Four Years Along: Emerging Traits of Programs in the NATSAP Practice Research Network (PRN)

Anita R. Tucker, Ph.D, LICSW
University of New Hampshire

Ryan Zelov
University of New Hampshire

Michael Young, Ph.D.
University of New Hampshire

Correspondence should be addressed to Dr. Anita Tucker, Department of Social Work, University of New Hampshire, Durham, NH 03824. E-mail: anita.tucker@unh.edu.

Abstract

In its fourth year of active data collection, the NATSAP Practice Research Network (PRN) is showing promising outcomes for the NATSAP programs actively participating; however the overall implications of this PRN are still emerging. This study specifically looked at the changes reported by youth and their parents from admission to discharge using the Y-OQ scales. Overall, significant mean differences with large effects sizes were found at discharge, with many changes large enough to show significant clinical change according to the Y-OQ benchmarks. Additional analyses suggested that gender and depression were related to rates of change in both residential and outdoor behavioral healthcare programs. Despite these findings and similar to PRNs overall, there are several limitations to these findings including large variations in the data, limited generalizability, attrition and missing data. Only with increased support both on the research and program side can the potential of this PRN be realized.

Keywords: *NATSAP, Practice Research Network, Outdoor Behavioral Healthcare (OBH), residential treatment centers (RTCs), Y-OQ*

Four Years Along: Emerging Traits of Programs in the NATSAP Practice Research Network (PRN)

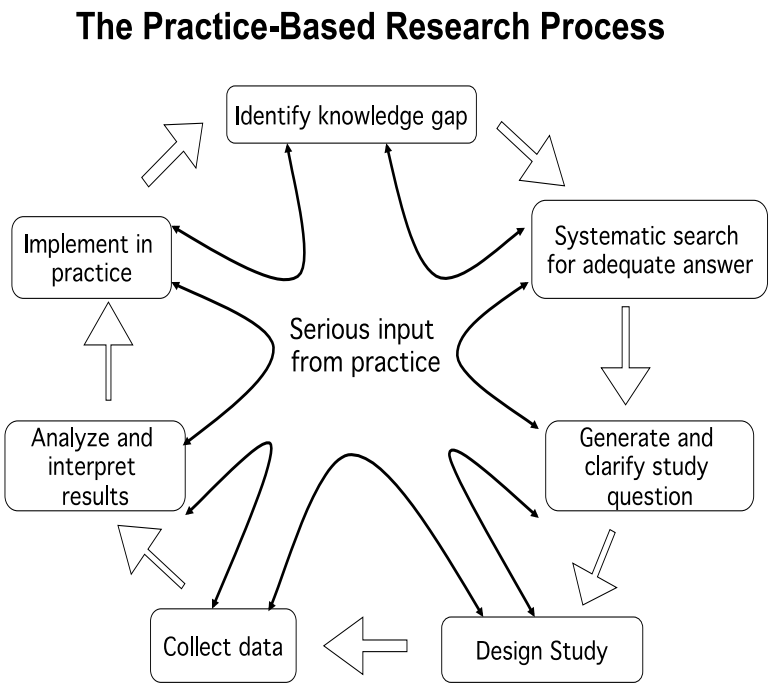
Established in 2007, the NATSAP Practice Research Network (PRN), also known as the NATSAP Research and Evaluation Network, was developed to respond to the call for research demonstrating the program effectiveness of NATSAP programs by clients, allied professional organizations, and federal agencies, such as those sponsoring the *Stop Child Abuse in Residential Programs for Teens Act of 2009* (Gass, 2006; Gass & Young, 2007; H.R. 911, 2009). The NATSAP PRN was seen as a cost effective tool to provide evidence-based outcomes for programs. These outcomes were available as aggregated organizational outcomes indicating what is (and is not) being accomplished by NATSAP programs as an industry group. The outcomes were also accessible by individual programs as credible and confidential feedback on the effectiveness of their particular programs compared to other programs. Four years later, the NATSAP PRN has established emerging support of the effectiveness of NATSAP programs, however many questions still remain about the "true" outcomes achieved by these programs. Despite these limitations, there are a variety of strengths such an approach has over other research efforts.

The use of PRNs (or sometimes called PBRNs for Practice Based Research Networks) to examine medical, clinical, mental health and educational issues is a well-established research methodology (McMillen, Lenze, Hawley, & Osborne, 2009). PRNs have been established in different fields, but especially in areas of health and mental health, in response to a call by consumers, government agencies and insurance companies to increase cost efficiency and the quality of care for clients (Luijsterburg, van den Bogaard, & de Vries Robbé, 2007). Some of the first important examples of PRN research were studies conducted by the RAND organization in the late 1980s. The outcomes of RAND's Medical Outcomes Study (MOS) determined US healthcare policies on the role of financing and reimbursement strategies for public and private care that are still used today (Gilbody, House, & Sheldon, 2002). In fact, "the enduring legacy of the MOS is the fact that patient centered measures of health status developed for the study eventually evolved into the short form 36 (SF 36) - now the most commonly used generic measure of health related quality of life" (Gilbody et al., 2002, p. 1).

While sharing many of the same research and statistical methods, there are several important differences between PRN research and more limited standard experimental and quasi-experimental research design practices. One of the main differences, and the major strength of the PRN approach, is the use of a network of collaborative providers (Gilbody et al., 2002; McMillan et al., 2009). By banding together, these providers create a more compelling research designs by offering a broader understanding

of client treatment than could be typically achieved by analyzing the work of a single organization. Another major difference is when data are interpreted and when these interpretations are used to inform practice. As seen in Figure 1, traditional experimental designs typically wait to implement conclusions drawn from their findings in just the final stages of the research process. PRN research looks to collect data and implement findings throughout all stages of the research process. This has been evident with the NATSAP PRN as several studies have collected and interpreted data early in the "life" of the NATSAP PRN (e.g., Young & Gass, 2008, 2010).

Figure 1: *The Practice Based Research Process*



Gilbody et al. (2002) highlight other critical differences that exist between outcome-based research with PRN databases and other traditional research outcome designs (e.g., quasi-experimental designs). PRN research evaluates interventions that are already in place in mental health care settings; collects data that are already in place, part of the treatment process and easily collected; uses clinical staff to collect data; and often collects data even before a specific research question is known. On the other hand, traditional experimental design research collects data only after the research question is known, implements new procedures

for clients as part of the research process and data are generally collected by specified researchers with their only responsibility being to research client outcomes. PRN research also can collect data in multiple settings of providers comparing different and sometimes competing interventions (rather than denying any intervention for some clients); analyzes existing interventions and normative data or established criteria for comparison; and can utilize research methods that are relatively inexpensive and conducted with real-life clients experiencing real-life issues. In contrast, more traditional experimental research conducts studies in one particular setting with one particular research question, involves a research process that is obtrusive and new to the clients, implements more controlled experimental designs with comparison or control groups who receive no treatment or a placebo, and the cost is relatively expensive (Gilbody et al., 2002). While in mental health settings they are often under used (McMillan et al., 2009), it is clear that PRNs may be very valuable to any research on the impact of mental health practices due to their pragmatic flexibility and their efficient practice informed agenda.

When using a practice research network database, several recommendations suggested by Rosenhack, Fontana and Stolar (1999) have been adapted for use with the NATSAP PRN database. These are to use: (1) large numbers of clients, (2) standardized instruments that are appropriate for the clinical condition being treated, (3) outcome measures that are valued by clients and funding agencies, (4) outcome measurements in multiple relevant domains, (5) extensive data in addition to outcome measures in order to support comparison procedures (e.g., large amounts of demographic data), (6) the collection of data in standardized intervals right after important events such as immediately after discharge, and highly valued standard collection periods (e.g., one year post discharge), and (7) aggressive steps to achieve the highest possible follow-up rates.

The purpose of this study was to present the current status of the outcome data of the NATSAP PRN four years along. This article will present the trends in the data, limitations of the database at this stage of its development, and potential directions for the future.

Methods

Measures

The NATSAP programs participating in this study collected psychosocial client information from multiple sources. The NATSAP PRN currently utilizes the Outcome Questionnaire Family of Instruments (OQ) (Burlingame et al., 2005; OQ Measures, 2011; Wells, Burlingame, & Rose, 2003). The Y-OQ-SR 2.0 and the Y-OQ 30 SR are self-report instruments completed by youth ages 11 to 19. The Y-OQ 2.0 and Y-OQ 30 instruments were also completed by parents and guardians at admission and discharge (Burlingame et al., 2005; Wells et al., 2003). The Y-OQ 2.0 assesses a variety of behavioral and emotional problems and possesses a variety of subscales outlined in Table 1. Unlike the Y-OQ

2.0, the Y-OQ 30 does not have a differentiation of subscales but is a shorter version that provides a global index score of youth's behavioral and emotional distress (Burlingame et al., 2005; Wells et al., 2003). The OQ assessments possess established normative scores with documented validity and reliability (Holloway, 2004; Jones, 2004; Lambert et al., 1996; Mueller, Lambert, & Burlingame, 1998; Wells et al., 2003). Programs participating in the NATSAP PRN had the option to use the Y-OQ 2.0 or the shorter Y-OQ 30 version.

Table 1
Y-OQ 2.0 Measure and Subscales

Youth Outcome Questionnaire	64 items
(1) **Interpersonal Distress:** Assesses change in emotional distress including anxiety, depression, fearfulness, hopelessness, and self harm.	
(2) **Somatic:** Assesses change in somatic distress typical in psychiatric presentation, including headaches, dizziness, stomachaches, nausea, and pain or weakness in joints.	
(3) **Interpersonal Relations:** Assesses change in the child's relationship with parents, other adults, and peers as well as the attitude towards others, interaction with friends, aggressiveness, arguing, and defiance.	
(4) **Critical Items:** Assesses inpatient services where short term stabilization is the primary change sought: changes in paranoia, obsessive-compulsive behavior, hallucinations, delusions, suicide, mania, and eating disorder issues.	
(5) **Social Problems:** Assesses changes in problematic behaviors that are socially related, including truancy, sexual problems, running away from home, destruction of property and substance abuse.	
(6) **Behavioral Dysfunction:** Assesses change in a child's ability to organize tasks, complete assignments, concentrate, handle frustration, including items on inattention, hyperactivity, and impulsivity.	

In addition to the standardized instruments, additional data were collected through customized questionnaires used with program staff (e.g., reasons for referral, referral source, admission date, gender, date of birth., and record of abuse), clients (e.g., attitude toward program and drug/alcohol use), and parent/guardians (e.g., previous treatment history, recent school performance, client drug/alcohol use). Copies of all questionnaires used can be viewed at the NATSAP website (http://natsap.org/research/natsap-research-and-evaluation-network/).

The Sample

Data were collected on 3,041 clients admitted to 23 residential programs between December 2007 and December 2010. All 23 of the programs were predominantly private-pay facilities and were all NATSAP

members. The current study, however only reported on those clients who have left treatment and for whom discharge data were collected via Y-OQ measures. This smaller study sample consists of 983 youth, or 32.3% of the total sample currently in the dataset. The clients in this study came predominantly from OBH programs (89.5%) and the remainder from RTCs (10.5%)[1] . A complete breakdown of the number of clients represented by the 11 programs from which the data were collected can be seen in Table 2. This table clearly shows how varying the degree of participation was for the different programs in the NATSAP PRN. In the RTC group, some programs only had one participant, and in the OBH group, one program provided 640 out of the total sample of 879 OBH clients.

Table 2
Data Collection from Participating Programs (N = 983)

	Program #	*n*	%
Residential Treatment Centers	Program 2	32	31.2
	Program 7	16	15.5
	Program 8	7	6.8
	Program 10	20	19.3
	Program 12	28	27.2
	Program 23	1	1.0
	Total	104	100.0
Outdoor Behavioral Health	Program 3	107	12.2
	Program 17	19	2.2
	Program 19	52	5.9
	Program 22	640	72.8
	Program 24	61	6.9
	Total	879	100.0

The majority of the study sample was male (67.4%) with 32.6% of the clients being female. The average age of the clients in this study sample was 15.8 years (*SD* = 1.7), with 93.7% of the clients between the ages of 13 and 18 years of age. For the clients for whom presenting issues were reported (*N* = 312), the most common presenting issues were alcohol and substance issues (57.4%), followed by depression (32.7%), Oppositional Defiant Disorder/Conduct Disorder (24.7%), and attention issues including Attention Deficit Hyperactivity Disorder or Attention Deficit Disorder (17.3%) (see Table 3). In most cases (89.1%), participants had two or more presenting issues. To see if there were statistical differences between the sample for which we have discharge data (*N* = 983) and the group for which we only have admission data (*N* = 972), independent

1) There were only data from 25 clients discharged from therapeutic boarding schools. The small sample from therapeutic boarding schools was expected given the lengths of stay are traditionally longer in these settings than the other two placement sites, so this sample was not included in this current study, but will be part of follow up studies once the dataset grows.

samples t-tests were completed comparing the admission means scores between these two groups on the youth self report forms and parent forms. At admission, there were no statistical differences between the means levels of functioning of the two groups on any of these measures. These findings support the notion that these groups were comparable and this study sample was typical of the overall NATSAP population.

Table 3
*Presenting Issues of Residential Participants (N = 312)**

Issue	n	%
Alcohol and Substance Abuse	179	57.4
Depression	102	32.7
Oppositional Defiant Disorder/ Conduct Disorder (ODD/CD)	77	24.7
Attention Issues (ADHD/ADD)	54	17.3
Anxiety	39	12.5
Trauma	31	9.9
Learning Disabilities	21	6.7
Autism	4	1.3

* NOTE: 89.1% of participants had 2 or more presenting issues

Findings

Youth Self Reports - Y-OQ

The only measure used to collect assessment and discharge data from youth in OBH programs was the Y-OQ 30-SR, with 879 youth completing this measure at admission and discharge. Unlike OBH programs, RTCs used the Y-OQ 2.0 SR with 104 youth completing the Y-OQ 2.O SR at discharge. Discharge data were collected from all students at the end of their programs. As noted, the Y-OQ 2.0 was only used at RTCs, hence there was a smaller amount of subscale data from the youth (N = 104) overall. Table 4 provides a complete description of the mean scores at admission and discharge on both Y-OQ measures. Paired samples t-tests were completed as well as effect sizes (d) and their confidence intervals for each analysis. Effect sizes measure the strength of a relationship across groups and are used to make numeric comparisons between different findings and their overall treatment effects. Effects sizes are considered to be small when .20 or less, medium at .50 and large when greater than .80 (Cohen, as cited by Gillis & Speelman, 2008). When looking at youth self report, statistically significant differences as well as large effect sizes were found on all measures (see Table 4). Higher scores correlate with higher levels of dysfunction in the lives of the youth. These findings were consistent with the changes reported for the residential youth who completed the Y-OQ 2.0 and reported statistically significant decreases in total scores (signifying increases in functioning) and all six

subscales from admission to discharge, as well as large effect sizes (see Table 4).

Table 4
Y-OQ YSR Mean Scores at Admission and Discharge

	$M_{Admission}$ (SD)	$M_{Discharge}$ (SD)	*t*	*d*	95% CI (lower – upper)
Y-OQ 2.0 Scores from Youth in RTCs *(N = 104)*					
Total Score	89.38 (34.0)**	**40.00 (37.2)**	10.74*	2.55	-3.87 – 9.08
Critical Items	8.98 (5.5)**	4.2 (3.8)	7.96*	1.99	.73 – 2.98
Behavioral Dysfunction	19.6 (8.2)	10.5 (8.6)	9.63*	1.86	-.64 – 4.06
Social Problems	10.2 (6.7)**	**2.1 (5.2)**	10.06*	1.42	.29 – 2.33
Interpersonal Relations	11.3 (8.3)**	**2.4 (7.9)**	9.74*	2.34	.94 – 3.65
Somatic	7.9 (5.2)**	4.6 (4.4)	6.15*	1.49	.38 – 2.14
Intrapersonal Distress	31.3 (12.9)**	**16.3 (12.5)**	9.27*	2.12	-.19 – 3.80
Y-OQ 30-SR from Youth in OBH *(N = 879)*	41.05 (17.1)**	**22.61 (15.1)**	27.84*	1.33	.20 – 2.33

* $p < .001$
** Scores above the clinical cut-off which reflects dysfunction.
Bold scores represent changes considered to be clinically significant.

To help track client outcomes as well as client progress, clinical cut-off scores were calculated by the instrument developers who compared scores from a normative sample to two clinical samples of inpatient and outpatient populations (Burlingame et al., 2005; Wells et al., 2003). Based on these cut-offs, all of the mean admission scores for the Y-OQ 2.0 SR and Y-OQ 30 SR were within the clinical range of dysfunction for the participants; however, after participating in their residential programs, all of the discharge means were considered to be within the non-clinical range of functioning. In addition to cut-off scores, a reliable change index (RCI) (Jacobsen & Truax, 1991) was derived for all Y-OQ measures to determine if clients had made significant changes in their symptoms, because statistical significance does not always equate with clinical significance. For an individual's total score to be considered clinically significant according to the Y-OQ 2.0 SR the change must be 18 points or greater (with varying levels for the subscales) and 10 points or greater for the Y-OQ 30 SR in additional to post treatment scores falling below the clinical cut-offs (Burlingame et al., 2003; OQ Measures, 2011). As shown in bold on Table 4, both total scores were considered to reflect scores of significant clinical improvement, as well as three subscales for the Y-OQ 2.0 SR (Social Problems, Interpersonal Relations, Intrapersonal Distress).

Parental Reports – Y-OQ

Similar to the youth self report data, OBH programs used the Y-OQ 30 with parents, while RTCs used the Y-OQ 2.0 with its subsequent sub-scales. Overall, admission and discharge data were available from 87 parents of youth in RTCs and 171 parents of youth in OBH programs for a total of 258 parents reporting (representing 26.2 % of the youth from whom there was self-report discharge data as well). Table 5 provides a complete description of the mean scores at admission and discharge on both Y-OQ measures. Paired samples t-tests were completed and statistically significant differences were found on all measures, as well as high effects sizes (see Table 5).

Table 5
Parent Y-OQ Means at Admission and Discharge

	$M_{Admission}$(*sd*)	$M_{Discharge}$(*sd*)	*t*	*d*	*95% CI* (lower – upper)
Y-OQ 2.0 Parent Scores from RTC Youth (N = 87)					
Total Score	98.8 (30.5)**	**30.5 (31.1)**	17.23*	1.48*	-5.06 – 8.63
Critical Items	11.9 (6.0)**	**2.3 (4.7)**	13.31*	1.12	.06 – 1.85
Behavioral Dysfunction	28.0 (11.9)**	**11.4 (10.5)**	12.57*	1.32	-.26 – 2.97
Social Problems	9.0 (5.4)**	**2.6 (4.3)**	9.45*	1.39	.11 – 2.39
Interpersonal Relations	14.0 (6.7)**	**1.5 (6.2)**	15.86*	1.34	-.25 – 2.86
Somatic	8.4 (5.3)**	3.3 (3.1)	9.61*	.83	-.17 – 1.68
Intrapersonal Distress	27.5 (11.0)**	**9.5 (8.0)**	14.12*	1.27	-1.21 – 3.67
Y-OQ 30 Parent Scores for OBH Youth (*N* = 171)	51.76 (19.6)**	**28.12 (17.0)**	13.278	1.81	-1.13 – 4.36

* $p < .001$
** Scores above the clinical cut-off which reflects dysfunction.
Bold scores represent changes considered to be clinically significant.

Based on the clinical cut off scores for the parent forms, at admission parents reported their children to be functioning at a level of clinical concern or deviant from a non-clinical population of peers on all of the measures. After participating in their residential programs, however, all of the discharge means were considered to be within the non-clinical range of functioning. Unlike what youth reported, parents reported not only statistically significant changes, but changes that were large enough to be considered clinically significant according to the measure's reliable change index (RCI) on almost all measures (Wells et al., 2005; OQ Measures). As shown in bold on Table 5, the means of all of the scores except Somatic were considered to reflect areas of functioning as reported by the parent in which the youth had shown clinically significant changes.

Comparisons by Gender, and Presenting Issues

This study was not only interested in the overall impact of the residential programs on youth functioning, it was also interested in exploring which factors may influence a change in functioning in residential clients. In order to do this and include all youth and parent report data in the analyses, a Y-OQ 30 equivalent score was computed for the 104 students who completed the Y-OQ 2.0 SR version and the 87 parents who completed the Y-OQ 2.0. This equivalent score was created by combining those questions from the larger 2.0 version which were similar to the questions on the short form Y-OQ 30. These scores were included with the 879 students and 171 parents who completed the Y-OQ 30 for a combined total score from 983 youth participants and 258 parents at assessment and discharge. Change scores were then computed for each youth and analyses were made to see if gender and presenting issues were related to the level of change seen in participants both from self and parent reports.

Although this research was also interested in differences due to program types, due to the lack of breadth of data coming from a variety of programs it would have been inappropriate to make these comparisons. This is especially true since a large majority of the OBH data came from one OBH program and most of the RTC data came similarly from one program As the data grows from more programs, such comparisons may be important to make in future analyses.

Gender

Before discussing differences by gender and presenting issues, it must be noted that on average according to the Y-OQ equivalent total scores, youth and parents in the study reported clinically significant improvements overall (M_{change}> 10) regardless of gender or presenting issues. In general, the youth in the NATSAP programs made significant gains from admission to discharge. In addition to this, some groups had significantly larger improvements compared to others; but the effect these differences had on rates of change varied as highlighted by varying effect sizes.

According to youth participants, it seemed that on average females (n = 301) improved more (M_{change} = 23.13, sd = 21.8) than the 623 males (M_{change} = 17.36, sd = 19.6) (t = 4.03, df = 92, p < .001), yet this difference was small to medium in terms of effect size (d =.28, CI = .14 - .42). According to parents, however, although females (M_{change} = 30.4, sd = 22.7, n = 116) did better than males (M_{change} = 26.10, sd = 23.2, n = 126), these changes were not statistically significant. One factor that is important to discuss when looking at gender influences is that according to Y-OQ self report admission scores, females on average were more acute ($M_{admission}$ = 45.5, sd = 18.7) than males who reported lower levels of dysfunction ($M_{admission}$ = 40.7, sd = 17.2) at admission. At discharge, however, males reported similar levels of psychological functioning ($M_{discharge}$ = 23.3, sd

= 15.7) as females ($M_{discharge}$ = 22.2, sd = 15.6)(t = .99, df = 928, p = .32). Hense while females had higher levels of improvement, on average males and females were no different in terms of functioning at discharge.

Presenting Issues

A variety of independent samples t-tests were completed with both youth and parent data to see if youth who presented with depression, alcohol/substance abuse, ODD/CD or attention issues had significantly larger changes than those without these issues. Only one significant difference was found. According to youth self reports as shown in Table 6, youth who were referred for depression reported to improve significantly more than youth without depression issues at intake (t = 2.13, df = 310, p = .034), yet the effect of being depressed on change was small (d = .25, CI = .01 - .49)

Table 6

Comparison of Mean Changes in Y-OQ Equivalent Scores by Program Type and Gender.

	Youth Self Report		Parent Report	
Depression and Gender	***M*$_{change}$ *(sd)***	***n***	***M*$_{change}$ *(sd)***	***n***
Depression	**23.86 (21.3)**[a]	102	35.29 (17.7)	45
Males	**20.10 (21.3)**[b]	63	**38.93 (19.5)**[d]	14
Females	**30.11 (20.4)**[b]	38	33.65 (16.8)	31
No Depression	**18.02 (20.5)**[a]	210	30.64 (20.8)	85
Males	**14.57 (19.7)**[c]	136	**24.69 (23.1)**[d,e]	39
Females	**24.74 (20.6)**[c]	70	**36.81 (16.3)**[e]	44

[a-e] $p < .05$

Additional analyses were conducted to see if presenting issues combined with gender were related to change. Gender and depression did appear to have an impact of rates of change, at least according to parent reports. As shown on Table 6, for males who presented with depression, their parents reported significantly higher rates of change than parents of males who did not present with depression (t = 2.05, df = 51, p = .045) with a medium effect size found (d = .64, CI = .02 – 1.26). However, male self reports did not support this finding and no significant differences were found between females with or without depression as reported by youth or parents. To further investigate the role of depression and gender, analyses were conducted to see if there were differences in rates of change as reported by youth and parents between males and females with depression, and males and females without depression. Female youth who presented with depression reported significantly higher levels of change than male youth with depression (t = 2.33, df = 99, p = .02, d = .48, CI = .07 - .89), and female youth without depression also reported significantly higher levels of change than males

without depression (t = 3.459, df = 204, p = .001, d = .51, CI = .22 - .80). Hence according to self report data, gender appeared to have had more of a role on change than depression.

Discussion

Based on these preliminary findings, the programs in the NATSAP PRN appear to be significantly impacting change in their clients, most of whom were adolescents. This was based not only on youth self report, but also on the reports provided by the parents of the youth. Based on the Y-OQ and Y-OQ SR measures, not only did the youth improve significantly from admission to discharge, all but one of their assessment scores were considered above the cut off for clinical dysfunction at assessment, and all below this cut off at discharge. In most instances youth scores also improved enough to be considered clinically significant. For youth self report Y-OQ 2.0, levels of social problems, interpersonal problems and intrapersonal distress all decreased to a level that reflected healthy, non-deviant behavior. Both the Y-OQ 2.0 total scores, as well as Y-OQ 30 SR total score similarly showed improvements considered clinically significant. Youth regardless of setting (RTC or OBH) reported on average to be significantly and clinically improved at discharge. These clinically significant changes as determined according to Y-OQ benchmarks, were also supported statistically by large effect sizes, all but one were above 1.0.

Parent and Youth Differences

Overall, Total Scores for the Y-OQ 2.0 as reported by parents of youth in RTCs and scores for the Y-OQ 30 as reported by parents of youth in OBH programs improved enough to be considered clinically significant similar to youth reports (Jacobson & Truax, 1991). Comparable to the youth subscale scores, the Y-OQ 2.0 parent subscale improvements were strong enough to be considered clinically significant for Social Problems, Interpersonal Relations and Intrapersonal Distress, reflective of the youth reports; however, parents also reported improvements in Critical Items and Behavioral Dysfunction, which were not reflected by youth self reports. Hence, the parents of youth in RTCs reported overall changes in more areas of functioning than their children.

In exploring this difference in perspectives between parents and youth, parents at admission reported higher average levels of dysfunction than their children. For example, the mean Total Scores reflected were 89.38 (Y-OQ 2.0) for RTC youth and 41.05 (Y-OQ 30) for OBH youth while parents' mean total scores at admission were 98.8 (Y-OQ 2.0) for RTC youth and 51.76 (Y-OQ 30) for OBH youth. This pattern was similar for all of the Y-OQ 2.0 subscales as well for RTC youth (see Tables 2 and 3). This variance between parental and youth self report scores was consistent with similar outcomes studies of residential programs (Behrens & Satterfield, 2006) and OBH programs (Russell,

2003; 2005) as well as earlier iterations of this dataset (Young & Gass, 2010). In addition, the admission score variances were similar to patterns that have been observed in other assessment instruments such as the ASEBA (Achenbach & Edelbrock, 1991) and the Social Skills Checklist (Gass, 2005). Based on this, it appears common for parents to see youth as more acute than they view themselves. This was not surprising considering in many instances parents played a key role in youth attending these programs, due to their concern for their child and their behaviors and monetary investments in their child's treatment.

Although admissions score variances between youth and parent reports were similar to previous research, unlike Russell (2003) who observed that parent and self-report scores were similar at discharge, the sample studied in this study showed that for OBH participants, parents reported youth to be functioning worse than youth reported at discharge, as reflected by higher mean Y-OQ 30 discharge scores (28.12) reported by parents than youth (22.61). This pattern, however, was not consistent for RTC participants. Youth in RTCs reported to be functioning worse at discharge than parents of RTC youth report, as reflected by mean Total Y-OQ 2.0 scores and subscales (see Tables 3 & 4). The reasons for this difference were unclear, since there is little information regarding youth and parent report variance in the literature as most studies using the Y-OQ have focused on parent reports and not included self-reports in their analysis (Russell, 2003).

One possible reason for these inconsistencies may be due to differences in the number of males and females in each program and how problems were manifested according to gender. For example, females are more likely to internalize their issues, which are not always visibly noticeably to parents, while males tend to externalize and engage in behaviors that parents can observe (Eschenbeck, Kohlman, & Lohaus, 2007; Maschi, Morgan, Bradley & Hatcher, 2010). Hence because females account for the majority of the findings for RTCs and males for OBH programs, parent perceptions of problems may differ from children between programs based more on the gender of their children than on specific program type. This analysis should be considered tentative until the data on males in RTCs and females in OBH programs grow.

Gender Differences

The results of this study also showed that mean changes reported by all female participants from admission to discharge were significantly larger than those reported by males. These findings are congruent with those found by Russell (2003), where females reported higher levels of improvement than males. It is interesting to note that based on self reports females on average entered the programs with higher levels of dysfunction than males, but were functioning at the same level at discharge. It is unclear why this difference at admission exists, perhaps females have a more realistic sense of their functioning at admission, and males are more likely

to minimize their issues. This perspective is actually supported by the fact that parents of males did not report significantly higher changes than the parents of females and that parents in some cases reported their children to be more acute than their children. Parents may in fact have a more accurate perception of their child's functioning. In addition, small sample sizes for males in RTCs, females in OBH programs, and parent data as well as large variances for all groups may have impacted the power of the study and sensitivity of the analyses to detect change, making the role of gender unclear. Only as the sample sizes grow within the NATSAP PRN will the exact nature of gender and program influences on change become clearer.

Presenting Issues

When the four most common presenting issues were examined (alcohol/substance abuse, attention issues, ODD/CD and depression), youth with attention issues, alcohol/substance abuse and ODD/CD on average significantly and clinically improved in functioning as reported by parents and youth at discharge, but their improvements were not any larger than other youth without these presenting issues. These findings were consistent when factors of program type and gender were controlled for; hence it appears that treatment was equally impactful regardless of these issues of attention, substance abuse or ODD/CD. The data showed that only youth with depression issues at intake improved significantly more than participants who did not have this as a presenting issue according to youth self reports. These findings were consistent with previous research on OBH programs (Russell, 2003), yet inconsistent with previous research that found that the absence of mood disorders was a stronger predictor of positive outcomes for residential youth (Behrens & Satterfield, 2006).

When controlling for gender, the findings were inconsistent between youth and parent reports. According to youth, females with or without depression reported higher levels of change with medium effect sizes than males, but depressed females and males did not report significantly higher levels of change than their non-depressed counterparts (see Table 6). This suggests that gender, not depression was related to mean levels of change at discharge according to youth reports. On the contrary, according to parent reports, only females without depression were shown to improve more than males without depression. In addition, males with depression were found to improve significantly more than males without depression with a medium effect size suggesting that both gender and the presence of depression played a role in overall mean changes in functioning. One of the challenges in understanding these findings on presenting issues was that they were based on data from only a smaller subset of youth from whom there was matched data from admission to discharge [n = 312 (youth); n = 130 (parent)]. Hence due to missing data, it was difficult to say with certainty the role that depression and gender have on outcomes.

Limitations and Future Directions

As has been shown through the findings of this study, it appears that NATSAP programs that have reported data have shown on average consistent and clinical improvements in clients according to both youth and parents. Despite the positive nature of these findings, it is important to note the large variances among these outcomes. Although the mean differences from admission to discharge were consistently large and significant, the standard deviations of these means were also large. For example, youth in RTCs reported on average a mean change of 28.68 with a standard deviation of 26.0. Based on this, approximately 68% of the RTC youth reported changes between 2.68 points and 54.68. Hence, when looked at individually, there were youth that did not have clinically significant improvements.

These variances also impacted effect sizes. Although large effect sizes were reported for pre post changes as shown in Tables 4 and 5, the confidence intervals were also large. Using the previous example of self reports from youth in RTCs, although the effect size comparing admission and discharge means was large at 2.55, the 95% confidence interval ranged from as low as -3.87 to as high as 9.08. Hence, the effect size could easily have been small to inconsequential or much larger than reported. So while the data in this study showed a trend towards improvement, youth experienced a large range of changes and in some instances a lack of improvement and/or worsening of symptoms. In fact, around 34% of the youth reported changes less than the 10 points considered necessary for clinically significant improvements, while the other 66% reported clinically significant positive changes in functioning. Clearly, success was not global for all of the participants and these findings should not be applied universally to all youth in these programs.

These findings were also limited by some of the challenges that many PRNs face including recruitment and generalizability, measurement validity, managing relationships with members and ongoing program support (McMillan et al., 2009). In terms of recruitment and generalizability, it is important to point out that the data included in this study came from only 11 of the 23 programs (47.8%) actively participating in the NATSAP PRN. In fact, 640 of the total sample of 983 were from one single OBH program. Not all programs in NATSAP are fully engaged in the NATSAP PRN and those which are engaged are at different levels of data entry. Hence, these findings should not be considered representative of all NATSAP programs, or even the 23 NATSAP programs that are part of the PRN. In fact, the OBH outcomes may be more due to one or two programs than OBH as a model overall. Though promising, these findings are only a glimpse of the future.

Another limitation of these findings has to do with the validity and reliability of data. While the OQ measures have shown to have consistently strong reliability and validity, a lack of consistent data entry in

terms of demographics and presenting issues at intake by programs, limited the ability to truly understand how these independent variables impacted changes in youth functioning. In addition, attrition at discharge limited the size of the matched data and the confidence in the findings. Since it was unclear why discharge assessments were not completed, it cannot be ruled out that those participants were more acute or did worse than others for whom discharge data were collected.

If the future potential of the NATSAP PRN is to be maximized, these challenges need to be addressed. As with many PRNs, a lack of data entry is more likely due to the demands of the task rather than belief in the importance of the project (McMillan et al., 2009). In order to minimize the time required by busy practitioners, McMillan et al. (2009) stress the importance of managing relationships with PRN members and the need for ongoing support on both the research and program sides. First of all regular communication and exchanges between members are key. This can require:

> *"a well designed and implemented infrastructure. Needed resources might include ample budgeted time from a project manager, website with interactive features, automated email notification systems, annual open meetings and newsletters detailing findings from previous PBRN studies and describing upcoming and underway studies" (McMillan et al., 2009, p. 313).*

While many of these key features already exist as part of the NATSAP PRN system, like conference calls and progress reports, improvements are underway. In order to deal with past challenges of the complicated nature of data input noted by member programs, a new system of data entry was implemented and put into effect in June 2011, allowing for a more streamlined data entry system with less redundancy. This new system in many cases will provide identical record keeping systems from which to draw demographic and client history data more easily, which will augment the strength of the fact that programs in the NATSAP PRN already use similar standardized outcome measures.

In addition to support from the research side, support from the program side is equally important. If the NATSAP PRN is to endure over time, as is the aim of the database, minimal institutional commitment is needed (Clotier, 2005). Ongoing program and database management are crucial, which could include practitioner incentives, the use of research assistants and possibly reimbursing clinicians for lost time, or budgeting in time to complete assessments (McMillan et al., 2009).

Clearly the NATSAP PRN has shown the potential to produce significant network-wide program outcomes. While the PRN has plenty of room for growth, the positive nature of the outcomes reported here were significant. Future areas of growth should focus on improving consistency of data entry especially around demographic and client history information as well as discharge data, and increasing the rate of participation of

pograms. The growth of the NATSAP PRN requires a high level of care and nurturing and without proper support and commitment, the full potential of this endeavor will not be reached.

References

Achenbach, T. M., & Edelbrock, C. (1991). The child behavior checklist and revised child behavior profile. Burlington, VT: University Associates in Psychiatry.

Behrens, E., & Satterfield, K. (2006, August). *Report of findings from a multi-center study of youth outcomes in private residential treatment*. Paper Presented at the 114th Annual Convention of the American Psychological Association. New Orleans, LA.

Burlingame, G. M., Wells, M. G., Cox, J. C., Lambert, M. J., Latkowski, M., & Justice, D. (2005). *Administration and scoring manual for the Y-OQ™*. Wharton, NJ: American Professional Credentialing Service.

Clothier, H. J. (2005). An evaluation of the Australian sentinel practice research network surveillance for influenza-like illness. *Communicable Diseases Intelligence*, *29*, 231–247.

Eschenbeck, H., Kohlman, C. W., & Lohaus, A. (2007). Gender differences in coping strategies in children and adolescents. *Journal of Individual Differences*, *28*(1), 18–26. doi:10.1027/1614-0001.28.1.18.

Gass, M. A. (2006). Evidenced-based research: Catalyst for action and future paradigm for NATSAP programs? *Journal of Therapeutic Schools and Programs*, *1*(2), 4-10.

Gass, M. A. (2005). *Statistical data on the social skills checklist* (SSC). Durham, NH: University of New Hampshire.

Gass, M. A., & Young, M. (2007). Dealing with issues of program effectiveness, cost benefit analysis, and treatment fidelity: The development of the NATSAP Research and Evaluation Network. *Journal of Therapeutic Schools and Programs*, *2*(1), 8-24.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Outcomes research in mental health: Systematic Review. *The British Journal of Psychiatry, 181*, 8-16. doi:10.1192/bjp.181.1.8

Gillis, H. L., & Speelman, E. (2008). Are challenge (ropes) courses an effective tool: A meta analysis. *Journal of Experiential Education*, *31*(2), 111-135.

Holloway, K. M. (2004). In search of therapy phases: A confirmatory factor analysis of change scores in the OQ-30. *Dissertation Abstracts International*, *65*(6-B), 3164. Retrieved from EBSCO host.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12-19.

Jones, M. D. (2004). Construct validity and construct stability of the Outcome Questionnaire-30: A longitudinal factor analysis. *Dissertation Abstracts International, 65*(6-B), 3166. Retrieved from *EBSCO host.*

Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., et al. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy, 3*, 249-258.

Luijsterburg, J., van den Bogaard, J., & de Vries Robbé, P. (2007). Variety in mental health research data: When does more become too much? *BMC Psychiatry (Online Version)*, 7:45. doi: 10.1186/1471-244X-7-45. Retrieved from http://

www.biomedcentral.com/1471-244X/7/45

Maschi, T., Morgan, K., Bradley, & C., Hatcher, S. S. (2007). Exploring gender differences on internalizing and externalizing behavior among maltreated youth: Implications for social work action. *Child and Adolescent Social Work Journal*, *25*, 531-547.

McMillen, J. C., Lenze, S.L., Hawley, K.M., & Osborne, V.A. (2009). Revisiting practice-based research networks as a platform for mental health services research, *Administration and Policy in Mental Health, 36*, 308-321. doi: 10.1007/s10488-009-0222-2

Mueller, R., Lambert, M. J., & Burlingame, G. M. (1998). Construct validity of the Outcome Questionnaire: A confirmatory factor analysis. *Journal of Personality Assessment, 70*, 248-262.

OQ Measures. (2011). Y-OQ® 30. Retrieved from http://www.oqmeasures.com/page.asp?PageId=75

Rosenheck, R. A., Fontanna, A., & Stolar, M. (1999). Assessing quality of care: administrative indicators and clinical outcomes in post traumatic stress disorder. *Medical Care, 37*, 180-188.

Russell, K. C. (2005). Two years later: A qualitative assessment of youth well-being and the role of aftercare in Outdoor Behavioral Healthcare treatment. *Child and Youth Care Forum, 34*(3), 209-239.

Russell, K. C. (2003). An assessment of outcomes in Outdoor Behavioral Healthcare treatment. *Child and Youth Care Forum, 32*(6), 355-381.

Stop Child Abuse in Residential Programs for Teens Act, H.R. 911, 111th Cong. (2009).

Wells, M. G., Burlingame, G. M., & Rose, P. (2003). *Manual for the Youth Outcome Questionnaire Self-Report*. Wharton, NJ: American Professional Credentialing Service.

Young, M & Gass, M. A. (2008). Current descriptions of National Association of Therapeutic Schools and Programs (NATSAP) members. *Journal of Therapeutic Schools and Programs, 3*(1), 162-186.

Young, M. & Gass, M. A. (2010). Preliminary data from the NATSAP Research and Evaluation Network: Client characteristics at admission. *Journal of Therapeutic Schools and Programs, 4*(1), 85-111.