Journal of Experiential Education
2016, Vol. 39(1) 5–14

DOI: 10.1177/1053825915607535
jee.sagepub.com
SAGE

Why Randomized Trials Are Challenging Within Adventure Therapy Research: Lessons Learned in Norway

Leiv Einar Gabrielsen[1], Carina Ribe Fernee[1], Gunnar Oland Aasen[1], and Leif Torvald Eskedal[1]

Abstract

There are few high-quality studies using randomized controlled trials (RCTs) in the adventure and wilderness therapy literature. Thus, a unison call is heard for more such studies to be carried out. This article presents a Norwegian wilderness therapy research project that planned to incorporate this "gold standard" that is regarded as the most scientific and rigorous approach available. We did not succeed. Mounting challenges led us to discard the RCT altogether and select other methodologies. Here, we account for the ethical, health outcome, practical, and empirical obstacles that we encountered when attempting to randomize at-risk adolescents into experiment and control groups. Our conclusion is that although RCTs may be superior in some aspects, they hold the potential to become bad science when the interventions are as complex and multi-faceted as adventure and wilderness therapy programs.

Keywords

adventure therapy, wilderness therapy, randomized controlled trial

A recent article in this journal calls for the increased use of comparison groups in adventure therapy studies (Norton et al., 2014). This wish is echoed throughout the literature on quantitative assessments of adventure and wilderness therapy programs (Becker, 2010; Clark, Marmol, Cooley, & Gathercoal, 2004; Gass, Gillis, & Russell, 2012; Wilson & Lipsey, 2000). Randomized experiments, also called randomized controlled trials (RCTs), are generally considered the gold standard of comparison

[1]Sorlandet Hospital Health Enterprise, Kristiansand, Norway

Corresponding Author:
Leiv Einar Gabrielsen, Department for Child and Adolescent Mental Health, Sorlandet Hospital Health Enterprise, Postboks 416, 4604 Kristiansand, Norway.
Email: Leiv.Einar.Gabrielsen@sshf.no

group designs (Shadish, Cook, & Campbell, 2002). This research approach is considered to produce evidence and puts forth claims of facts, truth, reality, and knowledge, words that Hacking (2000) calls elevator words.

We support the call for increased scientific verifiability within the field of adventure and wilderness therapy research. However, there are some pitfalls along the way, and these should be taken into account. This article is based on our experiences of an attempt to apply sample randomization to a clinical wilderness therapy research project in Norway. These experiences convinced us to abandon this approach, primarily due to ethical, health outcome, practical, and empirical consequences.

The RCT

When planning to scientifically document the efficacy or effectiveness of health interventions, the RCT is generally the design of choice (Shadish et al., 2002). The reason for this preference is that this approach, at least in theory, eliminates any other possible explanation of the outcome than the intervention studied. The RCT basically relies on establishing two or more groups of study subjects that are "identical" and then introducing the independent variable (e.g., the drug, the clinical procedure, the intervention) to one of the groups and not to the other. The group that is exposed to the independent variable is often labeled as the experimental group, and the group that is not exposed is labeled as the control group. The data that one wishes to study, that is, the dependent variables, are collected at a minimum before and after the intervention. The data gathering procedure is typically identical for both groups. The rationale for the RCT is obvious. In principle, when the only thing that separates groups is whether they have been subjected to the independent variable, any differences in post-intervention data are likely due to this variable alone. Hence, inferences can be made about causality.

The RCT inherently relies on the randomization of participants into groups. This is a complex procedure; however, conceptually, it is similar to tossing a coin. The randomization process minimizes allocation bias, and the threats from confounding variables are kept to a minimum. Confounding variables, or confounders, pose a serious threat to non-RCT studies. These variables, which are often unconsidered, correlate with both the dependent and independent variables. In studies without control groups, we are to a certain degree left to speculate about causal explanations for apparent health improvements. Continuous clinical measures, observations, interviews, and case studies all provide valuable insights into procedural health changes and subjective experiences. Statistically speaking, however, in regard to documenting effects and efficacy, we are left in a twilight zone.

Experiences From an RCT in a Norwegian Wilderness Therapy Project

Background

Although most Norwegians view themselves as avid outdoor people, until recently, the use of outdoor life, in general, and wilderness therapy (WT) in particular, has remained

underexplored in the mental health services for at-risk adolescents in Scandinavia (Fernee, Gabrielsen, Andersen, & Mesel, 2015). This is now changing, partly as a result of the project discussed here. In short, we developed a Norwegian version of WT called *Friluftsterapi*™—which translates as "therapy in the open air." *Friluftsterapi*™ draws on the diverse international practice and theory of WT, which has been adapted to the Norwegian sociocultural context by integrating the *friluftsliv* tradition (Henderson & Vikander, 2007). *Friluftsterapi*™ can be defined as "a specialized approach to mental health treatment that combines individual and group-based therapeutic work with basic outdoor life, engaging participants through ecological, physiological and psychological processes" (www.friluftsterapi.com). *Friluftsterapi*™ is developed as a stand-alone, outpatient and intermittent treatment program at the Department for child and adolescent mental health at Sorlandet Hospital in Southern Norway. Norwegian national health care is public, which ensures equal access to treatment. WT is offered as a voluntary group treatment for adolescents aged 16 to 18 years that have been referred to mental health services due to challenges such as depression, self-harming behavior, anxiety, and adjustment disorders. The WT program is facilitated over a 7- to 10-week period with a closed group structure, consisting of 8-day sessions with roughly 1-week intervals and two wilderness hikes of 3 and 6 days. A multi-disciplinary therapist team made up of three mental health professionals with outdoor competence accompanies the group of 8 to 10 adolescents for the duration of the treatment. The group is ideally an equal mix of boys and girls, which should also be reflected in the therapist team. Besides the first introductory day, the entire intervention takes place outdoors.

The Research

The ongoing accompanying research utilizes quantitative and qualitative techniques that are incorporated into a mixed-methods design. Starting with modest sample sizes, we are currently establishing a multi-center research project that will allow us to enroll more participants across the country. The WT programs will be run by multiple therapist teams in a variation of natural environments and at different times of the year. This design will ensure our statistical power, in addition to increasing the ecological validity and generalizability of our findings.

Initially, we conducted a basic RCT with the following two groups: (a) clients receiving WT and (b) clients receiving treatment as usual (TAU). The adolescents in the control group were offered WT at a later point if they required additional health care services. This resulted in the following procedure to recruit participants into the study:

1. Patients believed to be suitable for participation were briefly informed about the ongoing WT study by clinical staff.
2. If the adolescent's curiosity was triggered, he or she was invited to an initial meeting with an experienced clinician who represented the WT project team. Detailed information was provided about the WT intervention and treatment goals. The program was introduced as a clinical research project for which the

participants would be randomized to either WT or TAU 1 to 3 weeks later. This procedure ensured that the client had a 50% chance of joining or not joining a WT program.

3. During this same appointment, preliminary clinical assessments were carried out and the participants had the opportunity to ask questions. At this point, it was important that we reached a mutual agreement about his or her eligibility and motivation for attending the WT program.
4. If the patient desired to participate in the research project, the pretests were administered at the initial meeting.
5. Matched randomization took place as soon as possible after the pretests.

Lessons Learned

Ethical

All the adolescents who agreed to participate in the project expressed a desire to take part in the new WT treatment. This was the main prerequisite behind establishing "identical" experiment and control groups.[1] However, the randomized design allowed only half of them to receive their treatment of choice. All the potential participants were by definition vulnerable teenagers, the majority suffering from low self-esteem, insecurity, and/or disadvantageous attribution styles. Many of the control group adolescents experienced their assignment to the control group as yet another defeat along the already bumpy road of life. A colleague explained the randomization process as follows: "It felt like I first handed out a Christmas gift, then retrieved it to do a draw with the neighbor kid to decide who was actually to get it." In general, recruiting patients to research projects is often quite challenging, and teens who are referred to hospital specialized mental health treatment are often in a distressed state of mind, fatigued, defensive, or confused due to the nature of their problems. Hence, for this group of youngsters, randomization into groups was not a straightforward matter.

Even though considerable focus was awarded to simplicity and clarity in the initial conversation with the client, much information was provided. The topics and procedures that were covered included the following: (a) establishing an alliance and preliminary clinical assessments; (b) a review of the WT program; (c) an outline of the research project, including the forthcoming randomization; (d) an account of the alternative high-quality psychotherapy (TAU); (e) a summary of legal rights and ethical aspects in joining the research project; (f) a question and answer session; (g) a completion of the first part of a pretest questionnaire; and (h) a discussion of further appointments for physiological and psychological pretests.

Quite often, some of the provided information, even when acknowledged by the client at the time, appeared to be forgotten later. Furthermore, the participants' feedback demonstrated that the wait for the randomization results added insecurity and unpredictability to their already difficult circumstances. Finally, the administration of all the pretests in conjunction with the initial conversation might have caused further fatigue and a subsequent decrease in data validity. In addition, because we included

physiological heart rate variability data in our research, we logistically had to group participants to dates when experts in this field were available. If the participants were aware of the randomization outcome at this point, this grouping would, in principle, affect this and all other pretest data.

One could argue that a solution to the above problem would be to first conduct several frequent smaller meetings rather than one large meeting and then randomize immediately. We assume this would be a reasonable approach with inpatients or adolescents who are in continuous care or custody. However, in our case, all clients have voluntary outpatient status and thus lead their lives elsewhere. Our ideological and professional view is that it is important for the youth to maintain as many of his or her everyday activities (e.g., school, friends, family, hobbies) as possible while in treatment. We also know that frequent health care appointments pose practical difficulties for many individuals, often resulting in subsequent no-shows.

All research and clinical practice we do is governed by the ethical guidelines for the South-Eastern Norway Regional Health Authority. Salient and re-occurring in these guidelines is that the treatment offered must provide the patient with a feeling of predictability, involvement, and dignity. Although certainly not intended from our side, we acknowledge that the outcome of the randomization process itself can compromise these ethical standards.

Health Outcome

If we are able to identify and extract the "ideal" (i.e., those who will benefit the most) WT participants from a total clinical population, the mean health outcome will increase in both the remaining group and in the WT group. The rationale behind this statement is that WT participants might not benefit from conventional psychotherapy in the manner that other adolescents benefit, thus lowering the average outcome scores if they remain part of this group. A WT intervention, however, might meet their needs, thus increasing the expected results. The goal is for the clients to receive the most appropriate help, and evidence-informed interventions tell us more about group averages than individual benefits. A finding that many in a population respond to traditional psychotherapy still provides little information about those who do not.

If the above assumption is correct, a control group that consists of clients who would prefer WT will predictably report lower health advancements than the experiment group. This consequence has both ethical and methodological implications. First, these young people battle health problems that greatly reduce their overall quality of life. Willingly and knowingly offering less than what is believed to be the best possible treatment might not only be unethical but may also constitute some degree of malpractice. Second, this issue raises methodological concerns. For example, if we assume that the members of a control group are pre-conditioned to respond poorer to the health care that they are offered, is it right to have them in a control group in the first place?

To provide the best health care possible, we help young clients choose among a portfolio of various approaches to therapy. Each of these treatment modalities may be ideal for a specific person due to his or her challenges and counterproductive for

others. To provide the most beneficial health care for each individual client, we believe that a greater diversity of treatment approaches can potentially increase the overall health outcomes of a particular mental health service. Comparing health outcomes between these interventions, therefore, appears to be conceptually problematic.

Practical

We are aware that most WT programs are not hospital based. However, it is likely that nature-based interventions will increasingly be incorporated into government-run health care systems elsewhere as knowledge of the outcomes mounts (Berman & Davis-Berman, 2013). The benefits of working from within a hospital setting are numerous. The most significant advantages are arguably the professional recognition this entails and the availability of WT to teens regardless of their socioeconomic standings (or the standing of their guardians). The challenges, however, are primarily related to navigating health regulations that are not necessarily adapted to this line of work. The details of the current study are nation specific and further generalizations will not be discussed.

Empirical

Many of the above arguments are relevant in a pure empirical sense as well. It is difficult to see how experiment and control groups can be made identical when the participants in one group have their preferences met and the participants in the other group do not. We experienced the adolescents' disappointment when they were informed that they would not receive WT, but TAU. This assignment, at least for some time, decreased their motivation for genuinely engaging in the alternate treatment. A case can be made that this is a methodological error source that biases outcome differences in favor of WT.

If the control group is disadvantaged from the onset of treatment (i.e., not getting their primary choice of health care), an additional dilemma appears. TAU varies across practitioners and treatment approaches within the same clinical context. In other word, the control group interventions are heterogeneous. The main reason for this is that the control group is subjected to individual therapy, and invariably therapists are diverse in terms of both personality and choice of therapeutic approach. Spirito, Stanton, Donaldson, and Boergers (2002) elaborate when they conclude that "this variability makes interpretation of treatment results in clinical trials with treatment as usual comparison groups tenuous" (p. 41).

Finally, there is the question of validity. It is generally acknowledged that a strict study control in psychological research, such as an RCT, reduces studies' ecological validity. In other words, the more control the researcher has on perceivable factors that may "pollute" his research, the fewer generalizations can be made based on his findings. In the words of Waddell and Godderis (2005), "What works in research settings may not be the same as what works in practice" (p. 60). Ultimately, these are discussions of study efficacy and effectiveness as well as internal and external validity,

which are subjects of infinite complexity and on which many books and papers have been written.

Discussion and Some Possible Solutions

We attempted to conduct an RCT because we believe this approach has the prerequisite to provide some knowledge that other study designs cannot. In addition, evidence-based practice opens the door for funding and grants (Harper, 2010). However, as we are even more aware of now, there is a major practical difference between the independent variable of, for instance, a specific drug and a variable that is as complex as a voluminous, multi-faceted WT. In the latter, there is a range of variables at play, even before the onset of the intervention. Anticipation, visualization, motivation, and preparation are all mental processes that arguably take place within a client prior to a scheduled WT. In the control group, however, one can expect these processes to be an initial disappointment and then hope for the client's reorientation and motivation.[2]

If we were dealing with an independent variable that could remain undetected by both the test person and the administrator of this variable, a double-blind approach would be a good option (e.g., pharmaceutical trials). When neither the test person nor the tester is aware of who is subjected to the independent variable, most of the previously mentioned biases are canceled out. Many double-blind studies rely on the use of placebos, effectively excluding studies in which the human resource *is* the independent variable. An additional approach in medical trials is allowing the participants to act as their own controls. In this case, all participants are offered the variable of study but at different times along a longitudinal health intervention. The collection of continuous data will then reveal the hypothesized fluxes in outcome as the variable is introduced (or removed). This approach partially meets the ethical concern that not only those who are randomly assigned to the experiment group but all patients who participate in a possible health-promoting intervention should be allowed to benefit from the intervention. Within a WT-based approach, we encounter considerable practical difficulties in regard to adolescents with mental health challenges. This is especially the case if we recruit patients directly from admission to the health care facility. Starting an individual on TAU and then moving him or her to WT pose obvious problems. However, if a TAU is not working effectively, WT might then be a professionally and empirically better choice. In our study, the patients in the control group knew they could enter a WT program half a year later if they wished to do so and still were in need of treatment. For a troubled youngster with a reduced quality of life, "later" is similar to "never," and our promise of WT down the road did not seem to reconcile their disappointment.

Finally, the regression-discontinuity design (RDD; Thistlethwaite & Campbell, 1960) may be an alternative where randomization is not possible or the preferred option. This methodology elicits the causal effects of interventions by assigning a cutoff or threshold above or below which an intervention is assigned. As this approach has lower statistical power than the RCT, sample sizes should increase. To be able to perform the RDD, we would need access to one or more outcome measures that were administered pre–post throughout the clinic to all clients. These measures would have

to tap constructs that were generally relevant for all regardless of diagnosis, such as quality of life, sense of coherence, self-efficacy, and so on. At this point, we have been unable to collect such data, but aim to do so in a not-too-distant future.

In conclusion, with the approval from the Norwegian ethical committee for medical studies, we abandoned the use of control and comparison groups in this research project. We instead introduced a prior-to-program data collection in addition to the existing and identical pre-, post-, and 12-month follow-up data collections. Bowen and Neill (2013) refer to the period between entry into the study and the first day of the WT intervention as the lead-in period. In their seminal meta-analysis, the analysis of changes or lack of changes in data during this wait-list period is used to validate gains that occur during the WT. In addition, we added a daily and clinically relevant self-report measure, the state section of the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). This measure is collected 17 times for each participant and will be subjected to multi-level modeling, which according to Gelman (2006) can be helpful for causal inference.

To obtain comparative data with other studies, we translated and validated the Life Effectiveness Questionnaire (LEQ-H; Neill, 2008) and the Youth Outcome Questionnaire (Y-OQ-SR 2.0; Ridge, Warren, Burlingame, Wells, & Tumblin, 2009) into Norwegian. These scales are frequently used in Australian and U.S. research, respectively. Furthermore, our test battery comprises scales that are have recently been used in studies on 16- to 18-year old adolescents locally in Norway, thus providing access to more ecological valid data comparisons. Finally, we add a comprehensive volume of qualitative data, establishing a mixed-methods design. This allows for what Palinkas (2014) describes as a "thick description or depth of understanding to complement breadth of understanding afforded by quantitative methods" (p. 851).

At the time of writing, a multi-site study is being prepared. We also acknowledge that showing "economic" effectiveness arguably may be equally beneficial to the adventure therapy field compared with the more common focus on causality. Thus, we aim to include substantial and valid cost-benefit analysis in this work.

In principle, we support the call for more RCTs, but we also acknowledge why they are scarce. Ethical, practical, and empirical issues are major obstacles that must be navigated carefully for a controlled trial to be both an efficacious and a decent study approach. It is prudent to remind ourselves that high-quality evidence-based practice research within the broader nature-health paradigm may come in many shapes and forms. Regardless of whether an RCT is conducted, transparency and sober inferences should be the true hallmark of good science within adventure and WT research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The clinical Friluftsterapi™ research project that is referred to

in the text is financially supported by the Competence Development Fund of Southern Norway (CDFSN) and the Department of Child and Adolescent Mental Health (ABUP) at Sorlandet Hospital HE.

Notes

1. The assumption that randomization of small samples (which is the norm in most adventure and wilderness therapy studies) will lead to identical and balanced groups is, according to Grossman and Mackenzie (2005), irrational. They argue that the groups would likely be more balanced if they were selected by hand.
2. For a further discussion on using randomized controlled trials in complex interventions, see Wolff (2000).

References

Becker, S. P. (2010). Wilderness therapy: Ethical considerations for mental health professionals. *Child & Youth Care Forum*, *39*, 47-61.

Berman, D., & Davis-Berman, J. (2013). The role of therapeutic adventure in meeting the mental health needs of children and adolescents: Finding a niche in the health care systems of the United States and the United Kingdom. *Journal of Experiential Education*, *36*, 51-64.

Bowen, D. J., & Neill, J. T. (2013). A meta-analysis of adventure therapy outcomes and moderators. *The Open Psychology Journal*, *6*, 28-53.

Clark, J. P., Marmol, L. M., Cooley, R., & Gathercoal, K. (2004). The effects of wilderness therapy on the clinical concerns (on Axes I, II, and IV) of troubled adolescents. *Journal of Experiential Education*, *27*, 213-232.

Fernee, C. R., Gabrielsen, L. E., Andersen, A. J. W., & Mesel, T. (2015). *Therapy in the open air. Introducing wilderness therapy to the adolescent mental health services in Scandinavia*, *2*, e14.

Gass, M. A., Gillis, H. L., & Russell, K. C. (2012). *Adventure therapy: Theory, research, and practice*. New York, NY: Routledge.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, *48*, 432-435.

Grossman, J., & Mackenzie, F. K. (2005). The randomized controlled trial: Gold standard or merely standard? *Perspectives in Biology and Medicine*, *48*, 516-534.

Hacking, I. (2000). *The social construction of what?* Cambridge, MA: Harvard University Press.

Harper, N. J. (2010). Future paradigm or false idol: A cautionary tale of evidence-based practice for adventure education and therapy. *Journal of Experiential Education*, *33*, 38-55.

Henderson, B., & Vikander, N. (2007). *Nature first: Outdoor life the friluftsliv way*. Toronto, Ontario, Canada: Natural Heritage Books.

Neill, J. T. (2008). *Enhancing personal effectiveness: Impacts of outdoor education programs* (Unpublished doctoral dissertation). University of Western Sydney, NSW, Australia. Retrieved from http://wilderdom.com/wiki/Neill_2008_Enhancing_life_effectiveness:_The_impacts_of_outdoor_education_programs

Norton, C. L., Tucker, A., Russell, K. C., Bettmann, J. E., Gass, M. A., Gillis, H. L., & Behrens, E. (2014). Adventure therapy with youth. *Journal of Experiential Education*, *37*, 46-59.

Palinkas, L. A. (2014). Qualitative and mixed methods in mental health services and implementation research. *Journal of Clinical Child & Adolescent Psychology*, *43*, 851-861.

Ridge, N. W., Warren, J. S., Burlingame, G. M., Wells, M. G., & Tumblin, K. M. (2009). Reliability and validity of the Youth Outcome Questionnaire Self-Report. *Journal of Clinical Psychology*, *65*, 1115-1126.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Spirito, A., Stanton, C., Donaldson, D., & Boergers, J. (2002). Treatment-as-usual for adolescent suicide attempters: Implications for the choice of comparison groups in psychotherapy research. *Journal of Clinical Child & Adolescent Psychology*, *31*, 41-47.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *The Journal of Educational Psychology*, *51*, 309-317.

Waddell, C., & Godderis, R. (2005). Rethinking evidence-based practice for children's mental health. *Evidence-Based Mental Health*, *8*, 60-62.

Wilson, S. J., & Lipsey, M. W. (2000). Wilderness challenge programs for delinquent youth: A meta-analysis of outcome evaluations. *Evaluation and Program Planning*, *23*, 1-12.

Wolff, N. (2000). Using randomized controlled trials to evaluate socially complex services: Problems, challenges and recommendations. *The Journal of Mental Health Policy and Economics*, *3*, 97-109.

Author Biographies

Leiv Einar Gabrielsen, PhD in psychology. Has together with colleagues developed a Norwegian hospital based wilderness therapy intervention program offered to at-risk adolescents. Gabrielsen is project manager for the clinical research that accompanies this new intervention, as well as a clinician in the wilderness therapy groups.

Carina Ribe Fernee, MSc in Culture and health/Doctoral candidate. Primarily engaged in qualitative wilderness therapy research. Fernee is also a clinician in the wilderness therapy groups.

Gunnar Oland Aasen, Clinical child psychologist. Particularly involved in the development and clinical implementation of the Norwegian wilderness therapy intervention. Aasen is also a clinician in the wilderness therapy groups.

Leif Torvald Eskedal, PhD/MD, Primarily engaged in quantitative wilderness therapy research with a special interest in heart rate variability data. Eskedal is also a clinician in the wilderness therapy groups.