Identifying Mechanisms of Behavior Change in Psychosocial Alcohol Treatment Trials: Improving the Quality of Evidence from Mediation Analyses

JOHN W. FINNEY, PH.D.^{a,*}

^a*Center for Innovation to Implementation, Veterans Affairs Palo Alto Health Care System, Menlo Park, California*

ABSTRACT. Objective: In recent years, a shift has occurred from identifying variables that mediate the effects of psychosocial interventions for problematic alcohol use to trying to identify effective mechanisms of behavior change (MOBCs) that lead to better drinking-related and other outcomes. Although implementing interventions targeting specific MOBCs has considerable conceptual and intuitive appeal, it is important that strong empirical evidence be available that supports such implementation. One aim of this article is to consider some prior mediational analyses of the effects of alcohol-focused psychosocial treatments to illustrate findings and the types of evidence that typically have been presented. A second aim is to consider methodological approaches, specifically certain statistical analyses of observational data and experimental designs, that could provide more rigorous evidence regarding the causal effects of mechanisms of change (mediators) in

producing desired drinking-related outcomes. **Results:** Prior analyses that have examined the association of a mediator with an outcome, or the relationship between a mediator and an outcome with assigned treatment controlled, do not provide strong evidence regarding the causal effects of mediators on outcomes. Unfortunately, experimental designs that directly manipulate mediators seem difficult to apply in the alcohol treatment field as an approach to examine mediator/MOBC effects. **Conclusion:** To produce more compelling evidence, we need to investigate interventions that target a specific MOBC and then capitalize on random assignment to intervention/control conditions by using principal stratification or instrumental variable analyses to identify the effects of intervention-induced mediator change on outcomes. (*J. Stud. Alcohol Drugs*, 79, 163–170, 2018)

FOR MORE THAN TWO DECADES, identifying mediators that could account for how alcohol (and other drug) use disorder treatments exert their effects has been a focus of considerable attention in trials of psychosocial (and pharmacological) treatments (e.g., DiClemente et al., 1994; Finney, 1995; Morgenstern et al., 1996, 1997). More recently, a shift has occurred from identifying variables that mediate the effects of psychosocial interventions to trying to identify effective mechanisms of behavior change (MOBCs) that lead to better drinking-related and other outcomes. The idea is that more potent interventions then could be designed that produce even greater increases in effective mechanisms of change, thereby leading to even better drinking-related outcomes. An early proponent of targeting processes of change was Moos (2007): “Rather than focusing so heavily on understanding specific types or orientations of treatment, such as CBT [cognitive-behavioral therapy] or TSF [twelve-step facilitation], training should emphasize common treatment processes, such as promoting support, goal direction, and structure in treatment and clients’ life contexts, enhancing clients’ involvement in new rewarding activities, and building their self-efficacy and coping skills” (p. 119).

Implementing effective interventions that target specific processes of change has considerable appeal. For example,

in terms of conceptual frameworks of implementation research (e.g., Rogers, 2003), implementing MOBC-targeted interventions in routine care rates highly on adaptability, trialability, and lower complexity. In other words, new treatment approaches are more likely to be implemented by providers and treatment organizations if they can be adapted to local circumstances, can be tried out, and are less complex. Moreover, it should be easier to train providers in specific interventions, rather than more complicated, multifaceted treatment approaches, and providers should be more likely to use targeted MOBC interventions that could be integrated with, rather than supplant, their current treatment approaches (Gifford et al., 2012).

Although implementing MOBCs has considerable conceptual and intuitive appeal, it is important to consider the supporting empirical evidence before contemplating implementation. This article focuses on evidence from mediational analyses of the effects of psychosocial treatments for alcohol use disorders and misuse. Its first aim is to consider the nature and results of selected, but typical, analyses regarding the effectiveness of specific mechanisms of change. The second aim is to consider statistical/methodological approaches that would yield findings in which one could have greater confidence that causal effects of MOBCs on outcomes have been isolated.

Simple mediational model

Much of the existing work on mediators of alcohol treatment effects has been guided by the basic mediational model

Received: March 16, 2017. Revision: July 21, 2017.

*Correspondence may be sent to John W. Finney at the Center for Innovation to Implementation, Veterans Affairs Palo Alto Health Care System (152MPD), 795 Willow Road, Menlo Park, CA 94025, or via email at: jwf_448@yahoo.com.

FIGURE 1. The basic mediational model. *Note:* Indirect effect of T on O via M = ab ; direct effect of T on O = c' = $(c - ab)$; total effect of T (c) on O = indirect effect (ab) + direct effect (c')

depicted in Figure 1 that was popularized by Baron and Kenny (1986) and analyzed via three equations, as noted by MacKinnon et al. (2007):

$$O = i_1 + cT + e_1 \quad [1]$$

$$M = i_2 + aT + e_2 \quad [2]$$

$$O = i_3 + cT + bM + e_3 \quad [3]$$

In this model, regardless of whether a treatment (T) has a “total” effect (see coefficient c in Equation 1 and path c in Figure 1) on the outcome (O), analyses examine the effect of treatment on the mediator (M; see coefficient a in Equation 2 and path a in Figure 1) and the effect of the mediator on an outcome with treatment controlled (see coefficient b in Equation 3 and path b in Figure 1). Multiplying the a and b paths (ab) provides one estimate of the “indirect effect” of treatment on the outcome via the mediator. After the mediator is taken into account, any remaining effect of treatment on the outcome is referred to as its “direct effect” (see c' in Equation 3 and path c' in Figure 1). Although studies have examined (a) multiple mediators (e.g., Kelly et al., 2012), (b) whether mediational paths are moderated by other variables (e.g., Kelly & Hoepfner, 2013; Longabaugh & Wirtz, 2001), and (c) whether the positive effects of some mediational pathways are suppressed by others (e.g., MacKinnon et al., 2000; Wirtz, 2007), as well as other issues, the simple mediational model has been predominant in the alcohol treatment research field.

Obviously, full mediational analyses (Equations 1–3) address the indirect effects of a treatment via a mediator, but what is the value of focusing separately on the a and b paths? If the total effect of treatment is not significant, examining those paths can allow one to determine where the treatment process may have broken down (Chen, 1990; Finney, 1995; MacKinnon et al., 2002; Suchman, 1967). If the a path is not significant, it represents what Suchman

(1967) referred to as a “program failure” (the treatment was not successful in effecting the desired change on the mediator). If the b path is not significant, it represents what Suchman (1967) labeled a “theory failure” (i.e., the mediator does not affect the outcome in the way in which the theory underlying the intervention assumed that it would). Subsequently and more generally, Chen (1990) argued that the a path addresses an “action theory” and the b path addresses a “conceptual theory” indicating how the mediator is supposed to affect the outcome. The relevant point is that, in analyses to identify effective MOBCs, the b path from the mediator to the outcome is of particular interest.

Mediator–outcome relationships in alcohol treatment studies

A comprehensive review of relationships found between mediators and outcomes in psychosocial alcohol treatment research is beyond the scope of this article. However, some reviews and individual studies are considered here to illustrate analyses that frequently have been used to identify mediator–outcome relationships and the substantive findings they have generated. For example, Morgenstern and Longabaugh (2000) reviewed five studies of CBT that examined the relationship of coping skills to drinking-related outcomes. Analyses included partial, semipartial, and bivariate correlation analyses, and other multiple regression analyses. Of the 11 analyses of b paths reported by Morgenstern and Longabaugh (2000), 9 found that stronger coping skills were significantly related to better alcohol-related outcomes.

In a review of research on motivational interviewing for substance use disorders, Apodaca and Longabaugh (2009) identified four studies that examined one, and one study that examined two, of four potential motivational interviewing mechanisms of change: client change talk/intention, client involvement/engagement, client resistance, and client experi-

ence of discrepancy between goals/values and substance use behavior. Across studies, the b path, as addressed in various regression analyses (including analyses of variance and partial correlation analyses), had an average r of .23 (see Figure 1 in Apodaca & Longabaugh, 2009).

Rather than focusing on treatment condition in their review of motivational interviewing mechanisms of change, Magill and her colleagues (2014) examined linkages of therapists' motivational interviewing—consistent and -inconsistent behaviors with client behaviors, and then the relationship between during-treatment client behaviors and alcohol or other addiction outcomes at follow-up (i.e., the b path). Across studies using correlations (including Kendall's tau- b) and regression (including binomial and multilevel) analyses, the average b path was $r = .06$ for client change talk, $-.24$ for client sustain talk, and $.12$ for those studies using a composite of change and sustain talk. Although none of the studies included in the first two reviews above controlled for treatment condition in examining mediator–outcome relationships (see Equation 3), that approach has been used in other studies (e.g., Kiluk et al., 2010; Mensinger et al., 2007).

These reviews and studies certainly are not exhaustive, but their findings might seem to provide some positive evidence that such variables as increased coping skills, reduced sustain talk, etc., are effective mechanisms of change with respect to improving alcohol and other outcomes. Before accepting that conclusion, however, it is useful to consider the assumptions underlying mediational analyses.

Assumptions of the basic mediational model

The relevant assumptions were described cogently in an article by Jo (2008). The “constant effect” assumption is that there is no interaction between treatment and mediator in relation to the outcome, which should be tested by including a Treatment $\times$ Mediator product term in mediational analyses (Kraemer et al., 2002). Another assumption is that the relationship between the mediator and outcome is linear—in other words, that outcome scores are linearly higher or lower as mediator scores are higher or lower. This assumption needs to be met if the ab path is to be interpreted as the indirect effect of treatment on outcome via the mediator in full mediational analyses (Imai et al., 2010; Jo, 2008).

Most critical, however, is the assumption implied by the single-headed arrows in Figure 1 that the a and b paths represent causal effects. With respect to the a path, with a sufficiently large sample of individuals randomly assigned to more than one treatment/control condition, one can have reasonable confidence that it estimates the causal effect of treatment on the mediator. However, with respect to the b path, individuals typically are not randomly assigned to different levels of the mediator, so the mediator–outcome relationships that have been examined in alcohol treatment

research usually have been correlational, not causal (Jo, 2008).

To interpret b path results from correlational analyses as causal, one needs to subscribe to the third assumption outlined by Jo (2008)—the “ignorability of the observed mediator status” (p. 317). In other words, simple mediational analyses assume that individuals with the same status on the mediator have the same observed and unobserved characteristics, on average, regardless of treatment/control condition. In reality, individuals who improve on the mediator in the treatment condition may have different observed and unobserved characteristics than those who show mediator improvement in the control condition, even if they ultimately have the same score on the mediator (Jo, 2008). In this regard, Bullock et al. (2010) note that “Baron-Kenny estimates are prone not only to bias but to bias of the sort that overstates the extent of mediation” (p. 551). Why? Because other variables that affect the mediator also are likely to affect the outcome in the same direction.

What clearly is most relevant in evaluating if a b path analysis provides evidence for targeting a particular MOBC is whether the analysis was likely to have estimated the causal effect of the mediator on the outcome. In treatment research, the causal effect of interest is the effect on the outcome of treatment- or intervention-induced change on the mediator. Unfortunately, b path findings from most existing research do not address this effect. If a bivariate correlation or partial regression coefficient between the mediator and outcome is reported, it presumably embodies the treatment-induced effect plus the effects of other, noncontrolled confounders of mediator status and outcomes (Bullock et al., 2010; Jo, 2008). More problematic is that, although the a coefficient of the ab product in full mediational analyses of randomized controlled trial (RCT) data captures treatment-induced change on a mediator, the isolated b coefficient does not. As indicated in Equation 3, treatment condition is statistically controlled in the calculation of the b coefficient, thus rendering the mediator–outcome relationship independent of any treatment effect on the mediator. To generate compelling evidence supporting MOBCs, we need to estimate more rigorously the causal effects of treatment-induced mediator change on outcomes.

Identifying causal effects of mediators on outcomes

Causal inferences regarding mediators' effects on outcomes can be strengthened in a variety of ways (Kazdin & Nock, 2003). Here, however, the focus is on increasing confidence in causal inferences through (a) statistical techniques applied to observational (nonexperimental) data on mediator–outcome relationships or (b) experiments that manipulate mediators.

Statistical approaches. MacKinnon and Pirlott (2015) discussed four approaches to the statistical analysis of non-

experimental data that, when certain assumptions are met, can provide more confidence that the causal effects of mediators have been addressed. They include (a) comprehensive structural equation modeling (SEM), (b) inverse probability weighting analysis, (c) principal stratification analysis, and (d) instrumental variable models.

With comprehensive SEM, the idea is to control for as many relevant covariates as were assessed, with measurement error taken into account (Loehlin, 2017). For example, using SEM, Mason et al. (2009) controlled for age, alcohol abuse risk, and assignment to a Preparing for the Drug Free Years (PDFY) or control intervention. They concluded that increases in prosocial skills mediated the effect of PDFY on the reduced risk of an alcohol abuse diagnosis 11 years later for young women. Critical assumptions for SEM analyses are that all confounders—variables associated with the mediator and outcome—are in the model and that, if multiple mediators (each could be considered a potential “posttreatment confounder” of a focal mediator—Coffman & Zhong, 2012) are included, measures of all mediators are of equal quality. If all confounders are included in the analysis (a very strong assumption), the causal effect of the focal mediator would be isolated. Having a sufficiently large sample size for structural equation analyses often can be another substantial challenge (Wolf et al., 2013).

The inverse probability weighting approach controls for as many relevant covariates as were assessed but uses them to predict the mediator, not the outcome (Coffman & Zhong, 2012). To weight individuals on the extent to which their mediator scores are confounded (i.e., predicted by factors other than assignment to treatment/control condition), the results of two analyses are used to form a ratio. The first predicts the mediator only from treatment assignment, with the predicted value forming the numerator of the ratio. The second predicts the mediator from both treatment assignment and observed covariates, with the predicted value used as the denominator of the ratio.

For dichotomous mediators, probabilities are calculated for the predicted values of the numerator and denominator; for continuous mediators, conditional densities of the predicted values are calculated from a normal probability density function (see VanderWeele et al., 2011; Robins et al., 2000; Thoemmes & Ong, 2016). A regression analysis using individuals' inversely weighted scores on the mediator then predicts the outcome (for more details, see Coffman & Zhong, 2012; Thoemmes & Ong, 2016). If all relevant covariates have been included in the analysis (again, a very strong assumption), a causal effect of the mediator on the outcome is identified.

The principal stratification method (Gallop et al., 2009; Jo, 2008; Rubin, 2004) is most easily explained by considering a dichotomous response on a potential mediator, as described by Jo (2008; see Imai et al., 2010, for a discussion regarding continuous mediators). Given either “improve-

ment” or no improvement on a dichotomous mediator, four hypothetical types of participants are assumed. As labeled by Jo (2008), “never-changers” experience no improvement on the mediator in either the treatment or control condition. “Forward-changers” experience mediator improvement only in the experimental condition; “backward-changers” exhibit mediator improvement only in the control condition. Last, “always-changers” show improvement on the mediator in either the experimental or control condition. The principal stratification approach assumes that there are no backward-changers (individuals who show a positive mediator response only in the control condition). A second critical assumption, labeled the “exclusion restriction” assumption, is that the entire effect of treatment is transmitted through the mediator (i.e., that the direct effect of treatment is zero). Thus, the intervention to affect a mediator must be targeted quite narrowly.

The difference between the proportions of changers (always- and forward-changers) in the treatment and control (always-changers only) condition is assumed to reflect the proportion of individuals experiencing the desired change on the mediator as a result of the treatment (forward-changers). The difference in outcome for the two conditions is multiplied by the proportion of forward-changers to estimate the causally mediated treatment effect. For example, say that .50 of the participants in a treatment condition are abstinent at follow-up versus .20 in the control group. If the proportion of changers (always- and forward-changers) in the treatment group is .85, and the proportion of (always) changers in the control group is .15, then $.70 (.85 - .15)$ of the .30 difference in the abstinence rate between the treatment and control groups is assumed to be indirect via the mediator (MacKinnon & Pirlott, 2015).

An instrumental variable typically is thought of as one that mimics random assignment (Angrist & Krueger, 2001). Thus, random assignment is an ideal instrument to use in isolating the causal effect of mediators on outcomes. Participants' randomly assigned intervention/control condition is used to predict (a) the mediator and (b) the outcome (Winship & Morgan, 1999). Then, the predicted outcome from the second analysis is regressed on the predicted mediator from the first analysis, sometimes with covariates included (alternatively, a simultaneous, rather than two-step, analysis could be conducted). As diagrammed in Figure 2, the assumptions for an instrumental variable analysis (Newhouse & McClellan, 1998) are that (a) the instrument (random assignment to intervention/control condition) is strongly associated with the mediator, (b) the entire effect of the treatment assignment on the outcome is transmitted through the mediator (the “exclusion restriction” assumption), and (c) the instrument is unrelated to any other factors affecting the outcome (a tenable assumption with random assignment of a large number of participants).

An example of this approach is the research by Humphreys et al. (2014), who used instrumental variable

FIGURE 2. Assumptions for instrumental variable (IV) analyses. *Note:* Assumptions of this type of IV analysis are that (a) the IV is strongly associated with the mediator, (b) the effect of the IV on the outcome is solely through the mediator, and (c) the IV is associated with no variable affecting the outcome other than the mediator.

analyses to try to isolate the causal effect of Alcoholics Anonymous (AA) participation across five randomized AA facilitation trials. AA participation had a significant effect on drinking-related outcomes in five of the six analyses they conducted (one study provided two data sets). Drawbacks to the instrumental variable approach (Winship & Morgan, 1999) include the inability to test the exclusion restriction (e.g., that AA facilitation was narrowly focused and affected drinking-related outcomes only via greater AA participation). Also, instrumental variable analyses will produce large standard errors of estimates if the instrument has a weak relationship to the mediator or if small samples are used. Finally, Winship and Morgan (1999) note that an accurate estimate of the average effect of a mediator will be generated only if the mediator effect is constant across all research participants.

These weaknesses notwithstanding, principal stratification and instrumental variable analyses seem the most promising of the four approaches considered by MacKinnon and Pirlott (2015) to isolate the causal effects of MOBCs. For the structural equation and inverse probability weighting approaches to yield unbiased causal estimates of mediator effects, all relevant confounders/covariates must be included in the analyses. In actuality, it is difficult to identify, much less assess, all such variables. In contrast, the principal stratification and instrumental variable approaches do not rely on that difficult-to-meet assumption and capitalize on random assignment to the intervention/control condition to yield stronger causal inferences. However, other assumptions must be met, including having an intervention so targeted on a single mediator that the total effect of the intervention is indirect via that mediator.

Experimental approaches. Discussions (e.g., MacKinnon et al., 2007; MacKinnon & Pirlott, 2015) sometimes raise the possibility that mediational analyses, including evaluating the causal effect of a mediator on an outcome, can use experimental designs. As will be seen, however, these designs seem difficult to apply in alcohol treatment research.

The article by Spencer et al. (2005) in the social psychology field often is cited to illustrate two approaches to mediational analyses. One, the experimental-causal-chain design, examines in sequential experiments a mediator as (a) an effect of a treatment or an independent variable and then as (b) a cause of an outcome or a dependent variable. To illustrate this design, Spencer et al. (2005) described research by Word et al. (1974) that found that (a) the nonverbal behavior of the White “interviewer” participants was more distant (e.g., involved less immediacy) toward randomly assigned African American than toward White confederate “interviewees” and that (b) White participant “interviewees” in a second study who, at random, were responded to as the African American interviewees, as opposed to the White interviewees, had been in the first study performed worse in an interview. The results were interpreted as supporting the hypothesis that nonverbal behavior accounts for the relationship between held stereotypes and stereotype-consistent behavior in stereotyped others.

The moderation-of-process design involves manipulating the mediator as a moderator of the relationship between a treatment or an independent variable and an outcome or a dependent variable. To illustrate this design, Spencer et al. (2005) discussed a study by Zanna and Cooper (1974) that addressed whether cognitive dissonance arising from inconsistent attitudes and behaviors leads to attitude change

by producing aversive emotional arousal. Participants were randomly assigned to receive differing information about a placebo pill's effects. Largely because attitude change did not occur in the face of inconsistent behavior for participants who were told that the pill would make them tense or anxious, the results were interpreted as support for internal conflict between attitudes and behavior leading to attitude change via emotional arousal.

Additional experimental designs have been suggested by Imai and colleagues (2011, 2013). In their parallel design, participants are divided into two groups, with the members of one randomly assigned to either a treatment or control condition. The other group is randomly assigned to not only a treatment or control condition but also to different mediator levels. The example provided by Imai et al. (2013) involves the "ultimatum game" in which a "proposer" makes an offer to a "responder" to divide an amount of money (e.g., proposer keeps \$15, responder receives \$5). If the responder refuses the proposer's offers, both parties receive nothing. The proportion of "unfair" offers accepted is the outcome variable.

Based on prior relevant research by Knoch et al. (2006), Imai et al. (2013) suggested that a study could be conducted with one group randomly assigned to receive a certain proportion of unfair offers (to estimate the total effect of the "treatment") and a second group randomly assigned to receive the same proportion of unfair offers, but also randomly assigned to right prefrontal cortex deactivation with transcranial magnetic stimulation (TMS) or no deactivation with sham TMS. If participants randomly assigned to deactivation accepted more unfair offers than those without TMS deactivation, it would support right dorsolateral prefrontal cortex processes as mediating the relationship between receiving fair/unfair offers and accepting or rejecting them. Note that this example is similar to the moderation-of-process design discussed by Spencer et al. (2005).

Finally, in Imai and colleagues' (2013) crossover design, participants in the first phase are randomly assigned to a treatment or a comparison condition and are subsequently assessed on the mediator and later on the outcome. In the second phase, each participant is assigned to the treatment or the comparison condition he or she did not receive in Phase 1 and to the same mediator level assessed in Phase 1. Thus, the approach is to estimate the direct effect of the sequentially assigned treatments while holding the mediator constant. Drawing on research by Bertrand and Mullainathan (2004), Imai et al. (2013) suggest a hypothetical study involving the relationship between race and discrimination in hiring, as indexed by callbacks for interviews. Résumés of African Americans could be sent to potential employers to determine how many received callbacks. Later, the same résumés could be sent to the same employers with "White-sounding" names, manipulating the independent variable, but holding constant the potential mediator of qualifications. Any

difference in callback rates could be attributed to the "direct effect" of race, given the absence of an indirect effect via variation in qualifications.

One crucial and challenging assumption of this design is no carryover effect into Phase 2 of the condition participants received in Phase 1. In addition, the crossover design, like the parallel design, assumes that the effect of a mediator intervention on outcomes occurs only via the mediator—that is, there is no direct effect of the mediator manipulation, including participants' being unaware of the manipulation (Imai et al., 2013). Imai and colleagues also acknowledge that direct manipulation of mediators can be difficult; therefore, they propose additional parallel and crossover designs that "encourage" participants to adopt a mediator status—for example, to encourage anxiety by having research participants write about a situation that previously caused them anxiety. Note that the varying information about the effects of a placebo pill in the study of Zanna and Cooper (1974) represents what Imai et al. (2013) labeled "encouragement" to adopt a particular mediator status.

The examples provided for the experimental designs with direct mediator manipulation illustrate why such designs seem difficult to apply in alcohol (and other addiction) treatment research. All of them focus on very short-term changes on a mediator (e.g., behavior toward a stereotyped other in an interview, short-term brain region deactivation) and short-term outcomes (e.g., behavior of an interviewee, making a decision about an unfair offer). It is quite different, if not impossible, to randomly assign persons directly to different levels of the types of psychosocial mediators (e.g., sense of self-efficacy, coping skills, social support) that typically have been examined in psychosocial alcohol treatment research and, moreover, in a way that would be sufficiently enduring to have an effect on a more deeply ingrained condition, such as alcohol dependence.

With the exception of the crossover approach, the experimental designs considered above are for conducting full mediational analyses—that is, deriving unbiased or less biased estimates of indirect and direct effects of a treatment or an independent variable on an outcome or a dependent variable. With respect to the mediator–outcome effect component (path *b*), the key feature of their designs is a targeted intervention to affect a mediator, either directly or via encouragement. Thus, when encouragement rather than direct manipulation is adopted, the mediator manipulation is the same as the targeted interventions to affect a single mediator discussed earlier in the context of principal stratification and instrumental variable analyses.

Discussion

Clearly, more focus is needed on the causal effects of mediators on outcomes if we are to identify effective MOBCs from mediational analyses. To do so, we need to investigate

interventions that target a specific MOBC, such as coping skills. This recommendation is in stark contrast to that of Hodgson (2000):

[W]e must be asking the broad question: “How does this treatment lead to change?” rather than more simply “Does this motivational intervention change motivation?” or “Does this coping treatment result in changes in coping skills?” (p. 1716).

Hodgson’s advice is consistent with the fact that many psychosocial alcohol treatments are multifaceted. To understand how those treatments work, we need to study their multiple mediators. Attending to interventions that target a specific MOBC obviously is more narrowly focused. Which approach should be used? In fact, both approaches can produce valuable new knowledge.

This conclusion is consistent with the National Institute on Alcohol Abuse and Alcoholism’s (2013) Program Announcement PAR 14-051 that calls for studies that focus on specific mechanisms of change in relation to behavioral treatment outcomes, as well as for studies that simultaneously compare and test multiple mechanisms to determine their relative importance. However, with respect to studying specific MOBCs, the PAR recommends an experimental approach “in which the proposed mechanism of behavior change is directly manipulated or varied across groups within or across specific behavioral treatments.”

For the reasons considered earlier, experimental designs to isolate mediator effects via direct manipulation seem extremely difficult to apply in alcohol treatment research. Nevertheless, one cannot rule out the possibility that at some point experimental studies might be able “to identify causal mechanisms experimentally through clever manipulations and future technological developments” (Imai et al., 2013, p. 26). In the meantime, principal stratification or instrumental variable analyses of data from RCTs of targeted “encouragement” interventions seem likely to have broader application. Such analyses can provide stronger findings on the causal effects of specific MOBC interventions that may support their implementation in alcohol-focused care.

Acknowledgments

This article is based on a presentation by the author on “Implementing Evidenced-based Mechanisms of Behavior Change in Alcohol-focused Care” that was given at the 15th Annual Mechanisms of Behavior Change Conference in San Antonio, Texas, June 20, 2015. I thank Ann Batiza for her helpful comments on an earlier draft of this article.

References

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural ex-

- periments. *Journal of Economic Perspectives*, 15, 69–85. doi:10.1257/jep.15.4.69
- Apodaca, T. R., & Longabaugh, R. (2009). Mechanisms of change in motivational interviewing: A review and preliminary evaluation of the evidence. *Addiction*, 104, 705–715. doi:10.1111/j.1360-0443.2009.02527.x
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173–1182. doi:10.1037/0022-3514.51.6.1173
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94, 991–1013. doi:10.1257/0002828042002561
- Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what’s the mechanism? (don’t expect an easy answer). *Journal of Personality and Social Psychology*, 98, 550–558. doi:10.1037/a0018933
- Chen, H.-T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Coffman, D. L., & Zhong, W. (2012). Assessing mediation using marginal structural models in the presence of confounding and moderation. *Psychological Methods*, 17, 642–664. doi:10.1037/a0029311
- DiClemente, C. C., Carroll, K. M., Connors, G. J., & Kadden, R. M. (1994). Process assessment in treatment matching research. *Journal of Studies on Alcohol, Supplement 12*, 156–162. doi:10.15288/jsas.1994.s12.156
- Finney, J. W. (1995). Enhancing substance abuse treatment evaluations: Examining mediators and moderators of treatment effects. *Journal of Substance Abuse*, 7, 135–150. doi:10.1016/0899-3289(95)90310-0
- Gallop, R., Small, D. S., Lin, J. Y., Elliott, M. R., Joffe, M., & Ten Have, T. R. (2009). Mediation analysis with principal stratification. *Statistics in Medicine*, 28, 1108–1130. doi:10.1002/sim.3533
- Gifford, E. V., Tavakoli, S., Weingardt, K. R., Finney, J. W., Pierson, H. M., Rosen, C. S., . . . Curran, G. M. (2012). How do components of evidence-based psychological treatment cluster in practice? A survey and cluster analysis. *Journal of Substance Abuse Treatment*, 42, 45–55. doi:10.1016/j.jsat.2011.07.008
- Hodgson, R. (2000). Coping skills and Tiger Woods. *Addiction*, 95, 1716–1717.
- Humphreys, K., Blodgett, J. C., & Wagner, T. H. (2014). Estimating the efficacy of Alcoholics Anonymous without self-selection bias: An instrumental variables re-analysis of randomized clinical trials. *Alcoholism: Clinical and Experimental Research*, 38, 2688–2694. doi:10.1111/acer.12557
- Imai, K., Jo, B., & Stuart, E. A. (2011). Commentary: Using potential outcomes to understand causal mediation analysis. *Multivariate Behavioral Research*, 46, 861–873. doi:10.1080/00273171.2011.606743
- Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15, 309–334. doi:10.1037/a0020761
- Imai, K., Tingley, D., & Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176, 5–51. doi:10.1111/j.1467-985X.2012.01032.x
- Jo, B. (2008). Causal inference in randomized experiments with mediational processes. *Psychological Methods*, 13, 314–336. doi:10.1037/a0014207
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 44, 1116–1129. doi:10.1111/1469-7610.00195
- Kelly, J. F., & Hoepfner, B. B. (2013). Does Alcoholics Anonymous work differently for men and women? A moderated multiple-mediation analysis in a large clinical sample. *Drug and Alcohol Dependence*, 130, 186–193. doi:10.1016/j.drugalcdep.2012.11.005
- Kelly, J. F., Hoepfner, B., Stout, R. L., & Pagano, M. (2012). Determining the relative importance of the mechanisms of behavior change within

- Alcoholics Anonymous: A multiple mediator analysis. *Addiction*, 107, 289–299. doi:10.1111/j.1360-0443.2011.03593.x
- Kiluk, B. D., Nich, C., Babuscio, T., & Carroll, K. M. (2010). Quality versus quantity: Acquisition of coping skills following computerized cognitive-behavioral therapy for substance use disorders. *Addiction*, 105, 2120–2127. doi:10.1111/j.1360-0443.2010.03076.x
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 314, 829–832. doi:10.1126/science.1129156
- Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry*, 59, 877–883. doi:10.1001/archpsyc.59.10.877
- Loehlin, J. C. (2017). *Latent variable models: An introduction to factor, path, and structural analysis* (5th ed.). New York, NY: Routledge.
- Longabaugh, R., & Wirtz, P. W. (Eds.). (2001). *Project MATCH hypotheses: Results and causal chain analyses*. Project MATCH Monograph Series v. 8 (NIH Pub. No. 01-4238). Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593–614. doi:10.1146/annurev.psych.58.110405.085542
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, 1, 173–181. doi:10.1023/A:1026595011371
- MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review*, 19, 30–43. doi:10.1177/1088868314542878
- MacKinnon, D. P., Taborga, M. P., & Morgan-Lopez, A. A. (2002). Mediation designs for tobacco prevention research. *Drug and Alcohol Dependence*, 68, Supplement 1, S69–S83. doi:10.1016/S0376-8716(02)00216-8
- Magill, M., Gaume, J., Apodaca, T. R., Walthers, J., Mastroleo, N. R., Borsari, B., & Longabaugh, R. (2014). The technical hypothesis of motivational interviewing: A meta-analysis of MI's key causal model. *Journal of Consulting and Clinical Psychology*, 82, 973–983. doi:10.1037/a0036833
- Mason, W. A., Kosterman, R., Haggerty, K. P., Hawkins, J. D., Redmond, C., Spoth, R. L., & Shin, C. (2009). Gender moderation and social developmental mediation of the effect of a family-focused substance use preventive intervention on young adult alcohol abuse. *Addictive Behaviors*, 34, 599–605. doi:10.1016/j.addbeh.2009.03.032
- Mensing, J. L., Lynch, K. G., Ten Have, T. R., & McKay, J. R. (2007). Mediators of telephone-based continuing care for alcohol and cocaine dependence. *Journal of Consulting and Clinical Psychology*, 75, 775–784. doi:10.1037/0022-006X.75.5.775
- Moos, R. H. (2007). Theory-based processes that promote the remission of substance use disorders. *Clinical Psychology Review*, 27, 537–551. doi:10.1016/j.cpr.2006.12.006
- Morgenstern, J., Frey, R. M., McCrady, B. S., Labouvie, E., & Neighbors, C. J. (1996). Examining mediators of change in traditional chemical dependency treatment. *Journal of Studies on Alcohol*, 57, 53–64. doi:10.15288/jsa.1996.57.53
- Morgenstern, J., & Longabaugh, R. (2000). Cognitive-behavioral treatment for alcohol dependence: A review of evidence for its hypothesized mechanisms of action. *Addiction*, 95, 1475–1490. doi:10.1046/j.1360-0443.2000.951014753.x
- Morgenstern, J., Labouvie, E., McCrady, B. S., Kahler, C. W., & Frey, R. M. (1997). Affiliation with Alcoholics Anonymous after treatment: A study of its therapeutic effects and mechanisms of action. *Journal of Consulting and Clinical Psychology*, 65, 768–777. doi:10.1037/0022-006X.65.5.768
- National Institute on Alcohol Abuse and Alcoholism. (2013). *Program Announcement PAR 14-051—Mechanisms of Behavior Change in the Treatment of Alcohol Use Disorders*. Retrieved from <http://grants.nih.gov/grants/guide/pa-files/PAR-14-051.html>
- Newhouse, J. P., & McClellan, M. (1998). Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health*, 19, 17–34. doi:10.1146/annurev.publhealth.19.1.17
- Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11, 550–560. doi:10.1097/00001648-200009000-00011
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.
- Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, 31, 161–170. doi:10.1111/j.1467-9469.2004.02-123.x
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89, 845–851. doi:10.1037/0022-3514.89.6.845
- Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service & social action programs*. New York, NY: Russell Sage Foundation.
- Thoemmes, F., & Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, 4, 40–59. doi:10.1177/2167696815621645
- VanderWeele, T. J., Hawkey, L. C., Thisted, R. A., & Cacioppo, J. T. (2011). A marginal structural model analysis for loneliness: Implications for intervention trials and clinical practice. *Journal of Consulting and Clinical Psychology*, 79, 225–235. doi:10.1037/a0022610
- Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25, 659–706. doi:10.1146/annurev.soc.25.1.659
- Wirtz, P. W. (2007). Advances in causal chain development and testing in alcohol research: Mediation, suppression, moderation, mediated moderation, and moderated mediation. *Alcoholism: Clinical and Experimental Research*, 31, Supplement 3, 57s–63s. doi:10.1111/j.1530-0277.2007.00495.x
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement*, 76, 913–934. doi:10.1177/0013164413495237
- Word, C. O., Zanna, M. P., & Cooper, J. (1974). The nonverbal mediation of self-fulfilling prophecies in interracial interaction. *Journal of Experimental Social Psychology*, 10, 109–120. doi:10.1016/0022-1031(74)90059-6
- Zanna, M. P., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology*, 29, 703–709. doi:10.1037/h0036651