Standardized Observation Analogue Procedure in the Treatment of Severe Childhood Aggression Study

Sabrina N. Grondhuis, PhD,¹ Cristan A. Farmer, PhD,² L. Eugene Arnold, MEd, MD,³ Kenneth D. Gadow, PhD,⁴ Robert L. Findling, MD, MBA,⁵ Brooke S.G. Molina, PhD,⁶ David J. Kolko, PhD,⁶ Kristin A. Buchan-Page, BA,⁷ Robert R. Rice, Jr., PhD,⁷ Eric M. Butter, PhD,^{7,8} and Michael G. Aman, PhD⁷

Abstract

Objective: To explore blinded observational outcomes in the Treatment of Severe Childhood Aggression (TOSCA) study.

Methods: During a 9-week acute trial, children with severe physical aggression and attention-deficit/hyperactivity disorder received parent training + titrated psychostimulant for 3 weeks, and those who failed to show an optimal response during Week 4 through Week 6 received in addition either randomly assigned placebo (Basic treatment) or titrated risperidone (Augmented treatment). Child and parent behaviors were videotaped in a Standardized Observation Analogue Procedure (SOAP) designed to elicit problems and strengths in child and parent interactions. SOAPs were collected at baseline and Week 9 and 52 follow-up.

Results: During the acute 9-week trial, augmented treatment was associated with better outcomes than basic treatment for 3 of 13 measures: increased Child Compliance ($p=0.004$; significant after correction for multiple tests), greater use of positive Parent Reinforcement ($p=0.03$), and more Shared Enjoyment ($p=0.04$). At follow-up, when medication was no longer by randomized assignment, parents used more Alpha Commands and displayed fewer Parent Negative Behaviors, and the dyads showed more Shared Enjoyment regardless of original randomization. Thus, there were better parent-child interactions with Augmented treatment, and interactions improved overall at follow-up regardless of original treatment assignment.

Conclusions: The SOAP demonstrated sensitivity to behavior changes between short-term treatments for a few (but not most) measures. The acute treatment differences for Child Compliance and Child Negative Behavior are generally consistent with the moderate superiority of Augmented over Basic treatment previously reported for the primary study outcome.

Keywords: observational measures, clinical trial, disruptive behavioral disorders, aggression, ADHD

Introduction

ASSESSMENT OF BEHAVIORAL PROBLEMS in large-scale clinical trials is often limited to questionnaires, checklists, or structured interviews to determine efficacy of study treatment. Such approaches are attractive, as they are able to assess symptoms across domains, are easy to administer, and can solicit responses from caregivers, teachers, and children, as appropriate. In addition,

these assessment measures have typically undergone prior psychometric evaluation and are able to discern medication effects, symptom severity changes, and other psychological outcomes (e.g., quality of life; Handen et al. 2013). Although these types of measures may be sufficient for evaluating safety and efficacy, in some cases, direct observation of participants can provide additional insight into behavior changes and can be a valuable supplement to typical assessment strategies.

¹Department of Psychology and Neuroscience, Millsaps College, Jackson, Mississippi.

²Interim Research Program, Department of Pediatrics and Developmental Neuroscience Branch, National Institutes of Health, Bethesda, Maryland.

³Department of Psychiatry, ⁷Nisonger Center, The Ohio State University, Columbus, Ohio.

⁴Department of Psychiatry, State University of New York, Stony Brook, New York.

⁵Division of Child & Adolescent Psychiatry, Bloomberg Children's Center, Baltimore, Maryland.

⁶Department of Psychiatry & Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania.

⁸Nationwide Children's Hospital, The Ohio State University, Columbus, Ohio.

Clinical Trial Registry: Treatment of Severe Childhood Aggression (TOSCA Study), URL: <http://clinicaltrials.gov>, unique identifier: NCT00796302

Funding: This study was supported by grants from National Institute of Mental Health (NIMH) to The Ohio State University (R01 MH077907), Case Western Reserve University (R01 MH077750), the University of Pittsburgh (R01 MH077676), and Stony Brook University (R01 MH 077997). The project was supported by a National Institutes of Health (NIH) General Clinical Research Center grant M01RR10710 (State University of New York Stony Brook) and Clinical and Translational Science Awards from the National Center for Advancing Translational Sciences: grants 8UL1TR000090-05 (The Ohio State University) and UL1 RR024153 and UL1TR000005 (University of Pittsburgh). The content is solely the responsibility of the authors and does not necessarily represent the official views of the respective National Centers for Advancing Translational Sciences or the NIH.

Direct observation protocols have been used to assess the symptoms of a variety of disorders and mental health concerns such as pervasive developmental disorder (PDD; Johnson et al. 2009), attention-deficit/hyperactivity disorder (ADHD; Cunningham and Barkley 1979; Wells et al. 2006), obsessive-compulsive disorder, and family conflict (Mash and Foster 2001). Child behaviors can be viewed as the intersection between child temperament and ability levels, situational demands, and interactions with other persons within the environment (Cunningham and Barkley 1979). This paradigm acknowledges that parents influence the behaviors of children, which then in turn impact parental behaviors, and so on in a virtuous or vicious cycle, and underscores the necessity for evaluating parent-child interactions to more fully understand response to intervention.

Direct observation protocols can be used to replicate this fluid dynamic by placing the child participant in a semicontrolled situation with her/his primary caregiver, which is designed to elicit behaviors analogous to those seen in the natural environment (Haynes 2001). The protocol structure and intentional constraints and demands differentiate direct observation procedures from passive naturalistic observations that monitor without interference.

Furthermore, direct observation techniques have been used by clinicians for decades to acquire objective measurements of interactions between parents and their children and potentially limit the bias associated with informant report (Roberts and Hope 2001). For instance, Cunningham and Barkley (1979) viewed boys with and without hyperactivity (a forerunner of ADHD) interacting with their parents. They found that hyperactive boys were less likely to stay on task and comply with instructions, and that mothers of hyperactive boys attempted to control and force structure to all observed activities, even those of free play, when compared to their nonhyperactive counterparts.

Direct observation has been used to evaluate behavioral changes in intervention and developmental research. Wells et al. (2006) compared parent-child interactions in the Multimodal Treatment Study of Children with ADHD (the MTA). They compared four possible treatment groups and observed greater improvement in those who received both medication and behavior therapy (compared with treatment as usual in the community, behavior therapy, and drug therapy alone).

Handen et al. (2013) evaluated parent/child dyads in a study of children with PDDs and behavior problems. All of the subjects received risperidone, and 60% of participants also received parent training in behavior management (PT). Measures of inappropriate behaviors and compliance improved for children at the study endpoint across treatment groups, and parents in both treatments demonstrated more reinforcement and fewer repeated demands. However, parents of children who received both medication and PT exhibited more reinforcement and fewer restrictive statements compared to parents of those who received medication alone. These findings suggest the possibility of improvements in the dynamics of the parent-child relationship as a function of both behavioral and pharmacologic treatment and raise the following question: Can this improvement be observed for children with severe aggression and ADHD?

The Treatment of Severe Childhood Aggression (TOSCA) study evaluated augmentation with risperidone for children with ADHD, concurrent disruptive behavior, and aggression. In the acute phase of the 9-week protocol, all child participants received individually titrated stimulant medication (STIM, usually OROS methylphenidate) for their ADHD symptoms, while all parents received a 9-session PT course during only the acute trial to provide additional tools for managing their child's behavior. The PT course was the Community Parent Education Program (Cunningham, 1998), and

all therapists completed a training course and participated in regular conference calls to ensure consistency.

If the child did not demonstrate significant symptom improvement by end of the third week (defined as within 0.5 standard deviation of the normal range of 15 or less on the Nisonger Child Behavior Rating Form [NCBRF] Disruptive Total score and a Clinical Global Impressions [CGI] rating of "Very much improved"), an additional treatment was added randomly. This was either placebo (PBO) for the Basic treatment condition (STIM + PT + PBO) or risperidone (RIS) for the Augmented treatment condition (STIM + PT + RIS), and were individually titrated to optimal response.

Findings from caregiver-completed behavior rating scales indicated that Augmented treatment resulted in greater parent-reported improvement in aggression and disruptive behaviors than Basic treatment (Aman et al. 2014). The study also included a 12-week extension of double-blind treatment for clinical responders (Findling et al. 2017) and a 52-week follow-up for all study participants (Gadow et al. 2016). Disruptive behaviors were substantially decreased from baseline for both treatment groups at the end of the 12-week extension, although scores worsened between Week 9 and the extension end point (Findling et al. 2017). Similarly, the 52-week follow-up also indicated improvement in disruptive behaviors from baseline for all participants, but the added benefit of Augmented treatment over Basic treatment attenuated during the time adherence to the original treatment assignment was lost (Gadow et al. 2016).

In addition to the use of well-validated rating scales for assessing response to treatment, the TOSCA study also included three administrations of the Standardized Observation Analogue Procedure (SOAP), a direct observation protocol, for each study participant at baseline and Week 9 (end of the acute trial) and Week 52 follow-up. This procedure was a modified version of the observational assessment developed and used by the Research Units on Pediatric Psychopharmacology (RUPP) group (Johnson et al. 2009; Handen et al. 2013). The TOSCA study therefore provides an opportunity to assess whether the SOAP is sensitive to behavioral changes due to treatment or over time. This study is the first to report the findings from SOAP observations of child participants and their primary caregivers (evaluated both individually and as pairs) during the TOSCA study.

The primary aim of this study was to evaluate the effect of Augmented versus Basic intervention during the acute trial on SOAP outcomes, with a secondary aim of evaluating the maintenance of SOAP scores 52 weeks postparticipation. For this secondary aim, we do not distinguish between randomization groups, as no constraints were placed on the treatment regimen of participants during the follow-up period. Only 43% of children randomized to Augmented and 36% of Basic continued with that original treatment (Gadow et al. 2016); many participants instituted the alternative regimen or adopted an entirely different medication. Our *a priori* hypothesis was that Week 9 SOAP ratings would be significantly better in the Augmented group than Basic group at the end of the acute trial. We made no specific hypothesis regarding the secondary (exploratory) aim of assessing maintenance of SOAP scores 1 year postparticipation.

Methods

Participants

TOSCA was a multisite, randomized controlled trial that included 168 children recruited from four sites (Case Western Reserve University, Ohio State University, University of Pittsburgh, and Stony Brook University) with the approval of each site's institutional review board. Participants included in this study were 149 children, 6 to 12 years of age, inclusive, who consented to

enroll in the TOSCA trial, and who completed at least one SOAP session at baseline, and Week 9 or Week 52. All children had diagnoses of ADHD (any subtype; Combined type $n=150$, Predominantly Inattentive type $n=13$, and Predominantly Hyperactive-Impulsive type $n=5$) and either Conduct Disorder ($n=37$) or Oppositional Defiant Disorder (ODD, $n=112$) according to DSM-IV-TR (Diagnostic and Statistical Manual of Mental Disorders, 4th edition—Text Revision; American Psychiatric Association [APA] 2000) diagnostic criteria.

Children additionally had serious physical aggression as determined by scores on the Overt Aggression Scale-M (Coccaro et al. 1991) and the Nisonger Child Behavior Rating Form D-Total (NCBRF; Aman et al. 2008). The sample was predominantly male ($n=114$, 77%) with average intelligence (IQ mean=98.13, $SD=14.19$), 8.85 years of age at initial screening ($SD=1.97$). Distribution of ADHD subtypes, Conduct Disorder, sex, age, and IQ was not statistically different across study treatment arms. Additional information about the TOSCA trial, including full sample characteristics, study design, or findings, may be found in Aman et al. (2014).

SOAP protocol

The TOSCA protocol indicated that the SOAP was to be administered at the baseline appointment and Week 9 (the end of the acute trial) and Week 52 follow-up. There were some unavoidable deviations, such as if subjects terminated the study early or were lost to follow-up. Study coordinators gave caregivers a standard list of 25 commands and asked them to choose 10 commands to administer to their child. Commands included a range of difficulty levels to account for the age span of enrolled children, and caregivers were allowed to pick commands with which they believed their child could comply, given age and ability levels. Sample commands included *come stand by me*, *jump on one foot five times*, or *put these word flash cards in alphabetical order*.

To complete the SOAP, each caregiver was outfitted with a small earpiece so the study coordinator could communicate with them. The child and caregiver were then led to a room equipped with a video recorder and that had a variety of toys, a table, and chairs. For the first 10 minutes of the session, the dyads were supposed to interact with each other and the toys as they normally would. Starting at 10 minutes, the coordinator instructed the caregiver through the earpiece to administer one of the predetermined commands on the list. The child then had 60 seconds to comply, at which point the coordinator instructed the caregiver to give the next command, regardless of level of compliance or completion of the previous command. This portion continued for a total of 10 minutes, giving the parent time to deliver 10 commands. At 20 minutes, the SOAP session was finished.

All SOAP recordings were sent to the Ohio State University site for coding using ProCoderDV (Tapp 2003) by a team of blinded coders trained for inter-rater reliability. This team was led by the first author and included three supervised undergraduate students. Reliability of coding using Cohen's kappa was predominantly moderate to substantial for the acute trial SOAPS, the main focus of this study, and ranged from 0.453, $p=0.001$ for Commands at Week 9, to 0.828, $p<0.001$ for Week 0 Reinforcement. Kendall's tau-b for ordinal data, in the case of the overall interaction and behavior codes, were strong for the acute trial and ranged from 0.606, $p=0.010$ for Overall Joint Interaction at Week 0, to 0.842, $p<0.001$ for Overall Child Behavior at Week 0. The same coder watched and coded each tape twice, once for compliance markers and once for relational markers.

Compliance markers were objective and included when commands were given, type of command (*Alpha Commands* that were clear and direct, such as "Put away the toys" or *Beta Commands* that were vague and typically worded as a question rather than a command statement such as "Could you please put the toys away?"), presence or absence of Child Compliance, Time (delay) to Compliance, and presence of parental Reinforcement. Due to the structure of the SOAP, these codes should only have occurred during the final 10 minutes of the session.

Relational codes were more subjective, and included coding Negative Child Behavior (yelling, aggression, leaving the room, and so on), Child Verbal Outbursts (disrespectful or deviant talk), Negative Parent Behavior (e.g., disdainful or hostile comments and threats of aversive consequences for noncompliance), Positive Interactions between parent and child (e.g., paying attention to one another and conversation relevant to discussion topic), and Shared Enjoyment (time when both parent and child were enthusiastically expressing joy, while engaged in the same activity). Relational behaviors could occur at any point during the SOAP recording.

After watching the recording twice, the coder then assigned three global codes for overall behavior quality, one for the parent, one for the child, and one for overall interaction, which rated the global impression of the dyad (1–9, with 9 being the most positive score). In the case of discrepancies for compliance or relational behaviors, we adopted the values coded by the primary coders who did the most coding and thus had the most experience with a given subject.

Child outcomes included five variables: Child Compliance quantity, Time to Compliance, number of Negative Child Behaviors, number of Child Verbal Outbursts, and the rating of Overall Child Behavior. Parent outcomes also consisted of five variables: number of Commands given, number of Alpha Commands, instances of Reinforcement, number of Parent Negative Behaviors, and the rating of Overall Parent Behavior. Joint outcomes only measured three variables: Positive Interaction, Shared Enjoyment, and the rating of Overall Joint Interaction. Previous studies that included direct observation protocols (Johnson et al. 2009; Handen et al. 2013) reported three child outcomes: Child Compliance, Negative Child Behaviors, and Child Verbal Outbursts. Therefore, we selected these three variables as our primary outcomes and the remaining 10 are reported as exploratory outcomes.

Statistical analysis

To answer the two research questions, a single linear mixed model was used. This model predicted SOAP variables at Week 9 and 52, controlling for the baseline value, diagnosis, and site. Repeated measurement within subjects was accounted for using an unstructured residual covariance matrix. A fixed effect of drug, which was assigned values of Basic or Augmented at Week 9 (depending on randomization) and Open at Week 52 (for all participants), was the primary effect of interest. Through planned contrasts, we evaluated the difference between Basic and Augmented at Week 9 (primary aim), and the difference between Week 9 and 52 ratings for the entire sample (regardless of randomization; secondary aim).

In several cases, exploratory data analysis indicated that the variable had a Poisson (for count variables: Child Negative Behavior, Child Verbal Outbursts, parent Reinforcement, Parent Negative Behavior, and Commands) or exponential (for continuous variables: Time to Compliance, Positive Interactions, and Shared Enjoyment) distribution. In these cases, a generalized linear mixed model with Poisson or exponential distribution (i.e., log link) was substituted. Child Compliance was squared before

analysis to correct a left skew in the distribution. Residuals from each model were plotted and checked for normality and influential outliers.

For the primary outcomes (Child Compliance, Negative Child Behaviors, and Child Verbal Outbursts), alpha was set to 0.017 for the Basic versus Augmented comparison, to account for repeated testing. Because non-normal distributions were employed for several outcomes, a single effect size like Cohen's *d* was not applicable to all models and was not calculated. For the Week 9 to 52 comparison, and for all secondary outcomes, alpha was not corrected (to reflect their exploratory nature and to guard against type 2 error when lack of change is the finding of interest). Analyses were conducted in SAS/STAT Version 9.3.

Of 149 participants with SOAP data, six (4%) were missing baseline data, 70 (47%) were missing one of two postbaseline assessments, and 23 (15%) had no postbaseline data. Maximum likelihood estimation was used, which allows for the inclusion of participants with at least one valid observation (i.e., those missing one postbaseline assessment); however, mean imputation was used for the six participants missing baseline data. Those missing all postbaseline data cannot contribute to the model and were therefore excluded, for a final sample size of 126.

To explore how listwise deletion of this latter group might bias the results, the baseline demographics and primary SOAP variables were compared between participants without postbaseline data ($n=23$) and those with baseline data ($n=120$). The excluded participants were more likely to be from two sites (Case, 27% of participants excluded; Stony Brook, 26%; Ohio State, 11%; Pittsburgh, 5%; $p=0.02$ for overall chi-square), but children with conduct disorder (17%) were not more likely than those with ODD (14%) to be excluded ($p=0.68$). The rate of exclusion also did not differ between males (16%) and females (16%; $p=0.94$).

Controlling for site, the two groups did not differ in age (9.11 ± 2.21 vs. 8.74 ± 1.88 , respectively; $p=0.77$). The groups did differ on IQ (excluded: 94.52 ± 13.86 and retained: 98.75 ± 14.44 ; $p=0.02$), and excluded participants had significantly better scores on SOAP Compliance (excluded: 8.86 ± 2.62 and retained: 8.03 ± 2.61 ; $p=0.02$), SOAP Negative Child Behavior (excluded, median: 1, and retained, median: 2; $p=0.003$), and SOAP Child Verbal Outbursts (excluded, median: 0, and retained, median: 1; $p<0.0001$).

Results

A Spearman correlation matrix was generated for all 13 SOAP variables (Table 1). The primary SOAP outcomes (Compliance, Child Negative Behaviors, and Child Verbal Outbursts) were moderately correlated (i.e., $r \leq 0.50$). Compliance was strongly correlated with Time to Compliance ($r=-0.73$) and Overall Child Behavior ($r=-0.69$). The SOAP parent variables were generally uncorrelated or only weakly correlated with each other and with child variables.

Acute phase

Primary outcomes. Least-square estimated values for each of the outcomes are shown in Table 2. Augmented treatment was associated with significantly better Compliance scores at Week 9 ($t[109]=-2.96$, $p=0.004$); the significance survives Bonferroni correction for the three primary outcomes. There was no difference in Child Negative Behaviors ($z=-0.68$, $p=0.49$) or in Child Verbal Outbursts ($z=1.47$, $p=0.14$).

Secondary outcomes. There was no difference between treatments groups for the secondary child variables of Time to Compliance ($z=1.49$, $p=0.14$) or Overall Child Behavior

TABLE 1. SPEARMAN CORRELATIONS AMONG DEPENDENT VARIABLES

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Compliance	1.00	-0.45**	-0.50**	-0.73**	0.69**	0.39**	0.44**	0.02	-0.06	0.02	0.09	0.01	0.52**
2. Child negative behaviors	-0.45**	1.00	0.51**	0.39**	-0.66**	-0.12	-0.34**	0.03	0.29**	0	0.02	0.01	-0.46**
3. Child verbal outbursts	-0.50**	0.51**	1.00	0.47**	-0.58**	0	-0.33**	0.05	0.21*	0.17*	-0.10	0.13	-0.38**
4. Time to compliance	-0.73**	0.39**	0.47**	1.00	-0.62**	-0.04	-0.44**	-0.05	0.13	-0.06	0.01	-0.12	-0.45**
5. Overall child behavior	0.69**	-0.66**	-0.58**	-0.62**	1.00	0.11	0.36**	0.06	-0.07	0.21*	0.04	0.06	0.79**
6. Commands given	0.39**	-0.12	0	-0.04	0.11	1.00	0.23**	0.04	0.06	-0.05	0.11	-0.03	0.07
7. Alpha commands	0.44**	-0.34**	-0.33**	-0.44**	0.36**	0.23**	1.00	0	-0.18*	-0.04	0.01	0.01	0.23**
8. Reinforcement	0.02	0.03	0.05	-0.05	0.06	0.04	0	1.00	0.02	0.23**	-0.18*	0.13	0.13
9. Parent negative behaviors	-0.06	0.29**	0.21*	0.13	-0.07	0.06	-0.18*	0.02	1.00	-0.24**	0.14	-0.09	-0.16
10. Overall parent behaviors	0.02	0	0.17*	-0.06	0.21*	-0.05	-0.04	0.23**	-0.24**	1.00	-0.20*	0.25**	0.60**
11. Positive interactions	0.09	0.02	-0.10	0.01	0.04	0.11	0.01	-0.18*	0.14	-0.20*	1.00	0.01	-0.01
12. Shared enjoyment	0.01	0.01	0.13	-0.12	0.06	-0.03	0.01	0.13	-0.09	0.25**	0.01	1.00	0.23**
13. Overall joint interaction	0.52**	-0.46**	-0.38**	-0.45**	0.79**	0.07	0.23**	0.13	-0.16	0.60**	-0.01	0.23**	1.00

* $p < 0.05$.

** $p < 0.01$.

TABLE 2. OBSERVED AND MODEL-ESTIMATED MEANS (SE) FOR STANDARDIZED OBSERVATION ANALOGUE PROCEDURE OUTCOMES

	Baseline: full sample <i>M (SD) or Median (IQR)</i>	Week 9: augmented <i>M (SE)</i>	Week 9: basic <i>M (SE)</i>	Week 52: full sample <i>M (SE) [mean change]</i>
Primary outcomes				
Child compliance	8.04 (2.55)	9.60 (2.23)*	8.51 (2.18)*	9.21 (1.67) [+1.56]
Child negative behavior ^a	2 (1–4)	2.02 (0.33)	1.73 (0.32)	1.23 (0.20) [–34%]**
Child verbal outbursts ^a	1 (0–3)	0.60 (0.14)	0.91 (0.24)	0.60 (0.15) [–19%]
Secondary outcomes				
Time to compliance ^b	116 (76–217)	100.41 (13.03)	125.73 (13.47)	95.76 (9.67) [–15%]
Overall child behavior	5.72 (1.78)	2.57 (0.45)	2.56 (0.44)	2.56 (0.44) [–0.13]
Commands given ^a	10 (10–10)	9.98 (0.18)	9.84 (0.08)	9.93 (0.12) [0.2%]
Alpha commands	5.39 (2.57)	2.42 (0.54)	2.39 (0.52)	2.60 (0.52) [+1.01]**
Reinforcement ^a	0 (0–1)	1.45 (0.27)*	0.92 (0.15)*	1.08 (0.15) [–7%]
Parent negative behaviors ^a	0 (0–1)	0.87 (0.21)	0.86 (0.25)	0.29 (0.08) [–67%]**
Overall parent behaviors	6.44 (1.39)	2.55 (0.43)	2.60 (0.42)	2.58 (0.40) [+0.05]
Positive interactions ^b	355 (91–578)	393.70 (41.84)	402.80 (46.27)	418.67 (38.97) [+5%]
Shared enjoyment ^b	0 (0–6)	13.24 (4.00)*	5.01 (2.03)*	24.06 (6.06) [+195%]**
Overall joint Interaction	6.07 (1.35)	2.57 (0.42)	2.59 (0.40)	2.56 (0.40) [–0.30]

Observed means are shown for baseline variables, which were included as covariates, and least square mean estimated values (with standard errors) are shown for Week 9 and 52 outcomes. Models included diagnosis and site as covariates. Poisson and exponential models index mean change on a percentage scale rather than on the Y-scale.

^aAnalyzed with Poisson distribution. Baseline values are medians (IQR).

^bAnalyzed with Exponential distribution. Baseline values are medians (IQR).

*Significant for the acute trial.

**Significant for Week 52 follow-up.

($t[110] = -0.33, p = 0.75$). Reinforcement was used more for children in the Augmented treatment than for those receiving Basic treatment ($z = -2.13, p = 0.03$), although there were no significant differences in Commands ($z = -0.75, p = 0.45$), Alpha Commands ($t[108] = -0.53, p = 0.60$), Negative Parent Behaviors ($z = -0.03, p = 0.98$), or Overall Parent Behavior ($t[107] = 1.02, p = 0.31$). The Augmented group engaged in more Shared Enjoyment than the Basic group ($z = -2.08, p = 0.04$); however, there was no significant difference in Positive Interactions ($z = 0.16, p = 0.87$) or Overall Joint Interaction ($t[109] = 0.60, p = 0.55$).

Week 52 follow-up

Primary outcomes. For the entire sample, the change from Week 9 to 52 in Compliance was nonsignificant ($t[62.6] = 0.62, p = 0.54$). While Child Verbal Outbursts did not change from Week 9 to 52 ($z = -0.82, p = 0.41$), Child Negative Behaviors decreased slightly ($z = -2.58, p = 0.01$), showing further improvement.

Secondary outcomes

For the entire sample, the change from Week 9 to 52 in Time to Compliance was nonsignificant ($z = -1.54, p = 0.12$), as was the Overall Child Behavior rating ($t[88.4] = -0.08, p = 0.93$). Parental use of Alpha Commands increased between Week 9 and 52 ($t[97.9] = 3.50, p = 0.0007$), while Negative Parent Behaviors decreased ($z = -3.72, p = 0.0002$). There were no significant differences in quantity of Commands ($z = 0.14, p = 0.89$), Reinforcement ($z = -0.50, p = 0.61$), or the Overall Parent Behavior ($t[86.7] = 0.02, p = 0.99$). More Shared Enjoyment was seen at Week 52 than at Week 9 ($z = 2.78, p = 0.005$), but there were no changes in Positive Interactions ($z = 0.45, p = 0.65$) or Overall Joint Interaction ($t[88.8] = -0.57, p = 0.57$).

Discussion

This investigation expanded on previous TOSCA findings by using a direct observation protocol to evaluate behavioral changes

in children with ADHD and ODD or CD over the course of a randomized, controlled trial of risperidone (vs. placebo) added to stimulant and parent training. We divided our interest into three quantifiable areas: child behaviors, parent behaviors, and child-parent interactions.

The results of linear mixed models indicated a significant effect of Augmented versus Basic treatment at Week 9 in one of the three primary outcomes, Child Compliance, at a p -value (0.004) that survived correction for multiple tests. This was consistent with our study hypothesis, and with the original trial results demonstrating a greater reduction in parent-reported disruptive behaviors with Augmented treatment (Aman et al. 2014). The other two primary outcomes, Child Negative Behaviors and Child Verbal Outbursts, did not demonstrate significant treatment differences, which could be due to the small variability in those constructs or may suggest that the superiority of Augmented over Basic is not associated with decreased levels of observed disruptive behaviors, but rather increased compliance.

Several secondary outcomes were also tested. Among the parent secondary outcomes, only reinforcement improved more under Augmented than Basic. This finding could be due to children complying more willingly, complying in a timelier manner, or complying without additional negative behaviors or language—all instances that would warrant reinforcement for these desired behaviors. It seems unlikely that this finding could be due exclusively to parents in the Augmented group utilizing PT more appropriately; all parents received the same management training, and the difference is likely due to more change in child behavior in the Augmented treatment offering more opportunity for reinforcement. This provides an interesting example of child medication changing parent behavior.

There was more Shared Enjoyment between caregivers and children in the Augmented group than in the Basic group, which is likely attributable to similar explanations as for reinforcement. Greater Child Compliance and fewer Child Negative Behaviors or Child Verbal Outbursts may increase the pleasure and satisfaction

found in these dyadic interactions for both parties. No other meaningful differences between groups were observed in secondary outcomes.

This study also included a 52-week follow-up, during which time families could pursue any treatment with their health providers. Thus, the analyses on these data focused on maintenance of behavior from Week 9 to 52, regardless of randomization during the acute phase. Among the primary outcomes, Child Negative Behaviors was the only variable that changed, improving from Week 9 to 52. This could reflect optimization of treatment, spontaneous remission of symptoms, maturation, therapist attention, or perhaps longer-term implementation of parent training. While Compliance and Child Verbal Outbursts did not improve during the follow-up period, neither did they significantly worsen. Although the absence of change is not evidence for an effect, the result does suggest that children at least maintained whatever improvements had been achieved by Week 9 of the trial, remarkable in view of the fact that 23% of the Augmented group and 11% of Basic group were taking no medication at Week 52.

Some improvements were also observed in secondary outcomes between Week 9 and 52. Alpha commands from parents were used more frequently, which is perhaps attributable to skills learned in PT during the acute trial to manage their child's behavior, consistent with existing literature evaluating PT success (Kazdin 1997). Caregivers in our study were also observed using fewer Parent Negative Behaviors (e.g., hostile comments directed at the child) at the Week 52 follow-up, which may have permitted the parent to be more engaged in the SOAP exchange, as suggested by previous research (Cunningham and Barkley 1979). Shared Enjoyment was more frequent at Week 52. Although this could be attributed to both child and caregiver having a better understanding of the SOAP expectations, the decrease in both Child Negative Behavior and Parent Negative Behavior at that final time point may have contributed to the higher quality of joint engagement.

Our findings were consistent with those of previous studies using direct observation. We were able to construct an analogue testing environment for our child-parent dyads that elicited undesirable behaviors and then demonstrated sensitivity to change over time for some of our salient variables (Hayes 2001). Our finding of an effect of treatment and amount of reinforcement is in line with research from Wells et al. (2006), who found the MTA multimodal treatment resulted in greater gains in proactive parenting over time, consistent with primary outcomes improving further (from 9 months to 14 months) after cessation of intensive behavior treatment (Arnold et al. 2004). The same group found that reduction of negative parenting mediated teacher-rated response to medication at school (Hinshaw et al. 2000). Handen et al. (2013) also found that caregivers in their combination treatment group used more reinforcement than those who received medication alone and reported similar difficulty in detecting changes to treatment due to high variance within the SOAP constructs being measured. The SOAP appears to be sensitive to change across treatment and time for selected outcomes and may act as an objective compliment to other study assessments.

Limitations

Although we were able to identify some compelling findings using the SOAP, there were several limitations as well. First, as in any clinical study, there were some barriers to standardization in the SOAP. Caregivers sometimes made up their own instructions

rather than following the directions given by study coordinators. This may have resulted in recordings that had a different number of commands or commands given in a different time frame than was intended.

Real-world scenarios were at times difficult to control, such as the presence of a second (typically infant or toddler) child in the room, use of a language other than English, fire alarms sounding and obscuring speech, or study participants eloping. Occasionally, loss of recordings, through either parental refusal to participate in the SOAP or compatibility problems with the ProCoderDV software, occurred. The observations were only a brief snapshot of interactions between child and caregiver, and may not accurately reflect family dynamics or behavioral difficulties taking place beyond the confines of the study rooms. Finally, this study did not include a no-treatment comparison group and we are unable to disentangle the effects of individual therapies other than Basic versus Augmented.

Conclusions

Parent-child interactions are filled with personality dynamics, power struggles, and, in the case of the participants in this study, the overarching influence of physical aggression and other disruptive behavior. Our findings indicate that the SOAP was sensitive to behavioral changes, and its inclusion in TOSCA provided an objective complement to the more subjective ratings. Although only a few outcomes investigated indicated improvement due to Augmented treatment in the 9-week acute trial, the presence of some instances, such as the amount of Compliance, is consistent with the moderate improvement over Basic treatment that was reported by Aman et al. (2014) and Gadow et al. (2015).

The modest improvements that persisted through to Week 52 follow-up should encourage clinicians that efficacious treatment options are available for this clinically severe and hard-to-treat population. This additional assessment may provide a more complete view of treatment effects in clinical trials by supplementing information collected by parent report or clinician interviews. Future clinical researchers should consider including direct observations to add unbiased information to their investigations.

Clinical Significance

The SOAP detected a treatment effect in Compliance, one of the three primary outcomes, in addition to Reinforcement and Shared Enjoyment. The SOAP also identified favorable changes between Week 9 and 52 for Child Negative Behavior, Alpha Commands, Parent Negative Behavior, and Shared Enjoyment. The addition of objectively assessed interactions between parent and child provides an added level of confidence to the study findings.

Moreover, the SOAPS are an apt way for study personnel to observe whether caregivers are using the skills taught to them in PT appropriately. Although the SOAP was not a primary outcome measure in TOSCA, being able to see that certain PT lessons (such as reductions in negative behavior) were identifiable to a greater extent at the end of the trial suggests that the intervention was impacting families in positive directions. Objective ratings of behavior, such as those provided by the SOAP, are a less biased way to observe the dynamics between parent and child. This protocol provides an important supplement to informant assessments that are subject to respondent mood, stress levels, and perceived item impact on quality of life.

Acknowledgment

We thank Kyle Yost, BA, Annemarie Krug, BA, and Eva Allibone, BA, for their assistance with coding.

Disclosures

Dr. Arnold has received research funding from Forest, Lilly, Noven, Shire, Supernus, Roche, and YoungLiving (as well as NIH and Autism Speaks), has consulted with Pfizer, Tris Pharma, and Waypoint, and has been on advisory boards for Arbor, Ironshore, Otsuka, Pfizer, Roche, Seaside Therapeutics, and Shire. Dr. Findling receives or has received research support, acted as a consultant, and/or served on a speaker's bureau for Acadia, Aevi, Akili, Allcobra, Allergan, Amerex, American Academy of Child & Adolescent Psychiatry, American Psychiatric Press, Arbor, Bracket, Daiichi-Sankyo, EPharma Solutions, Forest, Genentech, Insys, Ironshore, KemPharm, Luminopia, Lundbeck, Merck, NIH, Neurim, Noven, Nuvelton, Otsuka, PCORI, Pfizer, Physicians Postgraduate Press, Receptor Life Sciences, Roche, Sage, Shire, Sunovion, Supernus Pharmaceuticals, Syneurx, Teva, TouchPoint, Tris, and Validus. Dr. Molina has received conference funding from Shire and NEOS Therapeutics. Dr. Butter has consulted with CogState Clinical Trials, Bracket Global, and Hoffman-LaRoche as well as received research funding from NIH, MCHB/HRSA, Autism Speaks, the Simons Foundation, Department of Defense, and the Organization for Autism Research. Dr. Aman has received research contracts, consulted with, served on advisory boards, or done investigator training for Aevi Genomic Medicine, AMO Pharma, Bracket Global, CogState, Inc., CogState Clinical Trials, Ltd., Confluence Pharmaceutica, Coronado Biosciences, Hoffman-La Roche, Johnson and Johnson, Lumos Pharma, MedAvante, Inc., MedAvante-Prophase, Ovid Therapeutics, ProPhase LLC, Supernus Pharmaceuticals, and Zynerva Pharmaceuticals. He receives royalties from Slosson Educational Publications. Drs. Grondhuis, Farmer, Gadow, Kolko, Rice, Jr., and Ms. Buchan-Page have no conflicts of interest to disclose.

References

- Aman MG, Bukstein OG, Gadow KD, Arnold LE, Molina BS, Mcnamara NK, Rundberg-Rivera EV, Li X, Kipp H, Schneider J, Butter EM, Baker J, Sprafkin J, Rice RR, Bangalore SS, Farmer CA, Austin AB, Buchan-Page KA, Brown NV, Hurt EA, Grondhuis SN, Findling RL: What does risperidone add to parent training and stimulant for severe aggression in child attention-deficit/hyperactivity disorder? *J Am Acad Child Adolesc Psychiatry* 53:47–60, 2014.
- Aman M, Leone S, Lecavalier L, Park L, Buican B, Coury D: The Nisonger Child Behavior Rating Form: Typical IQ version. *Int Clin Psychopharmacol* 23:232–242, 2008.
- Arnold LE, Chuang S, Davies M, Abikoff HB, Conners CK, Elliott GR, Greenhill LL, Hechtman L, Hinshaw SP, Hoza B, Jensen PS, Kraemer HC, Langworthy-Lam KS, March JS, Newcorn JH, Pelham WE, Severe JB, Swanson JM, Vitiello B, Wells KC, Wigal T: Nine months of multicomponent behavioral treatment for ADHD and effectiveness of MTA fading procedures. *J Abnorm Child Psychol* 32:39–51, 2004.
- Coccaro EF, Harvey PH, Kurpshaw-Lawrence E, Herbert JL, Berstein DP: Development of neuropharmacologically based behavioral assessments of impulsive aggressive behavior. *J Neuropsychiatry Clin Neurosci* 3:44–51, 1991.
- Cunningham CE, Barkley RA: The interactions of normal and hyperactive children with their mothers in free play and structured tasks. *Child Dev* 50:217–224, 1979.
- Cunningham, CE: Large group, community based, parenting courses. In: Barkley, RA, ed. *Attention Deficit Hyperactivity: A Handbook for Diagnosis and Treatment*. New York: Guilford Press, 1998, pp. 394–412.
- Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. Washington, DC: American Psychiatric Association; 2000.
- Findling RL, Townsend L, Brown NV, Arnold LE, Gadow KD, Kolko DJ, Mcnamara NK, Gary DS, Kaplin DB, Farmer CA, Kipp H, Williams C, Butter EM, Bukstein OG, Rice R, Buchan-Page K, Molina BS, Aman MG: The treatment of severe childhood aggression study: 12 weeks of extended, blinded treatment in clinical responders. *J Child Adolesc Psychopharmacol* 27:52–65, 2017.
- Gadow KD, Brown NV, Arnold LE, Buchan-Page KA, Bukstein OG, Butter E, Farmer CA, Findling RL, Kolko DJ, Molina BS, Rice RR, Schneider J, Aman MG: Severely aggressive children receiving stimulant medication versus stimulant and risperidone: 12-month follow-up of the TOSCA trial. *J Am Acad Child Adolesc Psychiatry* 55:469–478, 2016.
- Handen BL, Johnson CR, Butter EM, Lecavalier L, Scahill L, Aman MG, McDougall CJ, Arnold LE, Swiezy NB, Sukhodolsky DG, Mulick JA, White SW, Bearss K, Hollway JA, Stigler KA, Dziura J, Yu S, Sacco K, Vitiello B: Use of a direct observational measure in a trial of risperidone and parent training in children with pervasive developmental disorders. *J Dev Phys Disabil* 25:355–371, 2013.
- Haynes SN: Clinical applications of analogue behavioral observation: Dimensions of psychometric evaluation. *Psychol Assess* 13:73–85, 2001.
- Hinshaw SP, Owens EB, Wells KC, Kraemer HC, Abikoff HB, Arnold LE, Conners C, Elliott GL, Greenhill LL, Hechtman LT, Hoza B, Jensen PS, March JS, Newcorn JH, Pelham WE, Swanson JM, Vitiello B, Wigal T: Family processes and treatment outcome in the MTA: Negative/ineffective parenting practices in relation to multimodal treatment. *J Abnorm Child Psychol* 28:555–568, 2000.
- Johnson CR, Butter EM, Handen BL, Sukhodolsky DG, Mulick J, Lecavalier L, Aman MG, Arnold EL, Scahill L, Swiezy N, Sacco K, Stigler KA, McDougall CJ: Standardised Observation Analogue Procedure (SOAP) for assessing parent and child behaviours in clinical trials. *J Intellect Dev Disabil* 34:230–238, 2009.
- Kazdin AE: Parent management training: Evidence, outcomes, and issues. *J Am Acad Child Adolesc Psychiatry* 36:1349–1356, 1997.
- Mash EJ, Foster SL: Exporting analogue behavioral observation from research to clinical practice: Useful or cost-defective? *Psychol Assess* 13:86–98, 2001.
- Roberts MW, Hope DA: Clinic observations of structured parent-child interaction designed to evaluate externalizing disorders. *Psychol Assess* 13:46–58, 2001.
- Tapp J: ProCoder for Digital Video: User Manual. Nashville, TN: The John F. Kennedy Center, Vanderbilt University, 2003.
- Wells KC, Chi TC, Hinshaw SP, Epstein JN, Pfiffner L, Nebel-Schwalm M, Owens EB, Arnold LE, Abikoff HB, Conners CK, Elliott GR, Greenhill LL, Hechtman L, Hoza B, Jensen PS, March J, Newcorn, JH, Pelham WE, Severe JB, Swanson J, Vitiello B, Wigal T: Treatment-related changes in objectively measured parenting behaviors in the multimodal treatment study of children with attention-deficit/hyperactivity disorder. *J Consult Clin Psychol* 74:649–657, 2006.

Address correspondence to:

Sabrina N. Grondhuis, PhD

Department of Psychology and Neuroscience

Millsaps College

1701 N. State Street

Jackson, MS 39210

E-mail: sabrina.grondhuis@millsaps.edu