



ALLIANCE FOR THE

# Safe, Therapeutic & Appropriate Use of Residential Treatment

## Treatment Research Lacks Good Science

### *A detailed scientific critique of Behrens study findings*

By Nicole Bush, PhD, Robert Friedman, PhD, Charley Huffine, MD, Barbara Huff, and Phil Elberg, JD  
August, 2011

**We at ASTART have been concerned about the marketing of teen residential programs that highlights the findings from a study by Ellen Behrens and Kristin Satterfield.** Two reports are widely cited in youth residential treatment marketing and promotional materials: *Report of Findings from a Multi-Center Study of Youth Outcomes in Private Residential Treatment* (Aug 2006; available on the web) and *A Multi-Center, Longitudinal Study of Youth Outcomes in Private Residential Treatment Programs* (April 2007; not publicly available, summary of select findings available via marketing materials).

There is a dearth of research on the effectiveness of residential programs, and this study does provide some information for consideration. However, there are striking **conflicts of interest** in the research and **several flaws in the methodology** of the study that make its findings questionable. Further, industry websites make several claims about the findings and their meaning that go far beyond what the data shows, and that our experts believe are **misleading to parents, providers and youth**.

### CONFLICTS OF INTEREST

#### **Study funders have conflict of interest**

*This study was funded by a company that owns and operates for-profit residential programs, which is a conflict of interest.* The company also uses the studies' researcher to personally recruit customers for the programs, which clearly draws her objectivity into question.

When for-profit companies pay for research to confirm the effectiveness of their product or treatment, their goal is generally to find anything that might act as evidence for the effectiveness of their facilities. In such situations, studies are designed and data is very often analyzed and presented in a way that is most beneficial to the company, rather than what is most scientific or accurate, and it is quite difficult for the consumer to detect and understand this bias.

### **Researcher not “independent,” and Aspen not just a “participant”**

The Aspen Education Group (“Aspen”) website is misleading when it begins to describe the research by stating “Aspen Education Group *participated* in the nation’s first large-scale study of its kind” — Aspen funded and guided the research on their own programs, making them far more than just “participants.” Aspen’s website also claims that the study was conducted by an *independent* research company, yet Behrens’s company, Canyon Consulting, was hired and paid by Aspen to perform this research, which can influence objectivity.

Importantly, the 2006 paper does not indicate that *all* 9 participating programs were owned by Aspen, and any conflict of interest disclosures made are relatively hidden from the consumer (e.g. the study authors briefly list Aspen as a funder only at the very end of their manuscript reference list, embedded in a paragraph with unrelated content, and Aspen discloses that they funded the research subtly at the end of the webpage after presenting their eye-catching, selective interpretation of the findings).

### **Financial motives and incentives call research into question**

The findings presented in the reports go beyond the actual data and suggest the programs studied (or similar “struggling teen residential treatment programs”) are effective—which is likely to influence their earnings potential, and thus, the financial interests of the company should be strongly considered when viewing these results.

The programs involved in the research were Academy at Swift River, Aspen Ranch, Copper Canyon Academy, Mount Bachelor Academy, Stone Mountain School, Pine Ridge Academy, SunHawk Academy, Turnabout Ranch, and Youth Care, Inc. The widely publicized 2006 report does not acknowledge that the nine widely varying

programs have the same parent company (Aspen Education Group). The 2007 report does state this, but this report is not publically available and this information is not included in marketing materials that cite this research.

## **METHODOLOGY: INSUFFICIENT CONTROLS AND MEASURES**

### **Results as presented cannot be attributed to the “treatment”**

The study uses a pre-post design with no comparison group, so results cannot be attributed to the “treatment.” This study and the reports associated with it do not have control groups (a group of similar youth who needed treatment but did not receive it). As teens mature, many symptoms decrease naturally without any treatment.

Without any comparison group, it is not possible to determine what would have happened without any treatment. Many youth would decrease in symptoms many months to many years later as they mature or as depression remits naturally, regardless of treatment, and so declines in symptoms cannot be attributed to the Aspen programs given the study design (although Aspen routinely makes this claim when marketing their programs).

It is typical with study designs that use scores on behavior problem checklists (such as the Achenbach CBCL that they used to measure problems) will be higher at admission than at any other time—this is in fact the time of greatest crisis. So showing a decline in symptoms over time could really be unrelated to attendance in such programs.

### **The timing and quality of the outcomes examined make the findings questionable**

The 2006 findings are based on reports from the children and parents at two critical times--first when the child enters the program which is a time when the parents and child see things at their worst, and second when the child is being discharged, which may be a time of optimism. Parents were asked to report on their child’s status when they had not lived around them for months or years.

The 2006 report fails to describe the timing of the "discharge" assessment and Aspen program’s criteria for program completion. The problem with this is that *youth in these programs are NOT generally discharged unless they report decreases in depression and anger and show improved communication with parents*, thus parents are usually quite pleased

with the adolescent's behavior upon discharge (the authors do acknowledge this possibility on p. 13) and youth often report that they are not experiencing problems in order to leave the program. Also, parents have not been living with or regularly interacting with their child while he/she was away, so their perceptions of their child's improvement are likely not reflective of true functioning.

It is more optimal to report the assessment of youth and family functioning several months after discharge, as those seem to be times when (after a brief "honeymoon") many youth return to drug use, acting out behavior, and depression, etc. and significant increases in family discord occur. The 2007 longitudinal report (which is not publicly available) does report those findings, but as we note below, only in a biased sample, and the follow-up findings are not nearly as positive as the discharge findings presented—even with that biased sample.

### **No valid, independent measures of improvements in functioning**

Also, the authors describe effects as "change in functioning" which is misleading. Measures are only for changes in *perceptions* of functioning. There are no valid, independent measures of actual functioning. There is also no discussion of how the parents came to have perceptions of their children's functioning. The vast majority of these youth lived apart from their parents the entire treatment period and often had few if any home visits in between. How much time did parents actually spend with them during treatment or after discharge? Were parent reports based on what staff told them? Was this consistent across all programs? There may be other interpretations of these perceptions, other than that the child has made progress. The reports from the child and family may be influenced by the context and may have little to do with actual changes in the child/family.

### **METHODOLOGY: SAMPLE SELECTION AND ANALYSIS HIGHLY BIASED**

#### **The sample used in analyses is quite biased—making the findings biased**

Children who do poorly in Aspen programs are dropped from the analyses in both reports. Throwing out subjects from your analyses because your treatment did not work with them significantly biases your findings to be positive, and is a questionable practice.

It is noteworthy that in the 2006 the clinical teams at these nine programs classified 50 of the 551 youngsters who were in the discharged sample as "treatment beyond scope." This refers to a group of youngsters for whom the program was not a suitable match, and who were transferred to "a more appropriate setting." The authors report that this group did less well than the others, but their data were excluded from the analyses because "it was deemed that a program making an early referral for students who required alternative clinical care would constitute appropriate, ethical care rather than a 'failure' on the part of the program."

### **Youth who left the program are not included in the analysis**

Aspen programs enrolled those youth and "treated them" and it didn't work so they were sent elsewhere, thus they should have been included in the analyses. "Intention to treat analysis" (in which you count dropouts and non-responders to surveys as "failures of treatment" or at least as "more likely to be failures") is the standard for treatment studies. For example, if you have a drug that produces 100% success at lifting depression in the 13 people who didn't stop taking it due to side effects when 87% dropped out of the study, it's not exactly going to be approved by the FDA or become a widely-prescribed drug.

### **Missing data creates a significant bias in the data**

The authors misrepresent the sample size throughout the papers and they do not handle analysis of their missing data in an appropriate scientific manner. The results for the majority of youth—six out of ten—are dropped from the reporting.

In both the 2006 and 2007 reports, there is tremendous inconsistency in the sample used in analyses. In fact, most families did not complete the majority of assessments during the study, so their data is missing from outcome analyses. In fact, 60% of parents and 37% of youth did not complete discharge assessments, and, on average, 81% of the youth and 73% of parents assessed at intake didn't participate in the follow-up study assessments!

This "missing data" described above significantly biases the findings—generally, families most satisfied with treatment are the ones that complete all forms in a timely fashion and those who are dissatisfied with the services or continue to be in crisis do

not fill out questionnaires. Subsequently, the write ups of the findings are, at times, misleading in that they compare findings from the admission sample to those in the pre-post-test sample, suggesting they are the same group of youth across findings, which is not true because of the huge drop in sample size.

In the abstract for the 2006 paper, the authors talk about a "sample of nearly 1000 adolescents, from nine private residential programs." Later, their method section acknowledges that, for their analyses of changes from admission to discharge, their sample was actually only 403 adolescents and 211 parents, but they do not conduct analyses to discern whether it was the "best functioning youth" to begin with who completed both admission and discharge data--although that is easy to test and should be reported. This report does not address this when interpreting the findings or discuss it as a limitation of the study.

Similarly, the 2007 report starts off talking about a study of 1027 kids, but at the end there are reports from 138 kids and 250 parents (response rates at 12 months post-discharge of 13.5% for young people and 24.5% for parents). In Table 2, the authors report percent of surveys returned for each time period, and they later describe analyses to assess for response bias. The 2007 report does acknowledge that for those who actually completed discharge surveys (already a biased sample from those who were assessed at intake), those who didn't return follow-up status surveys reported less treatment satisfaction, less change in their children's problems from treatment, and higher problems at discharge (in fact, nearly 50% higher total problems for those youth)—which means that the results reported for the longitudinal follow up are based on a sample of youth with the lowest problems at discharge whose parents were most satisfied with the programs.

Importantly, *non-responders were also four times more likely to have pulled their children out of treatment against program advice*, which suggests they did not find the treatment optimal—and their children's outcomes are not included in the analyses.

### **Inadequate statistical methods were used to analyze the data**

A more appropriate model would have been to use a "nested" design for analyses (e.g. multilevel modeling), because the data was drawn from nine discrete sites, and proper

analyses should account for that in the statistical model. Otherwise, the effects of one program can drive effects for the entire analyses and lead to biased results, or some programs with clear harm can be "washed out" by programs that are helpful.

Indeed, in the 2007 report, the authors state that "curriculum and programming across sites was "very diverse" (p.3), which implies that effects across sites may be very different and should be considered. So nested analyses should be conducted to assess for that before you make statements about the entire sample and "all Aspen" or "private residential treatment programs."

## **METHODOLOGY: CLAIMS OF SUCCESS EXAGGERATED**

### **Even within the biased sample, substance use barely decreases**

It is important to note that, even in this biased sample, youth report a substantial increase in alcohol and drug use over the 12 month post-discharge period, and while this increase still doesn't bring them to the level it was reported to be at admission, it is pretty high and close to rates at admission (for alcohol, 3.02 at admission, 1.24 at discharge, and 2.66 one year post-discharge—for drugs it is 3.84 at admission, 1.29 at discharge, and 2.68 one year post-discharge).

**This suggests that youth still had significant substance use problems, at almost their original rate, after their lengthy and expensive treatment.** These results can only be seen by carefully reviewing the tables as the authors do not write about these results or discuss them in their paper, instead focusing their comments on outcomes with better change results.

### **There is precious little discussion of the treatment that is purported to be effective**

Reading the reports, one can find no discussion of what "treatment" took place. What were the treatment modalities used? "Residential" is a *place*, not a "treatment modality." There is no description making it clear what was done to lead to the gains they claim. And it is doubtful that all nine programs did the same thing, equally well or equally well for all youth, especially given the acknowledgment that the nine programs were highly varied in enrollees and treatment modality. If the gains were real, maybe a vacation cruise without drugs or a study abroad program would do just as well.



**Additionally, one of the nine programs evaluated, Mount Bachelor Academy, was recently shut down by the authorities for documented abuse of youth, citing, in particular, that the “treatment modality” itself was found to be psychologically damaging to the participants and conducted by unqualified staff who lacked mental health training.**

### **“Clinical team” and their credentials and methods not identified**

Finally, no mention is made in the report of who made up the "clinical team," or how they were trained to discern "discharge status"—typically there would be a report of how consistency across raters was established (or “reliability across ratings”). It is unusual for quality research to not describe this central measure of their study.

## **MISLEADING USE OF THE UNPUBLISHED FINDINGS IN MARKETING**

### **Weaknesses and limitations of the study are not explained**

The study authors do not adequately acknowledge the study weaknesses or alternative explanations for results. To their credit, Behrens and Satterfield acknowledge a few of their study’s limitations in their reports. They note the need for a control group and the need for further research to determine the merit of these findings, especially in light of the many surprising findings. They also acknowledge that parents may “underreport” their child’s symptoms at discharge if they are motivated to release their child from treatment prior to the time advised by program staff, which may bias outcome data in a positive fashion, misrepresenting the efficacy of the treatment. However, as described above, they often misrepresent their study sample size and largely fail to acknowledge the many methodological flaws in their study and alternative interpretations of surprising findings that might reflect weaknesses on the part of their funders’ programs.

### **The study authors and study funders overstate the findings**

There is insufficient caution about the findings in both the 2006 and 2007 reports by Behrens and Satterfield, and certainly by the for-profit industry websites. Despite the many concerns and flaws outlined above, the 2007 report states, “Clearly, the present study provides evidence of lasting benefit for youth in private residential treatment.”(p



16), and the Aspen website states “Aspen Education Group’s Residential Therapeutic Schools and Programs: Proven Effective.” These studies simply do not support these claims.

Scientific method and research standards articulate that one study, especially a study with poor methodology and biased analyses, cannot “provide evidence of lasting benefit” or “prove” treatments effective. The Aspen website also states that “Aspen’s programs helped teens to develop stronger emotional well-being” and that “teens behave better as a result of Aspen’s programs,” attributing any improvements (real or imagined) in youth health to Aspen programming, but as described above, research that lacks experimental design (such as having a control group) cannot determine the cause of changes in outcomes.

### **There are concerning findings that Aspen does not highlight on its website**

The omissions suggest that there is a one-size-fits-all treatment, and that outside factors—such as age, parental income or use of medication—have *no influence* on a child’s functioning during treatment.

First, it is striking that the only variables in analyses that predicted improvement over time in the regression analyses were things such as youth having “no mood disorder” and “low level of problems at entry” etc. This quote on their website *should raise intense red flags*: “In other words, *change in functioning during treatment does not depend on age, gender, ethnicity, parental income, number and type of problems, presence/absence of psychiatric medication, prior treatment, length of stay, or discharge status*” (2006, p. 12). **It is quite unusual for all of those factors to NOT relate to treatment effectiveness—making it likely that this study has invalid data.** The authors comment on this being a surprising finding, and in an Aspen website video about the 2007 report Behrens describes the findings as “remarkable,” but **it is more than surprising—it is alarming.**

It is generally accepted in the field of psychological research that there are not treatments that have universally positive effects for such a range of complex youth problems. For example even highly-focused interventions by the nation’s leaders in ADHD research with large-scale, multisite, expert-run interventions struggle to demonstrate sizable positive effects of treatment for ADHD. Moreover, the world’s

leading researchers on Depression generally only find successful remission of symptoms in one-third to one-half of their subjects, and not a 100 percent decline in symptoms.

Further, at admission, parents rated their youngsters as having more severe problems than the youngsters rated themselves as having, but this was reversed at discharge—then the youth rated themselves as having more serious problems than did their parents. The finding that parents rated their teens as having more serious problems at admission than the teens themselves did is very typical, but the finding at discharge that the teens rated themselves as having more serious problems is unusual. One interpretation is that the parents were clearly more satisfied consumers of service than were the adolescents themselves. Modern standards of practice articulate the importance of meeting the rights, needs, and perspectives of the youth undergoing them, so this youth perspective is important.

## **OTHER CONCERNS ABOUT SCIENTIFIC MERIT**

**This research has not been peer-reviewed or published in scientific journals, which are clear standards for evidence in psychological science.**

The Behrens and Satterfield reports have not been confirmed by any outside scientists or refereed competitive science publications as scientific evidence. “Peer-review” means experts on the topics that are investigated in the study evaluate the research and critique the findings to determine whether it is of high enough quality to be published. Research findings that have not been published in peer-reviewed journals are of questionable merit.

Although one report was apparently presented as a “poster” at the American Psychological Association (APA) meeting in 2006, conferences are only a forum for sharing findings, and presentation at one does not suggest that the APA approved of the study or that any objective scientists reviewed the study to discern whether the study was conducted properly and interpreted with appropriate caution.

The 2006 report also is published in an outdoor behavioral health trade journal, which appears to provide some enhanced credibility, but it must be noted that the criteria and motives for publication in trade journals can be very different than that of competitive

science journals. The second longitudinal study is presented in a video presentation or slide presentation on the Aspen website, and the accompanying paper can only be obtained by request to Ellen Behrens as it is not publically available. Given this, the findings from both reports should be interpreted with great caution.

### **These findings have not been replicated**

A major principal of scientific method is that research findings must be replicated by another independent researcher in order to consider study findings scientifically valid—otherwise one cannot determine whether a particular finding was due to chance (a fluke) or whether the first researcher’s bias was influencing results. Particularly in treatment research, it is very important to replicate findings, as the effects of therapies are quite complex and varied, and multiple studies are needed to understand when and how they work, and for whom. Without replication, the Behrens and Satterfield findings are questionable.

### **SUMMARY**

**In summary, this study would not be published in a reputable journal as it stands. There is excessive bias in the methodology and analyses that favor positive treatment outcomes, which is particularly concerning given that Aspen paid for the research. Further, Aspen appears to “cherry pick” the results that support their industry and programs and makes claims about the causes of change in children’s health that are not justified by this data. As it stands, this research, as is currently presented to the public, appears to be more marketing and promotion than scientific research on treatment efficacy, so it should be viewed with great caution.**

Although many studies require replication or have methodological flaws, this particular body of research has considerable weaknesses and obvious biases. Moreover, research published in rigorous, well-respected scientific journals suggests that the techniques used by some private residential wilderness programs (including techniques used by the programs evaluated in this research) use approaches that can be harmful to youth. Specifically, research has shown that approaches such as "scared straight," expressive-experiential psychotherapies (e.g., releasing anger or fear by yelling, criticism, etc), boot camp interventions, and peer-group interventions for conduct disorder are likely to harm some youth (Lilienfeld, 2007).

We provide our perspectives on this study and the use of its findings to offer alternative views on the data collected and presented. We are not and have not been privy to the research design process or writing of the results but base our comments on the information presented by the researchers and the funders in their written documents and web-based media presentations. We strongly believe that there is great need for research assessing the effectiveness of residential treatment programs for “struggling teens” and their families, and we would like to see such research funded and published. However, this need must be met by scientifically rigorous, unbiased research conducted by financially objective parties.