FROM: Longabaugh R, Wirtz PW (eds.). 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 (NIAAA), 2001.

Part I: Design and Methodology

Matching Hypotheses

Causal Chain Analysis

Matching Hypotheses

Richard Longabaugh, Ed.D., and Philip W. Wirtz, Ph.D.

Project MATCH was the largest randomized clinical trial of a psychosocial treatment for alcoholism ever undertaken, involving 1726 clients, 10 universities, and 10 clinical research units in a collaborative study with the National Institute on Alcohol Abuse and Alcoholism. Three treatment modalities were compared, and 21 client characteristics were tested as matching variables. For each matching variable, one or more hypotheses were developed to predict which treatments would be most and least helpful to clients who differed on that variable. The client population studied had a current DSM-III-R diagnosis of alcohol abuse or dependence. Clients had to be at least 18 years old and could not be currently dependent on sedative/hypnotic drugs, stimulants, cocaine, or opiates nor could they have used any intravenous drug in the prior 6 months. Clients were comprehensively assessed (Connors et al. 1994) prior to random assignment to one of the three treatments and then were followed at 3-month intervals for a year after treatment completion. In the outpatient arm of study, clients were again interviewed at 3 years after treatment completion.

The design chosen to test the matching hypotheses in Project MATCH was only one of several viable alternatives. While scientific considerations were primary in the eventual design selected, other factors came into play once scientific criteria were met. Examples of these factors are described below.

Treatments

A matching study might well have started by selecting a client variable to which to match (e.g., gender) and then embarked upon developing a treatment that would be tailored to gender needs, perhaps maximizing the chances of finding matching effects. We however selected our treatment modalities first and then chose the client attributes that might be differentially affected by assignments to these treatments. Thus, matching hypotheses were developed after treatments had been decided upon. We might have increased the robustness of matching effects had we first developed the hypotheses and then designed treatment modalities to maximize the differences between the treatments by embodying active ingredients thought to be differentially effective in interaction with the client attribute.

The treatments chosen were Cognitive-Behavioral Coping Skills Therapy (CBT; Kadden et al. 1992), Motivational Enhancement Therapy (MET; Miller et al. 1992), and Twelve Step

The author listing of this chapter is alphabetical. Both authors contributed equally.

Richard Longabaugh, Ed.D.
Center for Alcohol and Addiction Studies
Brown University, School of Medicine
800 Butler Drive, Potter Building, Room 204
Providence, RI 02906
E-mail: Richard_Longabaugh@Brown.edu

Philip W. Wirtz, Ph.D.
Department of Management Science
George Washington University
2115 G Street NW, #403
Washington, DC 20052
Email: pww@gwu.edu

Facilitation (TSF; Nowinski et al. 1992). They were selected on the basis of several criteria (Donovan et al. 1994) :

- Demonstrated clinical effectiveness
- Potential for revealing matching effects
- Applicability to existing treatment programs and client populations
- Distinctiveness from the other MATCH treatments selected
- Feasibility of implementation within the constraints of a research trial.

Cognitive-behavioral therapy had been shown to be effective in a great number of studies, was widely used in academic settings, had been delivered to alcohol dependent populations in a great many studies (Holder et al. 1991), and could be delivered within a short period. Several of the principal investigators had previously conducted randomized clinical trials involving variations of CBT.

While we had considered a no-treatment control group, ethical and practical considerations precluded this. Our search for a minimum treatment comparison group led to the selection of MET, a specific application of motivational interviewing (Miller and Rollnick 1991). Motivationally based brief intervention had been shown to be effective in a number of studies (Holder et al. 1991). Because of its philosophy of maximizing utilization of client resources, it was thought to provide a significant contrast from CBT, which assumed that the road to recovery was through teaching new skills to the client. Consequently, many of the matching predictions developed were predicated in whole or in part on the expectation that four sessions of MET would be insufficient for clients with severe problems in various areas.

TSF was selected because of the popularity of the Minnesota Model in the treatment field. Fundamental to the Minnesota model are the Twelve Steps of Alcoholics Anonymous and the integration of treatment with the client's involvement in AA. Despite the popularity of this treatment approach, its effectiveness had not received adequate rigorous testing (Miller and Hester 1986). The opportunity to provide a rigorous test of the model in the context of client-treatment matching was compelling. It was expected that its treatment philosophy, which included a reliance on support groups and a higher power, would differ markedly from the MET focus on utilization of client resources and CBT's focus on individual skill development guided and taught by the CBT therapist.

Other treatment modalities were seriously considered, for example, the Community Reinforcement Approach (CRA). It was decided not to select CRA in view of perceived difficulties in implementing it across 10 clinical research units (CRUs). One pharmacological intervention, naltrexone, was also considered but was judged to not have sufficient evidence for its effectiveness at that time to warrant its inclusion in a major multisite test of matching.

The study design compared treatment modalities within two distinct treatment settings (or arms), aftercare and outpatient. The CRUs successful in the competition for the cooperative agreement grants were located at sites that had access to either inpatient or freestanding outpatient programs. The outpatient arm involved clients who had not had an inpatient treatment immediately preceding their involvement in the MATCH trial. In the case of CRUs attached to inpatient units, an aftercare study was conducted because it was deemed infeasible to superimpose the MATCH treatment on the inpatient program. The study was not conducted as an aftercare versus outpatient matching study because clients could not be randomly assigned to treatment setting.

Dependent Variables

As the study was of alcohol treatment effectiveness, one or more measures of alcohol consumption were to be included as primary dependent variables. Among a large number considered, percentage of days abstinent (PDA) and average drinks per drinking day (DDD) were chosen. PDA was an easy selection as it had high usage in prior studies and provided a relatively straightforward measure of drinking frequency (Babor et al. 1994). The selection of DDD was much more difficult. The goal was to index

drinking intensity, a dimension of drinking typology that was empirically associated with, but conceptually independent of drinking frequency.

Several candidates were considered, the leading contender among them being percentage of heavy drinking days. This measure was not selected, however, because of difficulty in designating a "heavy drinking day". While earlier research had frequently used six or more drinks as a cutoff, changes in cultural practices in the United States suggested that this figure was now too high. Furthermore, it was becoming clear that males and females differed in the effects of consuming the same amount of alcohol. Also, body weight and other factors were refining the conception of risky alcohol consumption. We concluded that what constituted a heavy drinking day was a moving target. DDD was selected because it represented an index of absolute amount of alcohol consumed (measured in standard drinks), independent of what might eventually be determined to constitute hazardous drinking intensity.

DDD had the major disadvantage that clients who had no drinking days during a period would have no data point. As this was likely to be so for a sizable percentage of clients within any followup period, this was unsatisfactory. Instead we opted to include in the analyses clients who had zero drinks on a "drinking day". This retained all clients in the analysis who were successfully followed up but also ensured that there would be a hefty correlation between PDA and DDD during the posttreatment period. The decision to include "zero drink drinking days" also resulted in a heavy concentration of observations at zero drinks per drinking day. This led to a methodologically determined lack of independence between the two primary dependent variables.

Primary dependent variables were limited to two in order to preserve the power to detect credible differences. Using the same two made it possible to preserve a standard metric for comparing matching effects across different matching variables.

Another promising construct for a primary dependent variable was believed to be negative consequences of alcohol consumption. However a disadvantage was that, at the time, there was no standard measure of negative consequences (Zweben and Cissler 1996). Alcohol dependence was also considered but discarded because it was an ambiguous construct that was difficult to quantify. Total abstinence was excluded as a primary dependent variable because of its insensitivity to major changes in drinking patterns that did not involve total abstinence and because of the relatively low percentage of clients who would be likely to achieve total abstinence throughout the entire recorded followup period.

Hypothesis Selection

The development of client-treatment matching hypotheses in Project MATCH was a significant departure from previous research on matching which was accomplished via single-site, smaller scale studies (Mattson et al. 1994). Because of the nature of the multisite collaborative study, a Steering Committee (SC) was authorized to develop the research hypotheses and the design to test these hypotheses as well as to assume collective responsibility for conducting the study and reporting the findings. The 11 principal investigators and other senior investigators involved in the process published main findings papers under the corporate authorship identified as the Project MATCH Research Group, PMRG (Project MATCH 1997*a*).

At the outset, the SC decided that rather than testing one or two matching hypotheses, several would be tested. Members of the PMRG formed self-selected groups called matching hypothesis teams (MHTs). Typically each team was composed of three to five members, including at least one statistician/methodologist and one or more clinical scientists who were well versed in the substance of the matching variable. Each team was charged to develop predictions about matching a single client attribute to one or more of the three study treatments. These were then presented to the entire SC, which had the responsibility for deciding which matching variables and a priori hypotheses would be selected for testing in the trial.

So that the entire SC would be fully informed in making these decisions, a procedure and

review process was put into place. Each MHT developed a 10–20 page document presenting the rationale for selection of a given client variable and rationale for each of the matching predictions proposed. The rationale included a review of published empirical support for use of the variable in matching research and a theoretical justification for the matching predictions made. Central to the theoretical justification was a "causal chain" (discussed in detail in the next chapter) that would provide a description of the underlying process (or processes) postulated to be necessary for the matching prediction to be supported. The procedures for operationalizing the matching variable were specified as well as analytic methods to test the predictions.

Each MHT document was submitted to a matching hypothesis review team composed of the editors of this volume, RL and PWW, whose primary expertise was, respectively, substantive and methodological/statistical. This review team critiqued the document and determined if it was ready for review by the full SC. Eventually each matching hypothesis was discussed, critiqued, and voted upon by the full SC to determine whether it would be included as one of the matching hypotheses.

This process was intensive and lengthy, spanning approximately 2 years during the planning phase. First, the document developed by the MHT had to pass the team's own review. Then, critique by the review team was followed by one and usually several iterations with the MHT before the document was ready for consideration by the full SC. Typically, on first presentation, the SC did not vote acceptance but raised questions and returned the document to the MHT for further revisions. The revised manuscript would then again undergo one or more iterations between the review team and the MHT before being resubmitted to the full SC for further deliberation. At that point, the matching hypothesis would be voted in or out. This last step was not perfunctory. In total, 28 matching variables were proposed by MHTs to the SC. Of these, 21 were tested in matching predictions.

Many of the variables selected for matching had been previously researched in single-site studies by members of the SC: Alcohol Involvement (Miller, Rychtarik,), Alcohol Dependence (Babor, Cooney), Psychiatric Severity (Cooney, Kadden), Sociopathy and Antisocial Personality Disorder (Cooney, Kadden, Litt, Longabaugh), Cognitive Impairment (Cooney, Donovan, Kadden, Longabaugh), Motivation (Carbonari, DiClemente, Miller, Tonigan), Self-Efficacy (DiClemente), Typology (Babor, Litt), Social Support (Longabaugh, Stout), Social Functioning (Cooney, Kadden), Interpersonal Dependency (Longabaugh), and Alcoholics Anonymous (Tonigan).

Given the treatments selected and the demonstrated potential of most of the matching variables, a primary focus of the matching hypothesis teams was development of the rationale for making client-treatment matching predictions, that is, why one or the other of the three already selected treatments would be likely to have a differential effect on the client matching variable.

How many treatments were to be included in the individual matching hypotheses was left up to the matching hypothesis teams. Some teams chose to develop predictions that involved contrasting matching effects for all three treatments, while other teams selected just two of the three treatments in their contrasts. Still others included all three treatments in the matching predictions but aggregated two together and compared the aggregate against the third. Any such aggregation was based on theory; for example, one treatment included an active ingredient thought likely to especially effect clients with a given attribute, whereas the other two treatments lacked or deemphasized this active ingredient. Another factor was the MHT's belief in the likely strength of the predicted matching effect. Teams believing that the matching effect was likely to be robust were more apt to include more predictions.

The SC decided during this review process that not all matching hypotheses would have equal status. One group was designated as *primary matching hypotheses*. They were given a higher priority because the SC felt they had a more compelling rationale because either their prior empirical support was stronger or the theory underlying the expected matching effects was more persuasive.

A second group of client variables was designated to test *secondary matching hypotheses.* Matching predictions involving these variables were also developed a priori but were deemed of lesser priority because their empirical and/or theoretical rationale was seen as less persuasive. By the end of the review process, 9 matching variables were accepted by the SC as primary and the remaining 12 as secondary. Had our results supported this designation of primary versus secondary hypotheses, much more would have been made of the distinction. While earlier publications reported the results of primary (Project MATCH Research Group 1997*a*) and secondary (Project MATCH Research Group 1997*b*) matching hypotheses, that distinction has proved unnecessary and is not used in this monograph. Table 1 lists all of the Project MATCH matching hypotheses.

Client-Treatment Interactions

For our purposes, an interaction is said to occur when a differential response to two treatments occurs as a function of the degree to which a client possesses a particular characteristic. Project MATCH allowed matching effects to involve interactions that were either *ordinal* or *disordinal.* A *disordinal* interaction was judged to occur when the two slopes were observed to cross one another at some point along the measurable client attribute continuum (figure 1), such that clients at one interval on the continuum were found to have better drinking outcomes when assigned to one treatment, but clients at another interval on the continuum were found to have better drinking outcomes when assigned to the contrasting treatment.

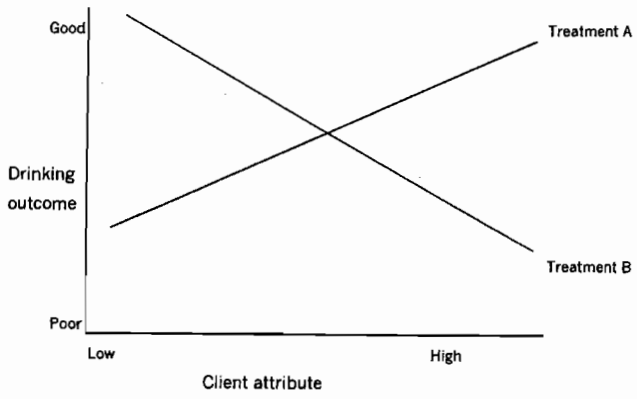

Figure 1. Disordinal interaction

If, in contrast, at one interval along the client attribute continuum, clients were found to have better drinking outcomes when assigned to one treatment but clients elsewhere on the continuum were found to have neither better nor worse drinking outcomes, then the interaction was designated as *ordinal* (figure 2).

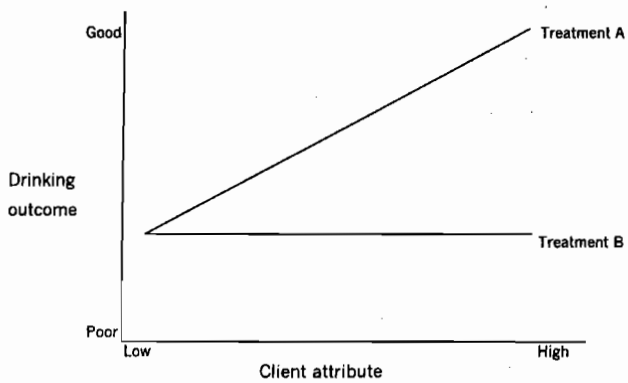

Figure 2. Ordinal interaction

In Project MATCH, we were interested in detecting both disordinal and ordinal interactions, even though disordinal interactions were more likely to have immediate wide-ranging implications for the field. Because of their theoretical value as well as their more limited immediate clinical value, interactions hypothesized to be ordinal were also approved by the SC.

All attribute-treatment interaction effects were hypothesized to be linear. Therefore, it was expected that the further up or down the range of the client attribute scale the actual value is, the more likely the treatments are to lead to different outcomes. Had a nonlinear interaction been hypothesized (or observed), the detection procedure would have become much more complex.

Ordinal Interactions at the Nonpredicted End of the Continuum

We did not specify as essential evidence for matching that the observed difference between the treatments occur at the end of the continuum where we expected the matching effect to

Table 1. Summary of hypothesized contrasts for each matching variable

Matching variable	Hypothesized contrast of slopes or differences between means[a]		
	CBT vs. MET	TSF vs. MET	CBT vs. TSF
Alcohol Dependence			TSF > CBT
Alcohol Involvement[b]	CBT[c] > MET	TSF[c] > MET	
Cognitive Impairment	CBT > MET	TSF > MET	TSF > CBT
Psychopathology and Severity	CBT > MET		CBT > TSF
Sociopathy and ASPD	CBT > MET	TSF > MET	CBT > TSF
A versus B Typology[b]	CBT[c] > MET	TSF[c] > MET	
Anger[b]	MET > CBT[c]	MET > TSF[c]	
Conceptual Level		MET > TSF	
Meaning Seeking[b]		TSF > MET[d]	TSF > CBT[d]
Religiosity[b]		TSF > MET[d]	TSF > CBT[d]
Interpersonal Dependency[b]		TSF > MET[d]	TSF > CBT[d]
Gender			CBT > TSF
Motivational Readiness	CBT > MET		
Problem Recognition[b]	CBT[c] > MET	TSF[c] > MET	
Self-Efficacy confidence	MET > CBT	MET > TSF	
Temptation minus confidence		TSF > MET	
Network Support for Drinking	CBT > MET	TSF > MET	
Prior AA[b]		TSF > MET[d]	TSF > CBT[d]
Poor Social Functioning	CBT > MET		CBT > TSF

[a] The hypothesized contrasts predict differences in slopes of the regression lines for each treatment on outcome as a function of client attribute. With the exception of the gender and typology attributes (which take on only discrete values), all contrasts take the form: The difference between the first treatment and the second becomes more positive (or less negative) with increasing values on the attribute. The gender and typology attributes take the form: The difference in means between the treatments is greater at one level of the attribute than at the other. Hypotheses did not test whether interactions were ordinal or disordinal.

[b] The rationale underlying the alcohol involvement, meaning seeking, typology, anger, religiosity, interpersonal dependency, problem recognition, and prior AA hypotheses assumes that, pertinent to the putative active ingredients involved in the hypothesized matching effect, two treatments are not different in their effect. Therefore, they were combined into a single condition which was then contrasted with the third treatment.

[c] Combined TSF and CBT treatment groups

[d] Combined MET and CBT treatment groups

be apparent. Thus, it was possible that some interactions would be observed at the "wrong end" of the variable continuum; this may be a somewhat difficult point to grasp and therefore can perhaps be best understood by an example.

In Project MATCH, the matching hypothesis team predicted that CBT would be more effective than TSF or MET for clients with higher psychopathology. This assertion was made with the knowledge that some components of the CBT treatment are designed to treat psychopathology while this is not so for either TSF or MET. Thus, the attribute-treatment interaction was predicted to occur because the CBT slope would be more positive than would the TSF or MET slopes, that is, as client psychopathology increased, it would be expected that CBT clients would have increasingly more favorable drinking outcomes than would TSF or MET clients. However, we instead found that with *decreasing* psychopathology, TSF clients had increasingly better outcomes than did CBT clients. The treatment triaging implication is to assign low psychopathology clients to TSF and not to CBT, rather than to assign high psychopathology clients to CBT and not to TSF.

But does this triage implication mean that the matching hypothesis is unsupported? Examination of the attribute-treatment interaction (see figure 2, page 88) reveals that the relative position of the two slopes is as predicted. As psychopathology increases, the outcomes of CBT clients are relatively better than at lower levels of psychopathology. In contrast, as psychopathology of TSF clients increases, their outcomes are not relatively better than at lower levels of psychopathology. As a result, the CBT slope is more positive than is the TSF slope, as was predicted from the rationale leading to the matching prediction.

Thus, the matching effect indicates that CBT is worse than TSF for clients without psychopathology. Clearly, evidence for and against the matching hypothesis is incomplete. It may be that TSF is a more effective treatment than CBT for the average alcohol dependent client because TSF has more of a particular active ingredient that would help all alcohol dependent clients irrespective of their psychopathology. However, because of CBT's unique effectiveness for clients with greater psychopathology, it is able to close the gap of effectiveness between TSF for clients with high psychopathology. Other explanations are also conceivable.

What is critical for the purpose of making the present point is that, as predicted, (1) the CBT slope was more positive than the TSF slope in the direction predicted at the required level of statistical significance and (2) at some interval along the client attribute continuum clients assigned to one of the two treatments would have better drinking outcomes than clients assigned to the other treatment, while this was not so for the remainder of the client sample.

In this example, all of the criteria specified for meeting the MATCH requirements for a matching effect were met. However, whether the theory underlying the matching effect is supported cannot be determined from outcome data alone. The causal chain analyses presented in this volume contribute to the interpretation of the matching results by examining further pertinent information.

Clinical Significance

If a disordinal interaction occurs, the clinical implications are obvious. The clinician would assign clients to the treatment that is best for clients with similar characteristics. All other factors being equal, those having a low score on the attribute should be assigned to treatment A, while those with a high score on the attribute should be assigned to treatment B.

Had Project MATCH found several disordinal interactions, the value to the treatment field would have likely been considerable. However, disordinal interactions were found for only three client attributes—client anger and network support for drinking in the outpatient arm of study (although ordinal interactions were hypothesized in each instance) and client dependence on alcohol in the aftercare arm of study.

The clinical benefits to the treatment field of an ordinal interaction are less obvious. A first reaction is, if clients with one score on the attribute do better in one treatment, but there is no difference between treatments on the outcome for clients with other scores on the attribute, then assign them all to the treatment that

has been found to be more effective for at least some of them, and no worse for the others.

However, the decision tree may become more complex when other considerations are taken into account. Two examples of considerations that may influence treatment choice involve practical issues. For example, if the treatment that is more effective for some requires more resources to deliver than the other, then this treatment might be reserved for only those clients who are likely to incrementally benefit. This is the rationale for assigning clients to different levels of treatment intensity, such as inpatient versus outpatient treatment.

Another consideration is the availability of the treatments contrasted. If a given program has only a limited number of therapists who have been trained to deliver the treatment that is more effective for some, but has other therapists available who can deliver the other treatment that is not less effective for the others, then the program might well assign clients with nondiscriminating scores to this other treatment.

Matching Hypotheses

As operationalized in Project MATCH, a client-treatment matching hypothesis is a prediction about the occurrence of a statistically significant interaction between a client attribute and a treatment modality, such that the predicted regression slopes of the contrasted treatments over the range of the client attribute are different from one another in a direction consistent with the theory. Graphically, this is evident when the slopes of the regression lines of the compared treatments on client outcomes are significantly different from one another in the hypothesized direction across the range of values of the client attribute. The criterion for statistical significance was achieved when a one-tailed prediction had an observed probability value equal to or less than 0.05, corrected for the number of contrasts tested for that matching variable.

It is important to note a criterion that was *not* included in the specification of our matching hypotheses. We did not require that the matching prediction specify the directionality of the individual slopes; rather the requirement was that the *difference* in slopes be consistent with theory-based expectations.

This decision was based on the absence of an untreated control group in Project MATCH. Thus, we were unable to tell whether a given client attribute would have prognostic value in the absence of treatment. Without this knowledge, we would be unable to say whether a slope that was negative was an indication that the treatment-client attribute combination was having an adverse affect on drinking outcome as compared to no treatment. Thus, if the slope of the line is descending for clients higher on the attribute scale, the only meaningful comparison possible is relative to the slope of the line for the comparison treatment. If both are descending, then the slope with the lesser decline would still be considered a relative match for that treatment. It is theoretically possible that both treatments may have been mismatches for this set of clients relative to no treatment at all. It must therefore be remembered that throughout this volume we are examining only client attribute-treatment combinations relative to other combinations of the attribute with different treatments.

Matching Effects

If the observed attribute by treatment interaction met these two criteria (statistically significant different slopes in a specified direction), then we tested whether the observed interaction had any clinical significance. An interaction that also had clinical significance was deemed a "matching effect". Clinical significance was asserted when we could identify one or more intervals along the client attribute axis in which we were at least 95-percent confident that a client having an attribute score within this interval would have an average drinking outcome superior to the contrasting treatment condition.

However, in order for the effect to be a matching effect it was also required that not all clients assigned to the one treatment condition would have better outcomes than if assigned to the other condition. In other words, at some point along the client attribute continuum, the two

slopes also had to either cross or be close enough to each other that we could not be 95-percent confident that clients with attribute scores at this point would have different drinking outcomes. If this latter criterion were not met, then we did not designate the attribute-treatment interaction to be a matching effect.

Clinical Versus Theoretical Value

In one respect, the criterion of clinical usefulness sets a higher standard for results to be judged of value than does the criterion of theoretical importance. For a client-treatment interaction to be clinically useful, it is not only necessary that attribute-treatment interactions be found, it is also minimally necessary that one subset of clients, but not another, be identified for whom the likelihood of a good outcome is enhanced by assignment to one or another treatment. The amount of clinical improvement resulting from differential treatment assignment is indicative of the clinical value to be attached to the matching effect. If the outcomes of a large number of clients could be enhanced, this result translates into clinical usefulness. Similarly, if the treatment outcomes of a smaller number of clients would be greatly improved by matching, this also would have greater clinical significance.

In contrast, for an attribute-treatment interaction to be theoretically interesting, it is only necessary that the slopes of the contrasting treatment by attribute lines be statistically significantly different from one another. This would include the case in which there was no matching effect discerned, even though there was a statistically significant interaction. This possible scenario is illustrated in figure 3, where it can be seen that at no level of the matching variable was the one treatment less effective than the other, yet an attribute-treatment interaction was observed, nested within a main effect of treatment.

Thus, it is not even necessary that the interaction account for clinically meaningful variance; what is important is that a lawful relationship has been established.

Figure 3. A client-treatment interaction without a matching effect.

If a client attribute-treatment interaction is statistically significant, but we are unable to identify the mechanisms producing this interaction, two effects result. First, it undermines the credibility of the effect, that is, was the relationship observed by chance or because of a lawful relation that exists between the variables. In other words, is it really true? Second, it also substantially limits the information available when attempting to replicate and/or generalize the effect with new populations. The scientific aim is to identify the set of conditions under which a client attribute-treatment interaction occurs, irrespective of whether a clinically important matching effect is observed.

Data Analysis

A Design Committee[1] was constituted for developing the plan for data analysis. Upon completion, the Project MATCH data analysis plan consisted of six major phases: (1) a preliminary phase in which descriptive and other analyses were performed to provide an overview of the data and to test assumptions for the main hypothesis tests; (2) confirmatory testing of the primary a priori matching hypotheses; supplementary analyses addressing (3) the secondary a priori matching hypotheses and (4)

[1] Design Committee members included Joseph Carbonari, Fran Del Boca, Mark Litt, Richard Longabaugh, Larry Muenz, Robert Rychtarik, Robert Stout (Chair), Scott Tonigan, and Philip Wirtz.

alternative analytical approaches for testing matches; (5) analyses directed at major issues other than patient-treatment matching; and (6) exploratory analyses aimed at detecting matching that was not predicted a priori. As the focus of this monograph is testing the a priori matching hypotheses, only the first four of these phases is reviewed.

Preliminary Analyses

Extensive analyses were performed to provide an overview of the data and test assumptions for the main hypothesis tests. Of primary concern at this stage were:

- Compliance
- Missing data
- Psychometric and distributional properties of variables that were critical to the primary objectives of the trial
- The effectiveness of urn randomization in producing equivalent treatment groups
- Detection of any site-based effects
- The relationships among the matching factors and among the primary dependent measures
- Assessment of the adequacy of the potential covariate candidates.

A number of key decisions resulted from these preliminary analyses. The distributional properties of the primary outcome variables suggested the need for an arcsin transformation of the frequency measure (percentage of days abstinent) and for a square root transformation of the intensity measure (drinks per drinking day). The two key outcome variables were found to be moderately correlated, but enough differences were found to merit separate analysis of each of them in the main analytical phase of the trial. Based on extensive analysis of missing data patterns, a "two-thirds" rule was adopted for aggregating daily data into summary weekly and monthly outcome indices, such that an individual's outcome indicators were recorded as "missing" on any week (during treatment) or month (following treatment) in which less than two-thirds of the drinking data were available for that individual.

Testing the Matching Hypothesis

For the two principal classes of a priori hypotheses addressed in Project MATCH (primary and secondary matching hypotheses), individual differences in response to alcohol treatment were modeled as a latent growth process (see, for example, Bryk and Raudenbush 1992; Laird and Ware 1982; Goldstein 1986; Strenio et al. 1983). Conceptually, this is roughly equivalent to creating a separate plot for each subject, with up to 12 points plotted (one for each followup period where data are available for the individual subject), with the outcome variable on the vertical axis and time on the horizontal axis. For each subject, a smooth curve (a latent growth curve) is then fit to these data so as to come as close as possible to the plotted points. Depending on investigator-specified constraints, this curve can be as simple as a straight line or as complex as an n^{th} order polynomial and is of the same level of complexity for all subjects.

Review of the matching hypotheses led to the decision that testing for quadratic polynomials would be sufficient. First, it could be argued that a matching effect would result from treatment that would persist in strength across the one year of posttreatment followup. Certainly this would be the hope of clinical investigators. On the other hand, given the oftentimes observed short-lived main effects of treatment, it was also quite plausible that matching effects present at the completion of treatment might subside with increasing time from treatment.

Still a third possibility considered was that matching effects might take time to emerge after treatment. This would be especially likely if ceiling effects were operative at treatment completion because all clients were doing very well. Only with time at risk from treatment completion would the differential effects of matching emerge. Finally, it was conceived as possible that it might take time for an effect to emerge, but that this effect would dissipate with longer intervals from treatment completion. A quadratic time variable would capture any of these possible scenarios. No greater levels of complexity were considered to be interpretable.

A further decision was made to split the period of observation into two distinct periods, within treatment and posttreatment. The primary window of observation was the period following scheduled treatment completion. (Treatment was scheduled to occur over 12 weeks). The Steering Committee believed that the posttreatment window was by far the most important time to consider. A matching effect that did not persist after the completion of treatment was considered to have little practical importance. Thus, the primary outcome period was the posttreatment period extending from the beginning of month 4 to the end of month 15. However, within-treatment matching effects were also to be investigated for their theoretical importance. If it were to be found that a matching effect occurred during treatment, this would support the theoretical rationale underlying the matching prediction, even though the observed matching effect would not substantially help clients.

It is interesting to note that this emphasis on posttreatment effects is not shared in randomized clinical trials of pharmacotherapeutic effects. In such studies, the end point for analysis is usually when the client is taken off the medication. For a parallel analysis in psychosocial treatment studies, the end point would also be at the completion of the therapy administration. Pharmacotherapy studies are more likely to focus on efficacy whereas psychosocial treatment outcome studies have traditionally focused more on effectiveness.

Model Parameters

The data analysis plan had no provision for systematically testing for nonlinear interaction effects between treatment modality and client attribute. This was because none of the matching hypotheses to be tested anticipated nonlinear matching effects. However, such relationships are not inconceivable. For example, it could be hypothesized that CBT would be more effective with clients who had moderate social skills than with clients who had either very good social skills or were markedly deficient. Those with extreme skills deficits might be unable to benefit sufficiently from social skills acquisition, whereas those with exceptionally good skills would not be in need of any further skills acquisition.

In contrast, TSF might be especially beneficial both for clients with especially poor social and exceptionally good social skills. Those with poor social skills might benefit from the group support offered to AA members, irrespective of their coping skills. Those with exceptionally good social skills might attain high status within AA because of these skills, which would help them to serve as role models for those less fortunate. These two nonlinear functions between a social skills matching variable and CBT and social skills and TSF could produce a nonlinear matching effect.

Some of the matching hypothesis teams did give consideration to the possibility of nonlinear matching effects, but it was decided that adding this complexity to an already seemingly very complex data analysis plan was unnecessary. Consequently, no comprehensive test of nonlinear matching effects was ever conducted, although some of the individual matching hypothesis teams did conduct exploratory analyses for such effects when linear matching effects were not observed.

For each matching hypothesis, each analysis was conducted twice (once for each of the two principal dependent measures) in each arm (outpatient, aftercare) and for each time period (within treatment and posttreatment) using the PROC MIXED procedure in SAS. In each analysis, the intercept, time, and time squared terms were entered as random factors in a linear model which also included the baseline value of the outcome variable, the matching variable, the treatment variable, the matching-by-treatment interaction, and the interactions of each of these with time and with time squared.

Analyses involving these time effects were centered at the midpoint of the associated period (within treatment, posttreatment) for two reasons. First, midpoint centering can reduce the implicit colinearity between an interaction term and the constituent variables of which it is a product (see Aiken and West 1991), thereby improving the power of the design to identify matching effects when they exist. Second, centering facilitates the interpretation of any significant attribute-treatment interaction which

does not change across time. Specifically, when time is centered at the midpoint of the period under investigation (and in the absence of a significant interaction of linear or quadratic time with the attribute-treatment effect), the coefficient of a significant attribute-treatment interaction represents the difference in slopes at the midpoint of the period.

Analyses involving discrete matching variables (e.g., gender, typology) were conducted in a manner similar to those associated with continuous matching variables, although the interpretation of the results varied slightly. Instead of focusing on between-treatment differences in the *slope* of the relationship between the dependent variable and the (continuous) matching variable, these analyses focused on between-treatment differences in the *mean* of the dependent variable at the two levels of the matching variable. Thus, the question became, for example, whether the difference in mean outcome between two treatments varied as a function of gender (possibly in interaction with linear or quadratic time). Similar to the analyses involving continuous matching variables, the hypotheses associated with the discrete matching variables were tested using Bonferroni-adjusted a priori directional contrasts.

Two sets of covariates were examined. The first set included the baseline value of the criterion drinking measure and its interaction with time. The second added a site-effect term. The objective of these analyses with the site-effect term added was to minimize the possibility that any discovered interactions were merely spurious reflections of differences between sites (i.e., to control for one potential source of internal invalidity). Conducting the analysis both ways—with and without the site in the model—was intended to facilitate discovery of hypothesized interaction relationships while at the same time allowing any such discovered relationships to be conservatively interpreted if they dissipated as a result of including site in the model.

Formal Testing Procedure

Each of the primary and secondary matching hypotheses was formally tested using a series of hypothesis-specific directional contrasts. Of primary interest was whether a given contrast was statistically significant (in the hypothesized direction), either by itself (directionally) or in interaction with the linear or quadratic time terms (nondirectionally). Contrasts for which a significant linear time and/or quadratic term appeared were then subjected to further testing on a month-by-month (posttreatment) or week-by-week (within treatment) basis in order to more adequately capture the timing of the increasing or decreasing matching effect. For any matching effect that interacted with time, each individual period was tested to judge whether the matching effect was present during that period. For examining these changing matching effects, $p<.05$ was judged to be a sufficient criterion for deciding whether matching was present within that single time period.

Although in theory only three data points are needed to estimate the parameters of any individual's latent growth curve under a quadratic polynomial specification, additional data points provide a much better assessment of the validity of the polynomial specification selected and of the error terms employed in the statistical tests. The flexibility offered by the Time Line Follow-Back procedure (Miller and Del Boca 1994) for assessing drinking behavior provided Project MATCH investigators with the opportunity to analyze drinking behavior across any time interval desired (including daily, if warranted). Recognizing that computational restrictions prohibited latent growth analysis of daily data across an 18-month period, 12 summary data points were felt to be sufficient to capture the essence of an individual's drinking behavior. For this reason, the latent growth analyses were based on the 12 *weekly* assessments of each participant's drinking behavior *during* treatment and on the 12 *monthly* assessments of each participant's drinking behavior *following* treatment.

For the a priori matching analyses, there was concern that strict application of a Bonferroni adjustment (accounting for both the primary and the secondary hypotheses)—including all of the hypotheses, two outcome variables per hypothesis, and multiple contrasts within some of the hypotheses—would result in an excessively conservative cutoff level that would unduly enhance the likelihood of a type 2 error. Since the

matching hypotheses were conceived of as conceptually independent of one another, it was decided to apply a trialwide Bonferroni correction to tests of significance within each hypothesis family, taking into account the two outcome variables and the number of proposed contrasts for that hypothesis, irrespective of whether the matching hypothesis was considered to be primary or secondary. If, for example, there were three hypotheses relating to a single matching variable, then those hypotheses were tested at a Bonferroni-corrected alpha level of 0.05/3. Because there were two dependent variables, the alpha level was further corrected by a factor of 2.

Latent Growth Analysis Versus Fixed-Effects MANOVA

At the time that Project MATCH was being designed, latent growth modeling (LGM) was largely unknown in alcohol research, although it was in widespread use in certain social science disciplines (notably educational psychology). The decision was made to focus on this analytical technique as the primary methodology for several reasons. First, there is extensive evidence of idiographic patterns of drinking among those dependent upon alcohol, and LGM permits individuals to manifest their unique patterns (within the confines of the selected polynomial form). Second, unlike classical general linear model formulations, LGM permits individual variation across time to be modeled as a random (rather than fixed) effect, thereby providing (1) estimation of the extent of between-individual variation across time and, as a result, (2) better estimation of the interaction effects and the error terms which form the basis for the statistical tests. Third, unlike multivariate formulations of the general linear model, individual subjects are permitted to have missing data at one or more time points and still be included in the analysis.

For these reasons, LGM was selected as the primary analytical approach for testing the Project MATCH primary and secondary matching hypotheses. However, because latent growth modeling was relatively new to the alcohol field, all major analyses were conducted a second time using the more widely understood fixed-effects general linear modeling procedure. Results of the two approaches consistently converged, with small differences due primarily to the differential assumptions, the slightly different sample size (classical general linear modeling requires no missing data), and cross-client variance in parameters.

In summary, the data analysis plan was quite elegant and appeared well suited to the hypotheses that it was developed to test. If matching was an active process in treatment, it seemed likely that this powerful design was well equipped to identify these effects, at least within the scope of possible effects envisioned.

Acknowledgments

This project was supported by a series of grants from the National Institute on Alcohol Abuse and Alcoholism as part of the cooperative agreement on matching patients to alcoholism treatments (U10AA 08443). We particularly thank Dr. Margaret E. Mattson and Jane K. Myers for their careful and thoughtful review and editing of this chapter. We also thank Julia Wolin for her assistance in the preparation of this chapter.

References

Aiken, L.G., and West, S.G. *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park: Sage, 1991.

Babor, T.F.; Longabaugh, R.; Zweben, A.; Fuller, R.; Stout, R.L.; Anton, R.F.; and Randall, C.L. Issues in the definition and measurement of drinking outcomes in alcoholism treatment research. In: Donovan, D., and Mattson, M.E. Alcoholism Treatment Matching Research: Methodological and Clinical Approaches. *Journal of Studies on Alcohol* Suppl. No. 12 (Dec.): 101–111, 1994.

Bryk, A.S., and Raudenbush, S.W. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications, 1992.

Connors, G.J.; Alen, J.P.; Cooney, N.L.; DiClemente, C.C.; Tonigan, J.S.; and Anton, R.F. Assessment issues and strategies in alcoholism treatment matching research. In: Donovan, D., and Mattson, M.E. Alcoholism Treatment Matching Research: Methodological and Clinical Approaches. *Journal of Studies on Alcohol* Suppl. No. 12 (Dec.):92–100, 1994.

Donovan, D.M.; Kadden, R.M.; DiClemente, C.C.; Carroll, K.M.; Longabaugh, R.; Zweben, A.; and Rychtarik, R. Issues in the selection and devel-

opment of therapies in alcoholism treatment matching research. In: Donovan, D., and Mattson, M.E. Alcoholism Treatment Matching Research: Methodological and Clinical Approaches. *Journal of Studies on Alcohol* Suppl. No. 12 (Dec.):138–148, 1994.

Goldstein, H.I. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 73(1):43–56, 1986.

Holder, H.D.; Longabaugh, R.; Miller, W.R.; and Robonis, A.V. The cost effectiveness of treatment for alcoholism: A first approximation. *Journal of Studies on Alcohol* 52:517–540, 1991.

Kadden, R.; Carroll, K.M.; Donovan, D.; Cooney, N.; Monti, P.; Abrams, D.; Litt, M.; and Hester, R. *Cognitive-Behavioral Coping Skills Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals With Alcohol Abuse and Dependence.* Project MATCH Monograph Series. Vol. 3. DHHS Pub. No. (ADM) 92–1895. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism, 1992.

Laird, N.M., and Ware, H. Random-effects models for longitudinal data. *Biometrics* 38:963–974, 1982.

Mattson, M.E.; Allen, J.P.; Longabaugh, R.; Nickless, C.; Connors, G.J.; and Kadden, R.M. A chronological review of empirical studies matching alcoholic clients to treatment. *Journal of Studies on Alcohol* Suppl. 12:16–29, 1994.

Miller, W.R., and Del Boca, F.K. Measurement of drinking behavior using the Form 90 family of instruments. *Journal of Studies on Alcohol* Suppl. 12:112–118, 1994.

Miller, W.R., and Hester, R.K. The effectiveness of alcoholism treatment: What research reveals. In: Miller, W.R., and Heather, N., eds. *Treating Addictive Behaviors: Processes of Change.* New York: Plenum Press, 1986. pp. 121–174.

Miller, W.R., and Rollnick, S. *Motivational Interviewing.* New York: Guilford Press, 1991.

Miller, W.R.; Zweben, A.; DiClemente, C.C.; and Rychtarik, R.G. *Motivational Enhancement Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals With Alcohol Abuse and Dependence.* Project MATCH Monograph Series. Vol. 2. DHHS Pub. No. (ADM) 92-1894. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism, 1992.

Nowinski, J.; Baker, S.; and Carroll, K. *Twelve Step Facilitation Therapy Manual: A Clinical Research Guide for Therapists Treating Individuals With Alcohol Abuse and Dependence.* Project MATCH Monograph Series. Vol. 1. NIH Pub. No. 94-3722. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism, 1992.

Donovan, D.M., and Mattson, M.E. Alcoholism treatment matching research: Methodological and clinical issues. *Journal of Studies on Alcohol* Supplement 12:5–14, 1994.

Sobell, L.C., and Sobell, M.B. Timeline follow-back: A technique for assessing self reported alcohol consumption. In: Litten, R.Z., and Allen, J.P., eds. *Measuring Alcohol Consumption: Psychosocial and Biochemical Methods.* Totowa, NJ: Humana Press, 1992. pp. 41–72.

Strenio, J.L.F.; Weisberg, H.I.; and Bryk, A.A. Empirical Bayes estimation of individual growth curve parameters and their relationship to covariates. *Biometrics* 39:71–86, 1983.

Zweben, A., and Cisler, R. Composite outcome measures in alcoholism treatment research: Problems and potentialities. *Substance Use and Misuse* 31:1783–1805, 1996.

Causal Chain Analysis

Richard Longabaugh, Ed.D., and Philip W. Wirtz, Ph.D.

ABSTRACT

The analytic strategies used to assess the theories underlying the a priori matching hypotheses share a core element in that they all employ a "causal chain analysis" to test their underlying theory. The aim of causal chain analysis is to identify the influence of the variables thought to mediate the effects of treatment on client outcomes. Thus, causal chain analysis is mediator analysis applied to understanding how treatment works. When the search for mediators is applied to treatment outcome studies, it involves subsets of potential variables that are descriptive of treatment and of client response to treatment. This chapter describes mediation analysis and its application to testing mediation of interaction effects. Three models for testing are identified and matching hypotheses are classified according to which models they employed in each causal chain analysis. A typology for classifying results of testing the matching hypotheses in subsequent chapters is offered. This typology includes the paradoxical result of finding a supportive causal chain in the absence of a matching effect.

The purposes of the causal chain analyses presented in this volume are to (1) provide further information that would either support the observed matching effect or challenge its credibility and (2) identify the variables contributing to an observed interaction or discover why the hypothesized attribute-treatment interaction did not occur. Thus, where matching hypotheses are supported, the aim of the causal chain analysis is to identify the process through which the interaction occurs. Delineation of the causal mechanism increases the credibility of the interaction, and the active ingredients of treatment so identified can be exported to other treatment interventions. When an attribute-treatment interaction hypothesis is not supported, the causal chain analysis permits identification of the source or sources of its failure, revealing unsupported assumptions about treatment processes.

Mediators Versus Moderators

By definition, a "mediated relationship" is one which is "dependent on, acting by, or connected through some intervening agency" (Webster's 1986). In its simplest form, if variable A (e.g., treatment modality) affects variable C (drinking outcome) indirectly (through the effect of variable A on variable B [e.g., treatment structure] and the consequent effect of variable B on variable C), then variable B (structure) is

The author listing of this chapter is alphabetical. Both authors contributed equally.

Richard Longabaugh, Ed.D.
Center for Alcohol and Addiction Studies
Brown University, School of Medicine
800 Butler Drive, Potter Building, Room 204
Providence, RI 02906
E-mail: Richard_Longabaugh@Brown.edu

Philip W. Wirtz, Ph.D.
Department of Management Science
George Washington University
2115 G Street NW, #403
Washington, DC 20052
Email: pww@gwu.edu

said to mediate the relationship between variables A (treatment modality) and C (drinking outcome). The term "mediator analysis" defines the formal process of empirically testing whether, for example, B mediates (or explains) the relationship between A and C.

The term "moderator" is intended to convey a quite different—but also in matching studies, quite important—phenomenon. Generically, if the magnitude of the relationship between variable A and variable C differs depending on the level of variable B, then B is said to moderate the relationship between A and C. If, for example, the effect of treatment (variable A) on percentage of days abstinent (variable C) differs depending on the level of anger manifested by the client at baseline (variable B), then anger is said to moderate the relationship between treatment and percentage of days abstinent.

The conjunction of mediators and moderators lies at the heart of the Project MATCH causal chain analysis. The matching hypotheses each represented a hypothesized moderator relationship. *The objective of the causal chain analysis was, for each such hypothesis, to scrutinize the logic underlying the purported moderator relationship, that is, to test the purported mediators of each hypothesized moderator relationship (i.e., each of the matching hypotheses).* In so doing, it is possible to both determine if the matching effects observed experimentally were supported for the hypothesized reasons and locate the locus of the failure in the causal chains of those matching hypotheses which were not supported.

In general, there are three common analytical approaches:

- An informal comparison of treatment-group slopes
- A formal test using the general linear model
- A formal test using a structural equation modeling approach

All three approaches involve comparing the models which include a purported mediator to models which do not include the purported mediator; the approaches differ in the rigor with which mediation can be tested and in the comprehensiveness of the underlying model.

The informal approach, which involves comparing the significance level of the attribute-by-treatment interaction before versus after controlling for one or more purported mediators, is the simplest to perform but provides no formal basis for concluding that mediation has occurred. The general linear model approach is elegantly suited to causal models characterized by a single attribute-by-treatment of interest; most of the Project MATCH hypotheses fall into this category. A structural modeling approach, which involves comparing constrained to unconstrained models, is ideal when testing for the simultaneous effect of several mediators (although it can be employed in simpler models as well).

Testing Mediation of Treatment Main Effects

Although Project MATCH focused on interaction effects, it is instructive to understand the procedure for testing mediation of main effects before attempting to broaden the scope to the more complex models involving mediation of interaction effects. Readers already familiar with the formal testing for mediation of main effects may safely skip this section.

Irrespective of the analytic strategy developed for testing mediation, the conceptual model for a full causal chain analysis involves the following steps (figure 1).

First, a defined treatment needs to be discriminated from a comparison treatment on the basis of measured treatment ingredients (link 1 in the diagram). For example, it may be hypothesized and observed that Twelve Step Facilitation (TSF) and Motivational Enhancement Therapy (MET) differ from one another in their emphasis on abstinence as a goal for treatment.

Second, this identified difference in treatment is related to a difference in client response to treatment. For example, clients receiving an abstinence focus are observed to achieve greater posttreatment abstinence than those receiving less of an abstinence focus (link 2 in the causal

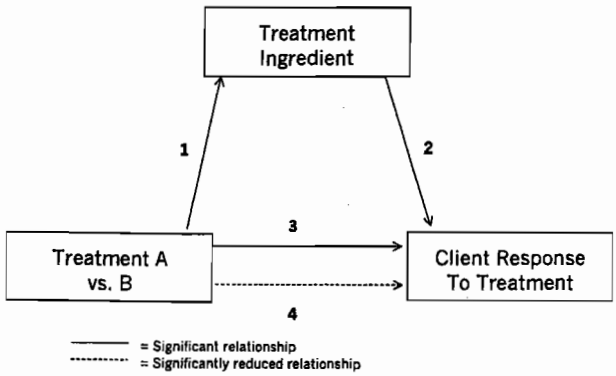

Figure 1. A generic causal chain analysis

chain). If it were also found that TSF was more effective than MET in producing client abstinence (link 3), then the next step in the causal chain analysis would be to examine whether the focus on abstinence accounts in whole, or in part, for TSF's incremental effectiveness (link 4). This would be demonstrated if it were observed that the significance of the main effect of TSF versus MET on abstinence was significantly reduced when the effect of focus on abstinence was partialed out from the relationship of treatment modality to client drinking outcome.

While this evidence is necessary for strong support for mediation, it is not sufficient evidence for causation. This is so because the entire causal chain analysis is conducted through an analysis of *associations* among the identified variables. The only variable that has been experimentally manipulated is treatment assignment. For unequivocal evidence of mediation, it would be necessary to experimentally manipulate each variable intervening between treatment assignment and drinking outcome to show that each in turn was dependent (or partially dependent) upon its predecessor in the causal chain. Such experimental manipulations cannot ordinarily be done in typical treatment outcome studies in which it is practical to manipulate only one variable, assignment to treatment condition. In a similar vein, some client variables of interest (age, gender, etc.) cannot be experimentally varied.

The case for causation can be strengthened by examination of competing mediator variable candidates. For maximum credibility, it is necessary to demonstrate that the identified mediating effect persists even after the effects of competing candidate variables have been removed. In our TSF versus MET example, this could be done analytically by showing that the mediator effects of focus on abstinence on the relationship between treatment and abstinence persisted even after the effects of competing candidates for mediation were removed. Because mediator analysis is correlational, and not experimental, the disciplined researcher can only conclude that the results obtained are consistent with the hypothesized mediation and the hypothesis has not been refuted.

Causal chain analyses in alcohol treatment outcome studies may quickly become complex, especially when a broad spectrum treatment such as Cognitive-Behavioral Therapy (CBT) is considered. CBT is often theorized to work indirectly to improve a client's drinking outcome (Longabaugh and Morgenstern 2000; Morgenstern and Longabaugh 2000). For example, CBT is intended to involve more time devoted to skills training than does MET and is expected to be predictive of differences in coping skills between CBT and MET clients at treatment completion. It is further hypothesized that improved coping skills will result in reduced drinking. Thus, for mediator analysis to support the theory, the causal chain analysis must test the links between: (1) treatment modality (CBT versus MET) and focus on skills training, (2) skills training and end of treatment improvement in client skills, and (3) improvement in client skills and drinking outcome. If it were found that (4) CBT was more effective than MET in reducing client drinking posttreatment, it would be necessary to show that (5) this greater effectiveness could be attributable to the causal chain, that is, when the effects of the linkages involved in the causal chain are removed from the relationship of treatment to drinking, this relationship is significantly reduced (figure 2). Finally, the evidence for mediation is strengthened if it can be shown that the mediated effect persists after ruling out alternative candidates for mediation, for example, the number of treatment sessions received in CBT versus MET.

Another complexity that frequently arises in causal chain analysis of treatment outcomes occurs because treatment effects are measured at

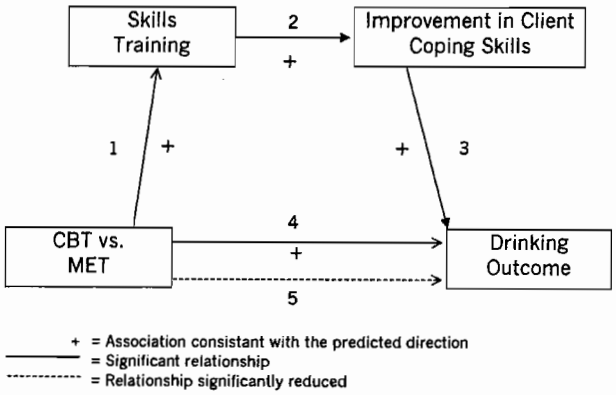

Figure 2. A hypothetical mediation analysis: CBT versus MET

intervals distally removed from the end of treatment, for example, clients are assessed a year after treatment completion. To be credible, a theory of treatment effectiveness is likely to postulate and test that client changes at the end of treatment are predictive of longer term drinking outcomes. Thus, the causal chain sequence specifies that assignment to alternative treatments leads to end-of-treatment differences in clients, themselves predictive of longer term drinking outcomes. In the TSF versus MET example described above, the causal chain would predict that the abstinence achieved by the client by the end of treatment would be predictive of reduced drinking during the year following treatment.

Not necessary for testing mediation, but highly desirable for demonstrating the generalizability of the mediator effects, would be demonstration that the effects of the putative mediator variable could be replicated in a new population. In Project MATCH, this opportunity was seemingly provided by the conduct of two independent studies—the aftercare treatment arm and the outpatient treatment arm. If it could be shown that the same mediator variables were operative in both arms of the study despite the difference in client populations and treatment context (i.e., standalone outpatient treatment versus aftercare following a more intensive inpatient or partial hospital experience), this would strongly support the belief that the mediator effects could be generalized across a broad spectrum of treatment settings.

Testing Mediation of Interaction Effects

While this approach has been widely adopted for testing simple mediation hypotheses, the formal test of a matching causal chain is one level more sophisticated, in that it requires testing for mediation of a *moderator* (i.e., interaction) effect rather than of *main* effect. The additional challenge imposed by a matching hypothesis is further heightened under a structural equation modeling approach, where detection of interactions involving latent constructs has proven particularly daunting.

When testing for a causal chain underlying main effects, it is necessary to show that the treatments compared differ in their implementation, for example, there was greater emphasis on an abstinence goal in TSF than in MET. When testing for causal chains to explain interactions, there is no single analogous step. Rather, for the causal chain to be supported, it is necessary that an interaction be observed in at least one step of the causal chain analysis and that the other links in the chain (which may involve either main or interaction effects) are also sustained.

Canonical Models

We describe three basic models which we believe are sufficient to account for the ways in which a causal chain can support a hypothesized client attribute-treatment interaction. Each of the models can be (and is in the following chapters) elaborated to involve considerably more complex causal chains. Nevertheless, at core they can be reduced to one of these three generic models, which we will refer to as canonical models. It should be emphasized that the value of distinguishing between canonical models is primarily descriptive.

Canonical Model 1

An element of the treatment process interacts with the client attribute.

The first model hypothesizes that treatments will differ in delivery of a specified treatment ingredient (figure 3). For example, TSF is hypothesized to have a greater focus on abstinence as a

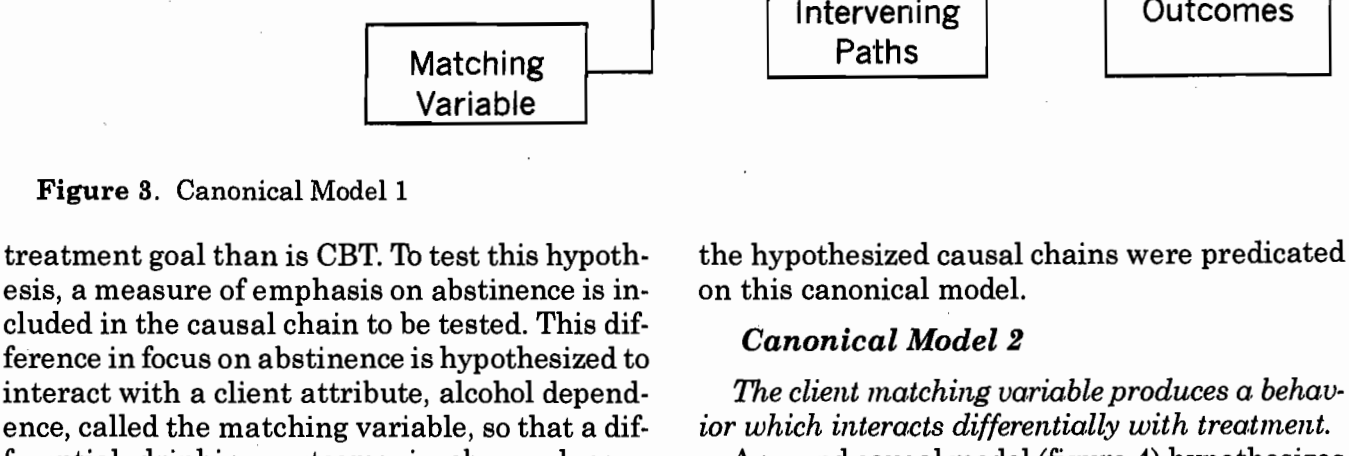

Figure 3. Canonical Model 1

treatment goal than is CBT. To test this hypothesis, a measure of emphasis on abstinence is included in the causal chain to be tested. This difference in focus on abstinence is hypothesized to interact with a client attribute, alcohol dependence, called the matching variable, so that a differential drinking outcome is observed as a function of the interaction of emphasis on abstinence. Clients with greater alcohol dependence benefit more from an emphasis on abstinence than clients with less alcohol dependence. Because TSF has a greater emphasis on abstinence than CBT, it is expected that TSF will be more effective with clients having greater alcohol dependence. Because CBT has a lesser focus on abstinence, it is expected that CBT will be more effective with clients with less alcohol dependence. In order for this causal chain to support an observed alcohol dependence by TSF versus CBT treatment interaction, at least three steps are necessary.

First, it needs to be established that TSF did involve a greater focus on an abstinence goal than CBT. Second, the *amount* of a client's dependence must be shown to interact with focus on abstinence to affect drinking outcomes, such that those with higher dependence had better drinking outcomes when their treatment had a greater emphasis on abstinence and/or that clients with low dependence had better drinking outcomes when their treatment had a lesser focus on abstinence. Third, it is necessary to demonstrate that this abstinence emphasis-dependence interaction accounted for the observed treatment modality-dependence interaction by showing that the significance of the latter interaction is significantly reduced when the effect of the former interaction is removed. Several of the hypothesized causal chains were predicated on this canonical model.

Canonical Model 2

The client matching variable produces a behavior which interacts differentially with treatment.

A second causal model (figure 4) hypothesizes that a client matching variable will lead to a certain kind of client behavior in treatment. This client behavior will interact with treatment modality to affect client outcome. MATCH hypothesized that clients differing in their anger prior to treatment would differ in their resistance to treatment, in that high anger clients would respond with greater resistance to treatment than would clients with low anger. To test this causal chain, indices of client resistance in treatment needed to be operationalized.

The next link in the causal chain was the hypothesis that this client resistance would interact with the type of treatment received, so that high anger clients assigned to MET would show less resistant behaviors than would high anger clients in CBT/TSF. The differential resistance to treatment would in turn result in different drinking outcomes, with those less resistant to treatment having better drinking outcomes than those more resistant.

In order for this causal chain to be supportive of an observed treatment modality-client attribute interaction, it is first necessary to show that clients high in anger differ from those low in anger in their behavioral resistance. Then, it is necessary to show an interaction of resistance with treatment modality such that high resistance clients in MET show greater reductions in drinking severity than do clients in any of the other treatment conditions. Finally, it is

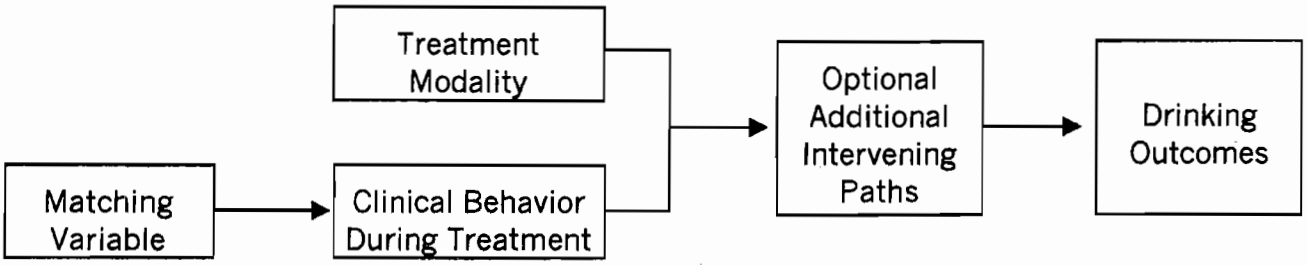

Figure 4. Canonical Model 2

necessary to show that the observed relationship between the combination of treatment modality-client anger and drinking outcome is significantly reduced when the effect of the treatment modality-client resistance interaction is removed.

Canonical Model 3A

Treatment modality interacts with a client matching variable to affect an intervening variable that is predictive of drinking outcome.

A third model postulates that a hypothesized treatment modality-client attribute interaction leads to a measured client response that in turn is predictive of drinking outcome (figure 5). For example, it was hypothesized that TSF would lead to greater AA participation by high meaning-seeking clients (but not by low meaning seekers) than would MET or CBT. AA participation in turn was expected to be predictive of better drinking outcomes. In this model, the combination of treatment modality and client attribute leads the client to engage in behavior otherwise unpredictable from treatment modality or client attribute by themselves. This changed behavior is in turn related to drinking outcome.

To test this causal chain, AA participation would be regressed on the product term of TSF versus MET and/or CBT by meaning seeking, and the product term would be observed to account for variance in AA participation over and above that accounted for by the two main effect variables, treatment assignment and meaning seeking. To validate the chain, AA participation would have to be shown predictive of drinking outcome. Finally, by partialing out the effect of AA participation on the initial meaning seeking by treatment assignment interaction, it would be expected that the latter effect would be significantly reduced.

Canonical Model 3B

The client matching variable modifies the delivery of the treatment modality to affect an intervening variable that is predictive of drinking outcome.

While it could be reasonably expected that one or the other of the three models described so far should be at the core of all of the causal chains hypothesized and tested, a second variation of Model 3 emerges. In this model, it is hypothesized that the client matching attribute will modify one of the treatments, such that the treatment actually changes as a function of the

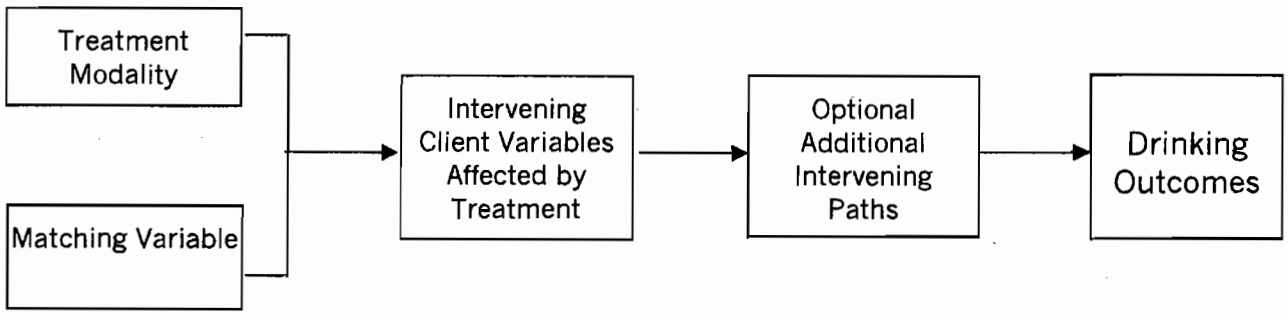

Figure 5. Canonical Model 3A. Treatment modality interacting with client matching variable to affect an intervening variable predictive of drinking outcome.

client attribute. This interaction of client attribute and treatment modality is predictive of a subsequent client behavior, which itself is associated with subsequent drinking outcome (figure 6).

As treatments are intended to be standardized across clients and delivered uniformly irrespective of client attributes, causal chain data supporting such a model would appear to challenge the integrity of the three treatments. However, this is not necessarily so, as there is room for such variability explicitly built into the CBT manual. The CBT treatment manual prescribes the delivery of eight core modules, but after completion of these modules, there is the option to select further modules from a menu of alternatives available. Thus, once the core sessions have been completed, the client and therapist can select from the remaining modules, the selection being dependent upon client choice or need. For example, a client might or might not be exposed to a mood management module.

In MATCH, it was hypothesized that CBT would be more effective then either TSF or MET for clients with greater psychopathology because CBT would provide high psychopathology clients with a greater exposure to treatment ingredients that reduce psychopathology. It was expected that this intervention would lead to a greater reduction in drinking than it would for low psychopathology clients.

In this scenario, CBT could be demonstrated to be different from the comparison treatment when delivered to a subset of clients, though not necessarily all (i.e., there would be an interaction observed in this first step of the causal chain). For example, if the mood management module in CBT was selected more often for (or by) high psychopathology clients, then this group would actually receive a somewhat greater exposure to CBT than would low psychopathology clients. Consequently, the CBT treatment for high psychopathology clients would differ from the CBT treatment provided for low psychopathology clients as a function of the psychopathology matching variable. This would establish a differential treatment experience for the subset of clients for whom it was hypothesized that CBT would be more effective. For this reason, it might be expected that a treatment-modality-by-client-attribute interaction would arise as a consequence of the matching variable affecting the treatment delivered. (This is in contrast to Model 1 in which the standard delivery of the treatment has a differential effect on clients varying in the matching variable.) If it were then shown that: (1) focus on mood management was predictive of reduced psychopathology for either all clients or just those psychopathologically impaired and (2) improved mood management was associated with better drinking outcomes either for all clients or just those psychopathologically impaired, this would provide evidence for the causal chain.

Summary

One or more of these canonical models are embedded in each of the hypothesized causal chains (table 1). Usually the causal chain is more complex and involves other linkages as well. In order for a causal chain to be supported, it is necessary that at least one interaction occur somewhere in the causal chain (which might be the original interaction affecting the mediator variable) and that connecting effects be carried

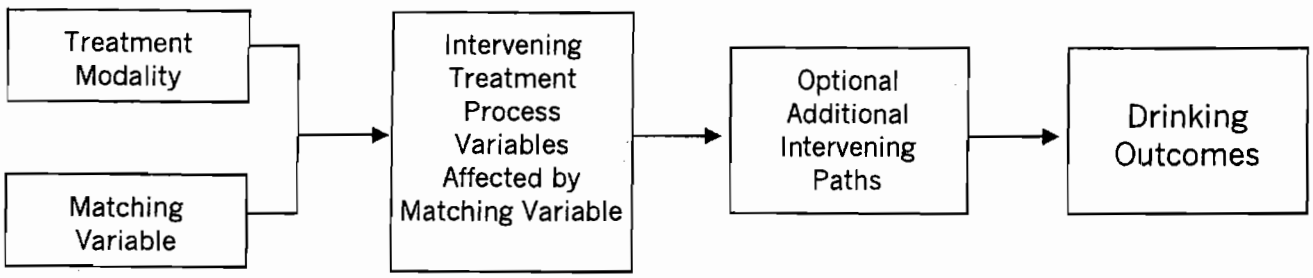

Figure 6. Canonical Model 3B. The client matching variable modifies the treatment modality to affect an intervening variable that is predictive of drinking outcome.

Table 1. Hypothesized mediator variables and Canonical Models for each causal chain

Matching variable	Hypothesized mediator	Canonical model*
Alcohol Dependence	Emphasis on abstinence	1
	AA involvement/Attendance	1
Alcohol Involvement	Within-Treatment Drinking	1&2
	Within-Treatment Self-Efficacy	1
	Negative Consequences of drinking	1
Cognitive Impairment	Amount of Therapy	1/3B
	Amount of Therapeutic structure	1
	AA involvement/Attendance in Treatment	1
Psychopathology and Psychiatric Severity	Emphasis on psychopathology	1/3A/3B
	Reduction in psychiatric symptoms	1/3A/3B
Sociopathy/ASPD	Amount of Therapeutic structure	1
	AA attendance	2
	Working Alliance	2
	Anger reduction	2,1
A versus B Typology	Amount of Therapeutic structure and cognitive change	1
	Working Alliance	3A
	Change in psychopathology	3A
Anger	Taking Steps	2
	Problem recognition	2
	Working Alliance	3
Conceptual Level	Amount of Therapeutic structure	1
Meaning Seeking	Working Alliance	3A
	Client satisfaction with Treatment	3A
	Percentage of Treatment sessions attended	3A
	AA involvement/attendance	3A
Religiosity	Working Alliance	3A
Interpersonal Dependency	Satisfaction with Treatment	3A,3A
	Percentage of Treatment sessions attended	3A,3A
Gender	Role Demands	1 & 2
	External Stresses	1 & 2
	Psychiatric severity	1 & 2
	Depression	1 & 2
	Self-esteem	1 & 2
Motivational Readiness	Working Alliance	3A
	Compliance	3A
	Change Processes	3A
	Increased Readiness to change	3A
	Self-efficacy for abstinence	3A
Problem Recognition	Working Alliance	3A
	Change in Problem Recognition	3A
Self-Efficacy	Change in Self-Efficacy	3A
Temptation minus Confidence	Change in Temptation minus Confidence	3A
Network Support for drinking	Coping with social pressure to drink	1
	Change in network support for drinking	1
	AA attendance	1
Prior AA	Working Alliance	3A
Social Functioning	Change in Social Functioning	3A

*Canonical models separated by a comma show that one model precedes the other in the causal chain. Models separated by an & were conceptualized as occurring simultaneously. A / between canonical classifications indicates that the mediational model allowed alternative canonical pathways through which the mediating effect could occur. Consensus classifications were made by the two authors on the basis of discussion following independent review and classification. Some causal chains may be misclassified due to a lack of total explicitness in their descriptions.

through the causal chain from treatment implementation to drinking outcome.

In the full causal chain underlying a matching hypothesis, differences in treatment delivery should not be assumed but should be tested. If this step is not conducted, one would not know if the lack of a differential client response is due to incorrect theory or if the clients in the treatments compared did not receive different doses of the putative active ingredients. Furthermore, some theories underlying the hypothesized client-treatment interactions make use of intervening client variables, such as increased motivation, as variables mediating the effect of treatment on drinking outcome. These need to be measured and their mediating influence tested as well. Finally, drinking outcomes are measured over one year following treatment completion. If a client-treatment interaction is observed to change over time (as some did), then the choice of time period for measuring drinking outcome becomes less obvious. As treatment effects have generally been found to be transient, the case can be made for measuring the drinking variable immediately following the end of treatment, rather than at a time further removed, when the effects of treatment may have been obscured by posttreatment events.

To conclude this section, we hypothesize that a moderating effect will be explained by a causal chain only when at least one of the connecting links in the causal chain involves an interaction and the remaining links involve either main effects or interactions. However, even if these conditions are met, only a complete test for mediation will establish whether the causal linkages are explanatory.

Paradoxical Findings

Under limited circumstances, causal chain linkages in the absence of client-treatment interactions may be observed. While the hypothesized client attribute-treatment modality interaction may not be detected, it may still be possible to find evidence for the causal chain developed to explain the expected but unobserved interaction. One explanation for this seemingly paradoxical finding is that different treatment modalities provide different pathways for clients with the same attribute to achieve equally good drinking outcomes.

A MATCH example may help to make this point clear (Kadden et al., this volume). It is hypothesized that CBT will be more helpful than TSF in reducing drinking for clients with antisocial personality disorders (ASPD). The logic behind this client-treatment modality interaction is as follows: ASPD clients are frequently lacking in coping skills. Drinking is their way of coping in situations where they lack the needed coping skills. Therefore, ASPD clients need training in coping to achieve their goals and avoid drinking. CBT trains clients to acquire and use these skills. When such skills are enacted instead of drinking, the ASPD client will have a good drinking outcome. Thus, the causal chain is: CBT involves coping skills training. Coping skills training leads to the acquisition of coping skills. Coping skills acquisition for ASPD clients will produce an interaction such that ASPD clients will have better drinking outcomes than non-ASPD clients.

Hypothetically, however, it could also have been hypothesized that TSF will be more helpful than CBT for ASPD clients, resulting from a different causal chain. TSF exhorts clients to attend AA. AA participation leads to better drinking outcomes. Generally, ASPD clients would be less likely to attend AA than non-ASPD clients. Nevertheless, because of the TSF emphasis on attending AA, more do so when assigned to TSF. ASPD clients are unlikely to receive acceptance and esteem from people in their natural social network because of the lack of consideration by the ASPD of the feelings and needs of others. However, in AA this behavior would not be cause for rejection because alcoholism is considered to be at the core of all such antisocial behaviors. The client may thus find acceptance in AA that would not be available outside of AA. This acceptance could lead to an increase in the ASPD client's concern and feelings for others. The reciprocated concern brings about increased self-esteem; the mutual concern increases social cohesiveness. The acceptance that ASPD clients find in AA, conditional upon their wanting to quit drinking, results in their increasing abstinence. In this scenario, TSF leads to AA participation, AA participation for the ASPD leads to greater

self-acceptance conditional upon sobriety, which in turn leads to better drinking outcomes.

Both of these causal chains could be equally operative—the former in CBT and the latter in TSF. However, if the only causal chain developed to support the hypothesized ASPD-treatment modality interaction was for CBT, the hypothesized client attribute-treatment interaction might not appear. This would not be because the underlying theory regarding the effect on drinking of ASPD client's acquiring coping skills was wrong, but because the causal chain underlying TSF's affect on ASPD clients was also right. In this scenario, TSF would have affected ASPD clients through an alternative (but unidentified) causal chain process that led to equally good drinking outcomes for ASPD clients treated in TSF. If this were the case, then a causal chain analysis could support coping skills training as a correlate of better drinking outcomes in CBT, despite the observation that the hypothesized interaction of CBT by ASPD did not materialize. Thus, ASPD clients would have equally good drinking outcomes in CBT and TSF, but for different reasons.

In such an event it would be instructive to reexamine the treatment modality initially believed to be inferior for clients with this attribute to assess whether an alternative causal chain was responsible for the unexpected effectiveness of the TSF treatment for the ASPD client. Such a discovery would not undermine the theory underlying the unobserved CBT–ASPD interaction, but it would provide one explanation as to why the interaction did not emerge. Such a discovery would be a considerable achievement for theories trying to explain treatment effectiveness.

A Classification Typology for MATCH Results

The joint examination of both attribute-treatment interaction effects and causal chain analysis leads to one of four outcomes (table 2).

First, in the best of circumstances, the hypothesized attribute-treatment interaction has occurred and a supporting causal chain has been identified. In this case, the theory underlying the matching hypothesis is supported and the credibility of the matching effect is strengthened. Not only are we able to explain why the treatment is especially effective for a defined client, we are also able to extract the identified treatment ingredient for incorporation into another treatment, hypothesizing that the active ingredient will increase the effectiveness for this type of client in this other treatment as well.

Table 2. A fourfold typology for classifying MATCH results

Matching hypothesis	Supporting causal chain	
	Identified	Not identified
Matching prediction supported	yes/yes	yes/no
Matching prediction not supported	no/yes	no/no

Second, it may be observed that the hypothesized attribute-treatment interaction has not occurred nor has the hypothesized causal chain been supported. In this circumstance the breakdown in the linkage in the causal chain will indicate whether the failure may be attributable to a failure in treatment implementation (e.g., where the treatments do not differ in their putative active ingredients) or a failure in the theory as to how treatment affects drinking (e.g., where the observed differences in treatments do not affect the clients as expected) (Finney and Moos 1992). Depending upon which of these is the case, future research may seek to strengthen the treatment or develop a better theory regarding treatment effects.

Third, it may be observed that the hypothesized attribute-treatment interaction has occurred, but the causal chain has not been supported, that is, the prediction is supported but not for the reasons we postulated. In this instance, the credibility of the observed interaction is undermined, with the belief that it may be the result of a type 1 error. Moreover, in terms of theory development, nothing has been added to the existing knowledge base. We may believe that we have produced an attribute-treatment interaction but we know not how. We are unable to

identify the active ingredients of treatment to be included in subsequent research.

Finally, it may be that a hypothesized attribute-treatment interaction does not occur even though the causal chain analysis suggests that it should have materialized. In this case, it is likely that alternative paths for reaching the same end are present in the comparison treatment but are unidentified. The next step for research would be to seek out these alternative pathways to better outcomes. This would permit the implementation of different treatment modalities with equal effectiveness by making within-modality adjustments to address the salient characteristics of these clients.

Acknowledgment

This project was supported by a series of grants from the National Institute on Alcohol Abuse and Alcoholism as part of the Cooperative Agreement on Matching Patients to Alcoholism Treatments.

References

Baron, R.M, and Kenny, D.A. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51(6):1173–1182, 1986.

Finney, J.W., and Moos, R.H. Four types of theory that can guide treatment evaluations. In: Chen, H.-T., and Rossi, P.H., eds. *Using Theory to Improve Program and Policy Evaluations.* New York: Greenwood Press, 1992. pp. 15–27.

Holmbeck, G.N. Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology* 65:599–610, 1997.

Longabaugh, R., and Morgenstern, J. Cognitive-Behavioral Coping-Skills Therapy for Alcohol Dependence: Current Status and Future Directions. *Alcohol Research and Health* 23(2):78–86, 2000.

Morgenstern, J., and Longabaugh, R. Cognitive behavioral treatment for alcohol dependence: A review of evidence for its hypothesized mechanisms of action. *Addiction* 95(10):1475–1490, 2000.

Webster's New World Dictionary, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1986.