome composite was also measured for the 90-day period prior to the 15-month follow-up (OTCM15). Drinking was captured on Form 90 (Miller and Del Boca, 1994) and problems were identified on the Drinker Inventory of Consequences (Miller et al., 1995). “Consistent AA attendance” for subjects was defined as some attendance during each successive 90-day period of treatment and follow-up during months 1-15, following the definition of Tonigan et al. (2003: table 11.2); all others were classified as “no attendance/inconsistent attendance.” The relevance of this categorization is supported by empirical findings indicating that consistent AA attendance over time is related to alcohol/drug use abstinence (e.g., Fiorentine and Hillhouse, 2003). Twelve step programs also stress the importance of consistent attendance to obtain maximum benefits.

Baseline variables available as covariates for propensity score matching were: sociodemographics, drinking and alcohol use consequences, employment status, psychiatric severity, cognitive impairment, conceptual level, meaning-seeking (a desire for greater life meaning; Crumbaugh, 1977), motivation for change, sociopathy, social support for drinking, alcoholism typology (type A/B), anger, criminal behavior, depression (Beck), antisocial personality disorder; assertion of autonomy; psychopathology; religiosity; self-efficacy-confidence (including alcohol abstinence self-efficacy), self-efficacy/temptation; social functioning; and readiness for change. A complete list of measures may be found in Project MATCH Research Group (1993) and Babor and Del Boca (2003).

2.4. Analysis Plan

2.4.1. The Basic Structure—We begin with the inference that the TSF intervention created a category of alcoholics who voluntarily choose to participate in AA, although they would not have “routinely” or “normally” done so had they been assigned to one of the other treatments. This is determined as follows. The proportion of consistent AA attenders in the combined CBT and MET conditions is 21.2% and the proportion in TSF is 34.7%; see Figure 1. (CBT and MET were pooled because there was no significant difference in consistent AA attendance—21.1% vs. 21.7%, respectively, based on the unimputed dataset, $p < .80$). Thus, we can infer that 13.5% ($34.7\% - 21.2\% = 13.5\%$, $n=79$) of the subjects in the TSF condition would not have attended AA consistently save for their assignment to TSF. We can term these 79 subjects the “Added” AA participants, while the other 123 are those who would routinely be “Expected” to attend AA in the absence of encouragement by TSF. The consistent AA attenders in TSF may be characterized as consisting of two latent subgroups: the Added AA and Expected AA subgroups.

The advantage of the random assignment feature of Project MATCH is that the AA participation of these “Added” AA subjects cannot be explained by “normal” AA self-selection factors, since the factor explaining their AA participation is known and is different, i.e., random assignment to TSF. Or to state it counterfactually, if the 582 clients assigned to TSF had instead been assigned to CBT or MET, only 21.2 % of them would have been

expected to attend AA consistently, rather than the observed 34.7%. This difference of 13.5% is statistically significant (chi-squared = 26.6, $df = 1$, $p < .001$) and the 95% confidence interval is (9.1%, 18.1%). This difference is consistent with the intent of the TSF intervention.

An analogous observation can be made about the inconsistent/no AA group in the CBT/MET condition. We know from the TSF condition that 65.3% of subjects could not be induced to attend AA consistently even with the encouragement of TSF (see Figure 1). Thus, we can infer that 13.2% (78.8% minus 65.6% = 13.2%, $n = 151$) of the inconsistent/non-AA subjects in CBT/MET would have attended AA consistently had they been exposed to TSF. (The discrepancy between 65.3% and 65.6% is due to rounding to whole persons.) We term the subjects in CBT/MET who would not be induced by TSF to attend AA consistently as the “Unreceptive” to AA subjects ($n = 750$) and those who would be induced to attend AA if exposed to TSF as the “Potential” AA participants ($n = 151$). The inconsistent/non-AA subjects in CBT/MET may be characterized as consisting of two latent subgroups: the Unreceptive to AA and the Potential AA subgroups.

If we could identify the latent subgroups, we could obtain an unbiased estimate of the effect of consistent AA attendance on drinking. The group comparison would be between alcoholics who consistently attended AA *solely* because they were exposed to TSF and alcoholics who were inconsistent/non-attenders *solely* because they were *not* exposed to TSF. Self-selection for consistent AA attendance would no longer present a confound in this comparison.

Our approach to constructing this group comparison builds on the technique of propensity score matching (PSM). A typical research situation is where one group has received a treatment and there is a second, non-randomly assigned group (“comparison group”) that has not received the treatment, but whose characteristics overlap with those of the treatment group. The two groups cannot be directly compared to compute a treatment effect because their characteristics differ, commonly termed selection bias. A predicted probability of treatment group membership is computed for all subjects and treatment group members are matched with comparison group members based on similar predicted probabilities of treatment group membership, termed the “propensity score.” The matched subjects are then used to compute a treatment effect, which has been adjusted for observed selection bias factors (Guo and Fraser, 2009; Rosenbaum and Rubin, 1983).

Propensity score matching (PSM) can be adapted to identify individuals belonging to the Project MATCH latent subgroups as described above. The basic idea is as follows. Using an array of baseline variables, we can determine the characteristics of the consistent AA subjects in CBT/MET and identify the consistent AA subjects in TSF who are most similar to those in CBT/MET; this would yield the Expected AA subgroup in TSF. The remaining consistent AA subjects in TSF would be the Added AA subgroup. Similarly, we can determine the characteristics of the inconsistent/non-AA subjects in TSF and identify the inconsistent/non-AA subjects in CBT/MET who are most similar to those in TSF; this would yield the Unreceptive to AA subgroup in CBT/MET. The remaining inconsistent/non-AA subjects in CBT/MET would be the Potential AA subgroup. The difference between the Added and Potential AA groups is then an estimate of AA’s effect on drinking outcomes.

2.4.2. Power Analysis—Most previous analyses of Project MATCH have analyzed the Aftercare and Outpatient subsamples separately. Given the estimated sizes of the Added ($n = 79$) and Potential ($n = 151$) latent groups in the data, the ability to detect a low medium effect size ($d = .40$) with a power = 0.80 could only be achieved by pooling the Aftercare and Outpatient subsamples (Cohen, 1988). Given the known pooled standard deviations of

PDA15 (0.47) and OTCM15 (1.32) in the data, this would allow the detection of an unstandardized PDA15 difference between the Added and Potential groups of 0.19 and an OTCM15 difference of 0.53.

2.4.3. Missing Data—Preliminary analysis involved creating a dataset based on data imputation for missing values to maximize statistical power. In the original dataset, there were 180 subjects with missing observations on some variables required for the analyses. The R package AMELIA II: Version 1.2-1.8 was used for multiple imputation of missing values based on a bootstrap EMB algorithm (Honaker et al., 2010). After data imputation, the dataset retained the full 1726 Project MATCH subjects; in terms of consistency, the breakdown of AA attendance among the treatment groups for the imputed dataset is virtually identical with the breakdown for the original dataset. The data layout in Figure 1 is based on the imputed dataset.

2.4.4. Effect Estimation—Logistic regressions for the PSM were performed by the R function GLM. Effect sizes for PSM were estimated with confidence intervals formed by a bootstrap-type resampling procedure written for this study by the authors. The analyses used the bootstrapped percentile confidence intervals as discussed in Efron and Tibshirani (1993). Briefly, for an estimator, these confidence intervals are based on the quantiles from the resampling distribution of the estimator; see also Hogg et al. (2012: chapter 4). The assumptions for effect estimation are that the subjects are independent of one another and that the random errors of responses have finite variance. The basic assumptions for this approach are the same as those for any propensity score matching analysis (Guo and Fraser, 2009).

The computational steps in the effect estimation are presented in detail in the supplementary materials for the article¹.

2.4.5 Validity Test—We conducted a validity test for our inferential procedures, consisting of the propensity analysis and bootstrapping as described above. We tried as nearly as possible to emulate our procedures over similarly generated data sets. The major exception was that the simulated data were generated from a known model. Monte Carlo checks for the validity of results of complex procedures on a real dataset are often performed using the same design of the original data set. For an example, see the validity studies in Kloke et al. (2009: section 4.1) and Chang et al. (1999: section 8.2). The details of our Monte Carlo simulation are in the supplementary materials for the article².

3. RESULTS

3.1. Estimated Effect of AA on Drinking Outcomes

The effect estimates for the primary (PDA15) and secondary (OTCM15) drinking outcomes are shown in Table 1. Effect estimates that include “0” within the two-tailed 95% confidence intervals are considered “not significant” by convention.

The first column gives the mean unstandardized differences in outcomes between the latent subgroups derived through PSM. All effects were statistically significant. Consistent with our hypothesis, the Added vs. Potential differences indicate that subjects who consistently attended AA solely due to their exposure to TSF had better outcomes than equivalent subjects who did not consistently attend AA, but would have so attended, had they been

¹Supplementary material can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

²Supplementary material can be found by accessing the online version of this paper at <http://dx.doi.org> and by entering doi:...

exposed to TSF. Also, the Added vs. Expected differences in the TSF condition indicate that the additional subjects who were induced to attend AA consistently solely due to their exposure to TSF benefited more from AA than those who would ordinarily have attended AA even without exposure to TSF. To summarize, the Added AA group shows higher alcohol abstinence and lower (better) scores on the outcome composite than either the Expected or Potential AA group.

The third column in Table 1 gives a measure of the standardized effects (Cohen's d). According to Cohen (1988), we might term an effect size of 0.42 as low medium, 0.53 as medium and 0.70 or 0.71 as high medium. The interpretation of such an effect size is, for instance, that the means of the Added vs. Potential AA groups for PDA15 differ by 0.70 of the pooled standard deviation of the groups. The standardized effect sizes of the Added vs. Potential AA differences are larger than the effect sizes of the Added vs. Expected AA differences.

We also computed the results without using missing data imputation. All results were significant except for the Added vs. Expected AA effect for the outcome composite (OTCM15).

3.2. Validity Test Results

The results of the validity test for our inferential procedure are as follows. For each simulated data set, we ran the same bootstrap-propensity analysis as we did on the imputed data set. We obtained the bootstrap confidence intervals, as we did for the imputed data.

We ran 50 simulations with the number of bootstraps set at 100. The results are summarized in Table 2. The empirical confidences are the percentage of times the bootstrap confidence intervals trapped the true effect 0. We obtained empirical confidences for nominal 95% and 90% confidence intervals. We have placed coverages of the 90% confidence intervals in parentheses. Thus, the empirical confidences are quite close to the nominal confidences. So at least from this limited test, our inferential procedures are valid.

4. DISCUSSION

Randomized controlled trials of mutual aid, including AA, are notoriously difficult to conduct. Correlational studies of mutual aid, even if they have a longitudinal design, are problematic to interpret due to potential confounds in the analysis that are difficult to control, including alcohol problem severity, psychological problem severity, motivation for change, prior AA experience, prior treatment experience, and social support for recovery, as mentioned previously. The present study is a novel application of propensity score matching which capitalizes on the random assignment feature of a large alcoholism treatment dataset, Project MATCH. Although that study did not attempt to randomly assign subjects to AA vs. no AA, the TSF condition did result (as intended) in directing a higher proportion of subjects to community AA meetings than the other two treatment conditions. Since those "excess" AA attenders in TSF would not have "ordinarily" selected themselves for AA, the MATCH data presented an opportunity to determine the effect of AA attendance on drinking, which is not confounded by ordinary self-selection factors. The study described the logic of a novel analysis based on PSM that separated out the effect of AA attendance from AA self-selection confounds. The study hypothesis was supported by the results; there was evidence that consistent AA attendance over 15 months led to better drinking outcomes on two measures, abstinence from alcohol and less problem drinking, when the Added and Potential AA latent groups were compared. The results were equivalent for the analyses with and without missing data imputation.

The analysis capitalized statistically on the excess (“Added AA”) number of consistent AA attenders in TSF to help draw a conclusion about the effect of consistent AA attendance on drinking, but the conclusion is not limited to consistent AA attenders in TSF only. Traditional analysis might calculate the main and interaction effects of treatment type and AA attendance on outcomes, but this is susceptible to bias due to differential self-selection for AA. Instead, we constructed a novel comparison, an Added AA latent group in TSF vs. a Potential AA latent group in CBT/MET, that controls for confounds due to self-selection into AA. Consequently, our conclusion that consistent AA attendance was associated with improved outcomes is not specific to a particular treatment, despite the fact that the Added AA individuals were all in TSF. One can think of it this way: the different effect of the treatments on AA attendance allowed the construction of an unbiased group comparison to estimate the effect of consistent AA attendance on drinking.

We did not state a hypothesis about outcomes for the Added vs. Expected AA groups in TSF. The results with missing data imputation showed the Added AA group to have better outcomes on both measures, whereas without imputation the Added AA group had better PDA15 outcome only. We suggest that the results with imputation are to be preferred, because imputation uses all of the data, increasing potential generalizability of the results, and also increases statistical power.

As to why the Added AA individuals apparently benefited more from AA than the Expected AA individuals who ordinarily would participate in AA, we can only speculate, since of course these groups were retrospectively identified. For example, it may be that the Added AA individuals held negative preconceptions about AA that were changed for the better upon participation, leading to above-average enthusiasm for AA and its principles.

These results have direct relevance to the treatment of alcohol-dependent individuals. The findings indicate that referring clients to AA will benefit them irrespective of whether the clients have the typical characteristics of AA participants, which clinicians may regard as “appropriate” for AA. In addition, AA stresses the importance of consistent attendance to obtain maximum benefits; the present study provides clinicians with additional empirical evidence supporting this principle.

This analysis was not suited to determining the mechanisms through which AA affects drinking outcomes, but prior analyses with the Project MATCH dataset have given insights into the possible mediators of such effects. These mediators include reduction of depression symptoms (Kelly et al., 2010), decreases in pro-drinking and increases in pro-abstinent social network ties (Kelly et al., 2011a), increases in spiritual practices (Kelly et al., 2011b), and increases in self-efficacy (Connors et al., 2001). The influence of these mediators, originally examined separately in studies, was confirmed in a recent integrative analysis of multiple mediators (Kelly et al., 2012). The results of the present study, which strengthen the evidence for AA effects on drinking, increase the importance of this prior research that identified putative mediators of AA effects.

TSF was found to be effective in improving AA attendance. There is clearly a potential for TSF to exert a synergistic effect on outcomes beyond increasing attendance; further research involving TSF and AA should examine this question.

The study has certain limitations. The effects of AA were found in the context of a treatment platform, albeit the treatment only took place within the first three months of the 15-month study period. Nevertheless, we cannot conclude that exactly the same AA effects would be found without that initial treatment experience. The study had a large amount of baseline data from which to identify possible covariates for the PSM; not all studies may have this. The logic of the analysis only obtains within the context of a randomized controlled trial

(RCT) that establishes treatment or other intervention groups with average equivalent characteristics. The MATCH sample is predominantly white males; it is unknown to what extent the results may apply specifically to female and minority group alcoholics. The study analyzed OTCM15 as an interval scale, but technically it better fits the definition of an ordinal scale. However, some additional statistical groundwork would be required to merge ordinal scales into the PSM model as described; this could be an avenue for future work. Finally, caution must be exercised about making causal inferences. Although the results indicate that the relationship between consistent AA attendance and drinking outcomes is not spurious, the direction of causality (or the existence of reciprocal causality) cannot be definitely established by the technique employed. Replication of this technique on additional appropriate datasets would be an important next step.

The applicability of the analysis can be extended to other RCTs in other domains. The analysis is applicable whenever randomization to study conditions results in an “excess” of subjects within one condition who are exposed to a substantively interesting intervention; that is, when more subjects are exposed to the intervention than would be exposed outside that study condition. For example, we may be interested in the effects of college attendance on young people, but we can’t randomly assign students to college or no college; students ordinarily self-select. (The analogy with AA is clear.) However, in a community where college-going rates are relatively low, we may be able to randomly assign high school volunteers to a special college facilitation intervention vs. no facilitation (or some other enrichment program). Assuming there is more college-going in the college facilitation condition than in the comparison condition, the analysis described in this paper could be performed to determine the unconfounded effects of college attendance on selected outcomes.

5. CONCLUSION

Randomized controlled trials of mutual aid, including Alcoholics Anonymous (AA), are difficult to conduct and correlational studies are problematic to interpret due to potential confounds. This novel propensity score matching method indicates that consistent AA attendance improves drinking outcomes, independent of “normal” confounding factors that occur in non-randomized studies. Since AA stresses the importance of consistent attendance to maximize its benefits, the present study provides clinicians with additional empirical evidence in support of this principle. Further research utilizing our PSM approach with similar datasets in the drug/alcohol (Timko and DeBenedetti, 2007; Walitzer et al., 2009; Litt et al., 2009) and related behavioral research fields is indicated and could yield important insights into the effects of mutual aid and other interventions when randomized designs are difficult to implement.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Therawat Wisadrattanapong and Charles Cleland assisted in database management during the first phase of the study.

Role of Funding Source

Funded by grant no. R21AA017906 from the National Institute on Alcohol Abuse and Alcoholism. (NIAAA). NIAAA had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

REFERENCES

- Alcoholics Anonymous. [accessed on March 1, 2012] Estimates of A.A. groups and members. Jan 1. 2011 2011 http://www.aa.org/en_pdfs/smf-53_en.pdf
- Babor, TF.; Del Boca, FK. *Treatment Matching in Alcoholism*. Cambridge University Press; United Kingdom: 2003.
- Bogenschutz MP. Individual and contextual factors that influence AA affiliation and outcomes. *Rec. Dev. Alcohol*. 2008; 18:413–433.
- Brandsma, JM.; Maultby, MC.; Welsh, RJ. *Outpatient Treatment of Alcoholism: A Review and Comparative Study*. University Park Press; Baltimore, MD: 1980.
- Chang W, McKean JW, Naranjo JD, Sheather SJ. High breakdown rank-based regression. *J. Am. Stat. Assoc.* 1999; 94:205–219.
- Chappel JN, DuPont RL. Twelve-step and mutual-help programs for addictive disorders. *Psychiatr. Clin. North Am.* 1999; 22:425–446. [PubMed: 10385942]
- Christenfeld NJ, Sloan RP, Carroll D, Greenland S. Risk factors, confounding, and the illusion of statistical control. *Psychosom. Med.* 2004; 66:868–875. [PubMed: 15564351]
- Cisler RA, Zweben A. Development of a composite measure for assessing alcohol treatment outcome: operationalization and validation. *Alcohol. Clin. Exp. Res.* 1999; 23:263–271. [PubMed: 10069555]
- Cohen, J. *Statistical Power Analysis for the Behavioral Science*. 2nd Ed. Lawrence Erlbaum Associates; New York: 1988.
- Connors GJ, Tonigan JS, Miller WR, MATCH Research Group. A longitudinal model of intake symptomatology, AA participation and outcome: retrospective study of the Project MATCH outpatient and aftercare samples. *J. Stud. Alcohol*. 2001; 62:817–825. [PubMed: 11838919]
- Crumbaugh JC. The Seeking of Noetic Goals Test (SONG): a complementary scale to the Purpose in Life Test (PIL). *J. Clin. Psychol.* 1977; 33:900–907. [PubMed: 893732]
- 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; 124:64–67.
- Efron, B.; Tibshirani, RJ. *An Introduction to the Bootstrap*. Chapman and Hall; New York: 1993.
- Emrick, CD.; Tonigan, JS.; Montgomery, H.; Little, L. Alcoholics Anonymous: what is currently known?. In: McCrady, BS.; Miller, WR., editors. *Research on Alcoholics Anonymous: Opportunities and Alternatives*. Rutgers Center for Alcohol Studies; New Brunswick, NJ: 1993. p. 41-77.
- Fiorentine R, Hillhouse MP. Why extensive participation in treatment and twelve-step programs is associated with the cessation of addictive behaviors: an application of the addicted-self model of recovery. *J. Addict. Dis.* 2003; 22:35–55. [PubMed: 12661978]
- 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]
- Gossop M, Harris J, Best D, Man LH, Manning V, Marshall J, Strang J. Is attendance at Alcoholics Anonymous meetings after inpatient treatment related to improved outcomes? A 6-month follow-up study. *Alcohol Alcohol*. 2003; 38:421–426. [PubMed: 12915517]
- Guo, SY.; Fraser, MW. *Propensity Score Analysis: Statistical Methods and Applications*. Sage; Los Angeles: 2009.
- Hasin D, Grant B. AA and other help seeking for alcohol problems: former drinkers in the U.S. general population. *J. Subst. Abuse*. 1995; 7:281–292. [PubMed: 8749788]
- Hogg, RV.; McKean, JW.; Craig, AT. *Introduction to Mathematical Statistics*. 7th Ed. Pearson-Hall; New Jersey: 2012.
- Honaker, J.; King, G.; Blackwell, M. [accessed on March 1, 2012] AMELIA II: A Program for Missing Data, CRAN. 2010. <http://cran.r-project.org/web/packages/Amelia/Amelia.pdf>
- Humphreys K, Mavis B, Stoffleymayr B. 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. *Eval. Prog. Plan.* 1996; 19:301–308.
- Humphreys, K. *Circles of Recovery*. Cambridge University Press; United Kingdom: 2004.
- Kaskutas LA, Ammon L, Delucchi K, Room R, Bond J, Weisner C. Alcoholics anonymous careers: patterns of AA involvement five years after treatment entry. *Alcohol. Clin. Exp. Res.* 2005; 29:1983–1990. [PubMed: 16340455]
- Kaskutas LA, Bond J, Humphreys K. Social networks as mediators of the effect of Alcoholics Anonymous. *Addiction.* 2002; 97:891–900. [PubMed: 12133128]
- Kelly JF, Stout RL, Magill M, Tonigan JS, Pagano ME. Mechanisms of behavior change in alcoholics anonymous: does Alcoholics Anonymous lead to better alcohol use outcomes by reducing depression symptoms? *Addiction.* 2010; 105:626–636. [PubMed: 20102345]
- Kelly JF, Stout RL, Magill M, Tonigan JS, Pagano ME. Spirituality in recovery: a lagged mediational analysis of alcoholics anonymous' principal theoretical mechanism of behavior change. *Alcohol. Clin. Exp. Res.* 2011a; 35:454–463. [PubMed: 21158876]
- Kelly JF, Stout RL, Magill M, Tonigan JS. The role of Alcoholics Anonymous in mobilizing adaptive social network changes: a prospective lagged mediational analysis. *Drug Alcohol Depend.* 2011b; 114:119–126. [PubMed: 21035276]
- Kelly JF, Hoepfner B, Stout RL, Pagano M. Determining the relative importance of the mechanisms of behavior change within Alcoholics Anonymous: a multiple mediator analysis. *Addiction.* 2012; 107:289–299. [PubMed: 21917054]
- Litt MD, Kadden RM, Kabela-Cormier E, Petry NM. Changing network support for drinking: network support project 2-year follow-up. *J. Consult. Clin. Psychol.* 2009; 77:229–242. [PubMed: 19309183]
- Kloke J, McKean JW, Rashid M. Rank-based estimation and associated inferences for linear models with cluster correlated errors. *J. Am. Stat. Assoc.* 2009; 104:384–390.
- Miller WR, Del Boca FK. Measurement of drinking behavior using the Form 90 family of instruments. *J. Stud. Alcohol Suppl.* 1994; 12:112–118. [PubMed: 7722987]
- Miller, WR.; Tonigan, JS.; Longabaugh, R. (DrInC). Vol. 4: Project MATCH Monograph Series. National Institute on Alcohol Abuse and Alcoholism; Rockville, MD: 1995. The Drinker Inventory of Consequences. NIH Publication No. 95-3911
- Morgenstern J, Labouvie E, McCrady BS, Kahler CW, Frey RM. Affiliation with Alcoholics Anonymous after treatment: a study of its therapeutic effects and mechanisms of action. *J. Consult. Clin. Psychol.* 1997; 65:768–777. [PubMed: 9337496]
- Morgenstern J, Bux DA, 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]
- Project MATCH Research Group. Project MATCH: rationale and methods for a multisite clinical trial matching patients to alcoholism treatment. *Alcohol. Clin. Exp. Res.* 1993; 17:1130–1145. [PubMed: 8116822]
- Project MATCH Research Group. Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *J. Stud. Alcohol.* 1997; 58:7–29. [PubMed: 8979210]
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983; 70:41–55.
- Schuckit MA, Tipp JE, Smith TL, Bucholz KK. Periods of abstinence following the onset of alcohol dependence in 1,853 men and women. *J. Stud. Alcohol.* 1997; 58:581–589. [PubMed: 9391917]
- Substance Abuse and Mental Health Services Administration. [accessed on August 7, 2012] Results from the 2010 national survey on drug use and health: summary of national findings. 2011. <http://www.samhsa.gov/data/NSDUH/2k10NSDUH/2k10Results.htm>
- 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]
- Tonigan JS, Toscova R, Miller WR. Meta-analysis of the literature on Alcoholics Anonymous: sample and study characteristics moderate findings. *J. Stud. Alcohol.* 1996; 57:65–72. [PubMed: 8747503]

- Tonigan, JC.; Conners, GJ.; Miller, WR. Participation and involvement in Alcoholics Anonymous. In: Babor, TF.; Del Boca, FK., editors. *Treatment Matching in Alcoholism*. Cambridge University Press; United Kingdom: 2003.
- Walitzer KS, Dermen KH, Barrick C. Facilitating involvement in Alcoholics Anonymous during out-patient treatment: a randomized clinical trial. *Addiction*. 2009; 104:391–401. [PubMed: 19207347]
- Walsh D, Hingson R, Merrigan D. A randomized trial of treatment option for alcohol-abusing workers. *New England J. Med*. 1991; 325:775–782. [PubMed: 1870651]
- Ye Y, Kaskutas LA. Using propensity scores to adjust for selection bias when assessing the effectiveness of Alcoholics Anonymous in observational studies. *Drug Alcohol Depend*. 2009; 104:56–64. [PubMed: 19457623]
- Zweben A, Cisler RA. Clinical and methodological utility of a composite outcome measure for alcohol treatment research. *Alcohol. Clin. Exp. Res*. 2003; 10:1680–1685. [PubMed: 14574241]

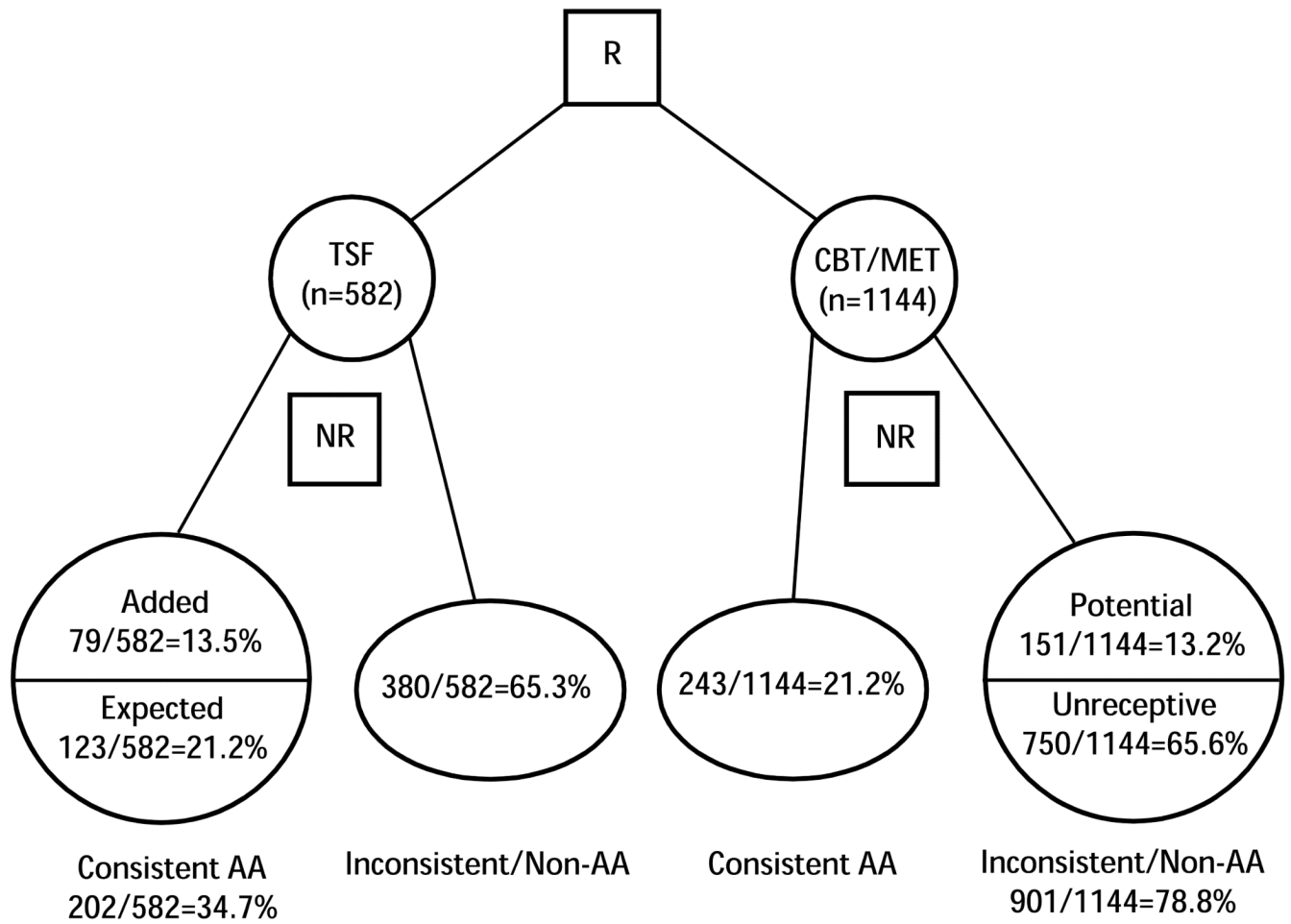


Figure 1.
Structure of Latent AA-related Subgroups in Project MATCH

Table 1
Drinking Outcome Effect Estimates and 95% Bootstrap Confidence Intervals (CIs)

	Unstandardized Effect (CI)	Standardized Effect (Cohen's d)
Outcome-PDA 15		
Added vs. Expected	0.075 (0.023, 0.126)	.42
Added vs. Potential	0.324 (0.198, 0.450)	.70
Outcome-OTCM15		
Added vs. Expected	-0.364 (-0.561, -0.167)	.53
Added vs. Potential	-1.046 (-1.447, -0.645)	.71

Bootstrap Simulation Results**Table 2**

	Confidence Interval (CI)	
	Added vs. Expected	Added vs. Potential
Nominal	95% (90%)	95% (90%)
Empirical	95% (88%)	98% (96%)