

For All You Do, This Article is For You: Thoughts on Optimizing and Evolving Treatment Evaluation

Laura Mills

Pine River Institute

Sarah Lewis

Center for Research, Assessment, and Treatment Efficacy

Abstract

Outdoor Behavioral Health programs and Residential or Therapeutic Boarding School treatment approaches have made great inroads when it comes to measuring outcomes. The information from their efforts can foster accountability to clients, evaluation-informed treatment and program decisions, validation for the hard work of agency staff, and demonstration of value to governance and funders. This paper was written to celebrate the dedication shown by agencies to engage with client outcome-oriented research and also to provoke thought as to how staff, agencies, and multi-agency collaborative teams can optimize and strategically plan their evaluation efforts.

Keywords: youth treatment evaluation, research methods, evaluating youth therapy, outdoor behavioral health, residential treatment for youth.

There is a wealth of literature in which theoreticians, researchers, and practitioners articulate the need to evaluate therapeutic effectiveness (e.g., Beautler, 2001; Horowitz, Lambert, & Strupp, 1997; Kazdin, 1996; Lambert, Hansen & Finch, 2001). Possibly in response to this need, or maybe arising from the dedication to being accountable to clients, therapeutic outcome measurement is starting to permeate contemporary clinical practice. Evaluating the success of clients benefits everyone. Clients and their families benefit from evaluation-informed program improvements and treatment planning; staff can validate the hard work they engage in every day; and financial stakeholders can justify their considerable investment, be it in one client or in an entire program. That said, evaluation must be undertaken with careful thought, thoughtful planning, and plan-driven execution. In this paper, we celebrate the evaluation efforts championed by Outdoor Behavioral Healthcare (OBH) Council and National Association of Therapeutic Schools and Programs (NATSAP) and discuss considerations for continued evaluation efforts.

Careful Thought

Considering the myriad tools, research methods, and analytic processes available, it is a daunting task to begin the process of evaluating treatment effectiveness. Sometimes, research questions are driven by external researcher interest, existing protocols that are already utilized in a particular field, or even the contents of a measurement instrument. When these are the foundations for evaluation protocol, programs run the risk of gathering information that does not resonate with their therapeutic goals and objectives, their client's experience or presentations, or their programmatic outcomes.

On the other hand, when carefully planned research questions, determined by well-defined outcome goals drive evaluation efforts, programs can succeed in making continual, evidence-based improvements to better serve their consumers. By guiding evaluation efforts this way, the research protocol and findings can align with treatment intentions and inform specific areas of quality assurance and improvement. Some questions to ask of clients, families, staff, and other stakeholders are, *'In what ways do we think our clients change?'*; *'How much do we think our clients will change?'*; *'For how long do we expect changes to be sustained?'* *'What are the overall goals and objectives of our programming?'* We highly recommend all programs engage in this intentional curiosity, for all of the reasons described above, as well as to foster agency-wide buy-in and ownership of evaluation processes and results.

TREATMENT EVALUATION

The OBH Council and NATSAP have done impressive work to overcome the daunting task of starting the evaluation process. They have championed initiatives that shepherd data collection for member programs within their outdoor behavioral healthcare and residential treatment / therapeutic boarding school (RTBS) approaches. Their efforts have resulted in selecting evaluation tools and liaising with relevant licensing agencies; sourcing infrastructure to member programs by way of an online data collection platform (i.e., Bestnotes CRM); providing support for programs to implement evaluation protocols (via the Remote Research Director Service of Petree Consulting Inc.); and providing resources to the OBH Center, housed at the University of New Hampshire, to securely house and manage data, support data analyses and dissemination findings, and examine the outcomes of member programs along with comparative cross-agency aggregate results.

The collaboration between industry affiliated scientists and OBH / RTBS practitioners represents an impressive merger. It allows for an empirical assessment of treatment outcomes such that programs can examine the quality and impact of their services while also identifying areas in need of improvement. This type of evaluation also validates treatment effectiveness to insurance, accreditation, and policy-making stakeholders, thereby sustaining the presence of OBH and RTBS approaches within the context of mental health service delivery.

Thoughtful Planning

In order to draw valid inferences regarding such a complex phenomenon as youth's response to treatment, rigor is required in the design and the execution of the research. The science of research methodology involves decision-making at every step of the process to assure that ambiguity is minimized and conclusions are based on valid findings. It is important to consider all facets of the process, because seemingly trivial issues can have profound impacts on the conclusions of the research. For example, a close look at post-treatment contact time-points should spur such questions as, *'If we measure outcomes right at the end of therapy, will improvements be due to our intervention or because the youth is happy to have completed the program?'* and, *'How long after therapy can we still associate outcomes to our process?'* Without addressing such questions, scientists may draw erroneous conclusions from their findings. Further, research results are influenced by innumerable factors, including but not limited to the subtle variations in the data collection process (e.g., lack of standardization in how participants are contacted, enrolled, and consented into the study). Thus, methodological science instructs us to consider all facets of the study design so that when a particular finding is obtained, we can reach an unambiguous conclusion (Kazdin, 2003). Naturally, the process of developing these standard protocols would provide a platform for excellent dialogue at OBH and RTBS knowledge exchange events.

Plan-Driven Execution

When agencies or agency groups develop standard protocols that optimize accurate interpretation of evaluation results, their curiosity and strategy is likely to evolve. For example, program stakeholders may be interested in the specific mechanisms of treatment that foster success (e.g., treatment progression or completion, duration, specific treatment goals or skill development, etc.). These types of questions can only be answered in contexts of treatment and evaluation fidelity. One way to ensure treatment fidelity is to adopt an evidence-based therapeutic approach. Following the example of medicine (Institute of Medicine, 2001), academic psychologists have spearheaded the move towards evidence-based practice in psychology (EBPP; American Psychological Association, 2005). Indeed, policy mandates for EBPP exist in many settings. For example, agencies providing federal-, state-, or insurance-funded mental health services are required, in some states, to provide evidenced-based treatments as part of their service delivery (Office of Applied Studies, 2008). EBPP helps ensure that clients are provided with therapeutic services that are known to be effective. However, this should not preclude the need to examine agency-level treatment outcomes. In fact, EBPP is the perfect platform to foster evaluation leading to ever increasing quality service delivery and wellness of clients, along with material for expanding knowledge in the field.

Another way to evolve agency evaluation is through collaboration with researchers whose interests align with the agency evaluation needs. For example, an agency serving youths with eating disorders would benefit from identifying a research collaborator with expertise in this area. This individual would

TREATMENT EVALUATION

consult on the current knowledge base, work with the agency to choose appropriate measurement constructs and instrumentation decisions, and help isolate and measure specific outcomes or mechanisms of change. This type of collaboration should be developed and nurtured with a long-term relationship in mind as opposed to study-specific involvement. Well-aligned, ongoing relationships allow for in-house development of sophisticated research which includes planning, implementation, data analysis, interpretation, and reporting, but also provides multiple opportunities for dissemination of findings and knowledge advancement.

Different types of methodological designs are used to answer different questions and should include considerations of population of inquiry, state of knowledge, and hypothesized relationships among variables (to name a few). Each type of design allows the investigator to answer different questions with different levels of rigor, and all designs have their strengths and weaknesses. For example, research in which the investigator has tight control over most study parameters tends to be less generalizable to the general population, and conversely, designs with less experimenter control tend to study heterogeneous samples (i.e., the types of clients more often seen in real-world treatment agencies). Four main types of methodological designs, with relevant strengths and limitations, are briefly reviewed below.

Case Studies

Imagine. You have an interesting client and want to share what you learned over the course of working together. You write a paper depicting the client's clinical presentation, as well as your treatment approach, challenges, solutions, and insights.

Benefits. Case studies help us gain insight into the behavior of individuals. The clinical presentation of clients to OBH and RTBS settings is nothing if not complex, so it's likely that other treatment providers will benefit from exploring parallels between their typical clients and your case. That said, case studies also provide a unique opportunity to delve into clients that are distinct or present with atypical historical, developmental, or personality characteristics. This provides a platform to share knowledge or describe modifications to existing therapeutic interventions, perhaps that of stepping outside of usual or well-known treatment parameters. Further, case studies can be used to supplement and enliven treatment-relevant information gathered from a larger study group. Finally, case studies remind us that there are real people with real problems, who come to treatment providers with real hopes of feeling better.

Drawbacks. The primary shortcoming of case studies is that there is no way to determine if the treatment is responsible for changes in the client, as this approach relies heavily on anecdotal information. If the client's health and behavior improved, it may have happened without therapy, or due to other changes in their environment. Secondly, treatment providers can't assume that the approach taken for the case will have a similar effect on their clients, even if there are striking similarities between the two (referred to in the scientific community as a lack of replication). Finally, case studies tell a story from the perspective of the author which invites the possibility of bias, no matter how impartial the author attempts to be.

Quasi-Experimental Studies

Quasi-experiments describe research in which study participants *already exist* within a specific group that is being studied (Kazdin, 2003). Sexual assault victims, adults on the autism spectrum, or teenagers with substance use disorders are a few examples. Both single-site and multiple group studies, described below, fall in this category.

Single-Site Studies. Imagine. You have data on the health and behavior of your clients measured pre-treatment, post-treatment, and at one year follow-up. You have this information for most of your clients over the most recent three years and can see that there are decreases in problematic behavior and increases in health and prosocial behavior for the group as a whole. Your study can help you understand the relationship between time and your outcome(s) of interest, with the understanding that within that time, the client experienced therapy.

Benefits. Pretest-posttest designs are a great place to start evaluating the effectiveness of programs

TREATMENT EVALUATION

(Kazdin, 2003). Results help validate staff for their hard work and allow programs to identify areas for evaluation-informed program planning and improvement. Another benefit is that pretest-posttest designs tend to give us information about populations that interest us – namely, our clients. This type of work can also be used to communicate with clients and families the types of outcomes they can reasonably expect from the program. When agencies engage with this type of evaluation, it demonstrates a dedication to being accountable, both for the well-being of clients and for providing the highest quality and most effective therapeutic services.

Drawbacks. The primary drawback to single-site quasi-experimental studies conjures a phrase common to the field, *'correlation does not imply causation'*. In other words, we can measure change but cannot conclude that the treatment program or intervention caused the change (Kazdin, 2003). Also, in the absence of a comparison condition or control group, we cannot rule-out that other factors (e.g., historical effects, maturation, instrumentation, statistical regression) accounted for the pattern of findings in the study.

Comparison (Multiple Group) Studies.

Imagine. You have a group of youths who completed your program over the span of two years. You also have a list of similar youths who applied to your program but never enrolled (comparison group). You administered health and behavior surveys to all participants at the time of application and at treatment completion (and a similar time frame for the no-treatment group). You can now compare whether symptomatic changes over time are different for the treatment versus the no-treatment group.

Benefits. Primarily with quasi-experiments, if the outcomes are different for the different groups, therapy is one of the factors that *may* be responsible for these differences (Kazdin, 2003). Another benefit is that participants in quasi-experiments often resemble the types of clients seen in the field, as they tend to have varied clinical presentations, histories, and needs. Quasi-experiments can be conducted in 'real world' treatment programs – a factor that benefits clients greatly. Quasi-experiments are also a great way for an agency with more than one program to compare programs and utilize outcomes for treatment planning, knowledge exchange, and quality improvement. In fact, with this type of design, one could compare different groups across any number of factors such as gender, age, treatment engagement, comorbid disorders, presenting problems, and a host of variables that might promise to foster change.

Drawbacks. One *cannot* conclude that treatment (or whichever comparative variable) was the main agent of change in a quasi-experiment. This is because groups may be fundamentally different (given the lack of random assignment to experimental conditions), and *that* difference may be the cornerstone of change. Using our example, if youths who applied but didn't enroll in your program declined treatment due to family financial status, socioeconomics may be the factor that accounted for differential outcomes between the treatment and the no-treatment groups. In other words, with this type of design, one cannot conclusively rule out alternative explanations for research findings. Researchers can, however, minimize potential confounds by using matched samples (where multiple groups are matched on relevant characteristics such as age, gender, psychiatric condition and severity, etc.), as well as the use of sophisticated statistical techniques (e.g., utilizing co-variables or structural modeling approaches). For example, there are procedures to minimize the possibility that income was the change factor, if income is a measured variable (Gaudiano, Dalrymple, Weinstock, & Lohr, 2015).

Experimental Research (Group designs)

Imagine. Two hundred youths want to attend your next treatment session. All youths are 16-17 years old and have experience with substance abuse, but no history of mental health issues, no behavioral or relationship problems, and no prior treatment. You measure substance use, health, and behavior of all 200 youths and then randomly assign each person either to the treatment or the no-treatment group. You again measure the health and behavior of all 200 youths at regular intervals.

In experimental research, the investigator is manipulating (controlling) the delivery of a procedure or treatment. A sample of people are selected from the general population and these individuals are *randomly* assigned to experimental conditions; thus, the groups are functionally equal. The population

TREATMENT EVALUATION

is similar across defined characteristics such as age, singular presenting problem, treatment-seeking behavior, etc. Random assignment gives confidence that any individual differences among participants is evenly dispersed across the groups. Now, the only difference between groups is the experimental delivery of a treatment intervention (versus a placebo or control group). One can now examine whether the treatment condition reliably results in changes in participants, and to what extent.

Benefits. Experimental designs are the only types of research where we can conclude that the manipulated variable (treatment) caused differences between groups on the measured outcome (Kazdin, 2003). This is because for these equal groups, treatment was the only functional difference, so it *must* have been the cause of different outcomes. One can even infer that those changes can be generalized to the relevant population.

Drawbacks. Limitations associated with RCTs include the prohibitive cost as they involve extensive training of staff, standardization across research conditions, and oversight by the primary investigator. Further, this type of design tends to require highly operationalized treatment and comparison control conditions, often specified in a manual. Treatment delivery is standardized and therapists are regularly assessed for their adherence to study treatment (fidelity to the therapeutic model). Second, RCT sample selection is based on carefully defined population characteristics and thus quite specific inclusion and exclusion criteria for the study. Because of this, the population tends to look different from clients seen in real-world settings. Although RCT results can be generalized, generalization only extends to the narrowly defined population. This means there may be little *external validity*, or in other words, we may not expect the findings of a RCT to apply to all clients in OBH and RTBS settings. Following our above example, the findings of the experiment would apply to youth, age 16-17 years old, with a history of a specific type of substance abuse, in conjunction with no mental health, behavioral, or relational problems, and no prior treatment. Third, the sample selection may not be possible based on the types of clients who present at OBH and RTBS programs. Lastly, and most importantly, there may be ethical problems in withholding treatment from some individuals for the sake of conducting research, particularly once it is clear that the intervention is effective and could improve the lives of clients.

Final Thoughts and Recommendations

The intent of this paper was threefold. First, we wanted to honor the strides taken in the OBH and RTBS fields: by agencies collecting the data, organizations facilitating this work, and the scientists turning this work into knowledge. Second, we hoped to raise awareness of ways to start or evolve current evaluation practices, with careful thought, thoughtful planning, and plan-driven execution. Third, we provided a brief overview of the types of studies that can be conducted such that stakeholders can consider the benefits and drawbacks to each when they are planning their own evaluation or learning from conference presentations, research papers, and popular media reports. To summarize, we provide four recommendations for research endeavors:

1. Ensure that your evaluation protocol resonates with your agency's mission-driven values and therapeutic objectives.
2. Pay careful attention to factors that can erroneously account for treatment outcomes and attempt to minimize alternative explanations.
3. Share learnings about the process and outcomes of evaluation endeavors at knowledge exchange events to foster a broad perspective of dialogue and learning.
4. Anticipate new questions and strategic partnerships that will foster ever-increasing sophistication of your evaluation, and thereby, the quality of service delivery within your agency.

References

- American Psychological Association (2005). Report of the Presidential Task Force on Evidence-Based Practice available online at <http://www.apa.org/practice/ebreport.pdf>.
- Beutler, L. E. (2001). Comparisons among quality assurance systems: From outcome assessment to clinical utility. *Journal of Consulting and Clinical Psychology, 69*(2), 197-204.
- Gaudiano, G., Dalrymple, K., Weinstock, L., & Lohr, J. (2015). The Science of psychotherapy: Developing, testing, and promoting evidence-based treatments. In S Lilienfeld, S Lynn, & J Lohr (Eds.), *Science and Pseudoscience in Clinical Psychology*, New York: Guilford Press.
- Horowitz, L., Lambert, M., & Strupp, H. (Eds.) (1997). *Measuring patient change in mood, anxiety, and personality disorders: Toward a core battery*. Washington, DC: American Psychological Association Press.
- Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.
- Kazdin, A. E. (1996). Evaluation in private practice: Introduction to the series. *Clinical Psychology: Science and Practice, 3*(2), 146-159.
- Kazdin, A. E. (2003). *Research design in clinical psychology* (4th Ed.). Boston, MA: Allyn & Bacon.
- Lambert, M. J., Hansen, N. B., & Finch, A. E. (2001). Patient-focused research: Using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology, 69*(2), 159-172.
- Office of Applied Studies. (2008). Results from the 2007 *National Survey on Drug Use and Mental Health: National findings*. Rockville, MD: Substance Abuse and Mental Health Services Administration.