Published in final edited form as:

J Am Acad Child Adolesc Psychiatry. 2012 November ; 51(11): 1173–1184. doi:10.1016/j.jaac.2012.08.028.

RUPP AUTISM NETWORK RANDOMIZED CLINICAL TRIAL OF PARENT TRAINING AND MEDICATION: ONE-YEAR FOLLOW-UP

L. Eugene Arnold, M.D., Michael G. Aman, Ph.D., Xiaobai Li, Ph.D., Eric Butter, Ph.D., Kristina Humphries, M.S., Laurence Scahill, R.N., Ph.D., Luc Lecavalier, Ph.D., Christopher J. McDougle, M.D., Naomi B. Swiezy, Ph.D., Benjamin Handen, Ph.D., Krystina Wilson, B.A., and Kimberly A. Stigler, M.D.

Abstract

Objective—To follow up on a 3-site 24-week randomized clinical trial (N=124) comparing antipsychotic medication alone (MED) to antipsychotic plus parent training (PT) in behavior management (COMB) for autism spectrum disorders with severe behavior problems. COMB had shown a significant advantage for child behavioral noncompliance ($p=.006$, $d=.34$), irritability ($p=.01$, $d=.48$), and hyperactivity/noncompliance ($p=.04$, $d=.55$), with a lower medication dose.

Method: A year after each participant's termination we priority-mailed an assessment packet—with a return-addressed envelope; a phone call alerted the family. Failure to return packets within a month elicited recontact and offers to resend.

Results—Eighty-seven of 124 families (70.2%) participated in follow-up. The improvement difference between treatments attenuated from post-treatment to follow-up for noncompliance ($d=0.32$ to $d=0.12$) and irritability ($d=0.46$ to $d=0.03$). Follow-up differences were nonsignificant. (The noncompliance difference was nonsignificant also at post-treatment for these 87.) 67% of COMB and 53% of MED were still taking risperidone, the original study medication. Most had needed dose adjustments or additional medication, and COMB no longer had a significantly lower dose. All COMB families but only 39% of MED reported seeking PT post-treatment. Daily living skills improvement during treatment predicted noncompliance improvement at follow-up for COMB, but noncompliance deterioration, and especially hyperactivity/noncompliance deterioration for MED-only children.

Conclusions—Study treatment experience/familiarity greatly influenced follow-up treatment: those who had received PT reported seeking it while those who had not experienced it tended not to seek it. The superiority of COMB over MED at post-treatment attenuated by over half at follow-up.

Keywords

autism; antipsychotic; parent training; follow-up; clinical trial

INTRODUCTION/BACKGROUND

This article reports a follow-up of a three-site 24-week randomized clinical trial¹ that compared antipsychotic medication alone (MED, risperidone; aripiprazole for risperidone nonresponders) to the same medication regimen plus parent training in behavior management (combination=COMB) for 124 children with autism spectrum disorders (ASDs) and severe irritability (aggression, severe tantrums, self-injury). The highly

significant efficacy of the 2 drugs for irritability of autism had already been proven previously (e.g., <http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2006/ucm108759.htm>) and this study examined the additive benefit from parent training. The main-study findings showed both treatment groups improving dramatically from baseline to 24 weeks, with a significant advantage of COMB for child noncompliance with instructions ($p=.006$, $d=.34$), irritability ($p=.01$, $d=.48$), and hyperactivity/noncompliance ($p=.04$, $d=.55$). More details are reported by Aman et al¹. This study follows up on the sample a year after the study treatment ended (18 months after baseline) to check for enduring effects.

The response curve reported by Aman et al¹ showed that although the COMB group improved more slowly (initially lagging the MED group), they caught up by 16 weeks and surpassed the MED group at 24 weeks. [EDITOR, THIS IS THE FIRST OF TWO TIMES THIS GRAPH IS REFERRED TO; SHOULD IT BE REPRODUCED FOR THE CONVENIENCE OF THE READERS?] This raised a question regarding what would happen after study treatment ended: would there be further improvement over the MED group, maintenance of the advantage, or attenuation or even dissipation of the advantage?

The only hypothesis was that the advantage of COMB would still be detectable a year later, but considerably attenuated as many in the MED-alone group received parent training (PT) in the community similar to what the COMB group had received in the study. Additional explorations were also pursued, including a possible mediator.

METHOD

The Research Units on Pediatric Psychopharmacology (RUPP) Autism Network conducted a three-site trial of the additive benefit of parent training (PT) in children taking well-managed FDA-approved medication for irritability of autism (manifested in tantrums, aggression, and/or self-injury). In a randomized, parallel-groups clinical trial, 124 children, aged 4 through 13 years with autism spectrum disorders (ASDs, autistic disorder, Asperger's disorder, pervasive developmental disorder not otherwise specified) accompanied by severe tantrums, self-injury, and aggression were randomized 3:2 to COMB ($n = 75$) or MED alone ($n = 49$) for 24 weeks. All participants received risperidone monotherapy from 0.5 to 3.5 mg/day, with switch to aripiprazole if risperidone was ineffective (aripiprazole started at 2 mg and could be adjusted up to 15 mg). Parents in the COMB group also received a mean of 10.9 PT sessions using a manual developed for the purpose² and successfully piloted before the main study³. The PT intervention comprised 11 core treatment sessions, three optional sessions, and up to three booster sessions, all individual and lasting 60–90 minutes each. Topics included preventive approaches, positive reinforcement, compliance, and teaching functional communication skills and specific adaptive skills. Direct instruction, activity sheets, video vignettes, and behavior rehearsal were all utilized. Parents were assigned individualized homework between sessions and were taught to collect data on children's behavior.

The primary measure of child behavioral compliance was the Home Situations Questionnaire (HSQ) score⁴ adapted to the ASD population⁵. Two of the secondary measures were the Aberrant Behavior Checklist⁶ Irritability and Hyperactivity/Noncompliance subscales. The results of the 24-week study were reported previously¹: the primary outcome measure (HSQ) showed a significant effect on child compliance with instructions ($p=0.006$, effect size $d=0.34$).

The study coordinator contacted all families approximately 18 months after study baseline, about a year after they terminated the main study, using documents and procedures approved by each institution's IRB. The coordinator mailed the assessment packet by priority mail or

courier with a return-addressed envelope, and phoned to alert the family. Those who did not return the packet within a month were recontacted to make sure they received it, and an offer was made to resend it if necessary. The usual methods of tracing lost families were followed, including working through their current clinicians.

Measures

The following measures were filled out by the primary caregivers, who also filled out the baseline measures:

1. The **Home Situations Questionnaire (HSQ)**⁴, the primary outcome measure for both the original study and the follow-up, is a parent-rated scale that captures noncompliant behavior in everyday circumstances (e.g., getting dressed for school, when caregiver is occupied on the telephone) over the previous two weeks. Questions answered affirmatively (i.e., noncompliance) are then scored on a Likert scale of 1 to 9 (ranging from mild to very severe). We chose this as the primary outcome measure because our PT package focused on reducing noncompliance to promote daily living skills. We added five items to make the instrument more pertinent to children with ASDs (e.g., “When there is an unexpected change in daily routine”), making a total of 25 items. The HSQ score was calculated by adding values from those rated on the Likert scale and dividing by 25. To assess internal consistency of the 25-item HSQ, we computed coefficient alpha⁷ for the baseline ratings; alpha was 0.92, suggesting excellent internal consistency.
2. The **Aberrant Behavior Checklist (ABC)**^{6,8,9} is a 58-item scale empirically developed to assess treatment effects in individuals with developmental disabilities. Its subscales are designated as Irritability (15 items), Social Withdrawal (16 items), Stereotypic Behavior (7 items), Hyperactivity/Noncompliance (16 items), and Inappropriate Speech (4 items). Psychometric characteristics of the ABC are very good.
3. A **Medication and Services-Use History** was devised for this follow-up, covering the period from study endpoint to the follow-up. It included questions about current use of behavioral management techniques and advice, other nonpharmacological treatments, medicines taken since leaving study treatment, and changes in dose or drug. See Appendix I for a copy of the form.
4. The *Vineland Adaptive Behavior Scales (Vineland)*¹⁰ is a parent interview assessing functional skills in three domains (Communication, Socialization, and Daily Living Skills) and an overall Adaptive Composite Score. Vineland items are scored 0 (behavior not performed), 1 (performed sometimes) or 2 (performed regularly). The instrument provides both standard scores and Age Equivalent scores. The latter, measured over time, yields an estimate of improvement in months.

Statistical Analysis

Baseline data (demographics, clinical characteristics) and outcomes collected at the end of the main 24-week study for subjects with follow-up were compared with those for subjects without follow-up. Categorical variables were summarized in frequencies and percentages. For continuous variables, mean, standard deviation, median, and range were reported. For contingency tables, Fisher’s exact test was used to calculate the p-values. Two-sample *t*-test was used to examine the difference in continuous outcome measures between independent groups if deemed appropriate. Otherwise, Wilcoxon ranked sum test was used. For the 87 subjects with follow-up, the HSQ mean scores and the five ABC subscales scores were summarized according to the treatment groups (COMB vs. MED) assigned in the study for

baseline, end of study treatment, and end of follow-up. Difference in change scores from baseline was assessed by two-sample *t* test and the corresponding effect size was presented. Parents' use of behavior management skills and the subjects' medication history were described.

Analysis of covariance (ANCOVA) was used to check the predictor/mediator effect of change in Vineland Adaptive Behavior Scale age-equivalent scores from baseline to 24 weeks on the change in behavioral compliance (HSQ mean) and irritability (ABC irritability subscale) from week 24 to follow-up. In the models, we included the Vineland score change as covariate, the treatment groups (COMB vs. MED), and their interaction. The analysis was repeated for the effects of four Vineland change scores: adaptive behavior composite standard score and age-equivalent gains for communication, daily living skills, and socialization. For these exploratory analyses, the significance level was set at two-sided $P < 0.05$ for all tests.

RESULTS

Of the 124 children who were randomized and analyzed for the original report (Aman et al, 2009), follow-up data were obtained from 87 (70.2%). The mean length of follow-up was 18.5 ± 5.2 months from original baseline (median 17 months, range 12.5–37.5 months). Table 1 compares those who contributed follow-up data to those who were unavailable. In general, differences between participants and those lost to follow-up were not significant statistically or clinically. Although the difference in number of yes responses on the HSQ (number of situations leading to noncompliant behavior) was significantly different at baseline, it was not at end of study treatment. The mean severity on HSQ also showed the same tendency (follow-up participants worse at baseline, not different at 24 weeks), but not significantly so. Interestingly, all eight of the Asperger's disorder cases in the original study participated in the follow-up, in contrast to those with autistic disorder and pervasive developmental disorder not otherwise specified.

Retention was not significantly different by treatment group: 36 of 49 (73%) assigned to MED and 51 of 75 (68%) assigned to COMB participated in follow-up. The duration of follow-up was also not significantly different by treatment assignment: The mean (sd), median, and range for MED were 17.8 (4.7), 16.3, and 13–32.6 months, and for COMB were 19.0 (5.5), 17.3, and 12.5–37.5 months.

For the 87 available at follow-up, Table 2 compares the original two randomized groups on the HSQ (primary outcome) and the 5 subscales of the ABC. Note that although COMB continues to show a nominal advantage over MED at follow-up, this attenuated to half the difference found at 24 weeks and was not significant. The failure of statistical significance may be at least partly a power problem because for these 87 subjects, considerably reduced in number from the 124 in the original 24-week intent-to-treat analysis, even the 24-week comparison was not significant.

Figure 1 shows the trajectory of HSQ scores from baseline to follow-up a year after end of study treatment for the 87 with follow-up. Note the decrease in noncompliance symptoms from baseline to week 24 (end of study treatment), largely a medication effect, but with nominally more improvement for the group also getting PT. Importantly, the improvement is largely maintained over the following year, with slight regression towards baseline, slightly greater in the COMB group. Although COMB visually separates from MED at both 24 weeks and FU, the separation is not significant at either time for the subgroup available at follow-up.

Table 3 shows use of medication and behavior management techniques at follow-up by assigned treatment group. The differences in medication use were negligible and nonsignificant. For those still taking risperidone, the dose was no longer lower for COMB as it had been at 24 weeks. However, there was a highly significant difference in reported seeking of parent training and using the behavior management techniques learned in PT. All (100%) of those originally assigned to PT (COMB group) reported continuing to seek more PT, while only 39% of those originally assigned to MED-alone reported seeking it ($p<.0001$) despite a routine recommendation at week 24 to do so, accompanied by phone numbers for resources from which to obtain PT. At follow-up 94% of those originally assigned to COMB were implementing behavior management techniques, compared to 53% of those assigned to MED ($p<.0001$). Among those who received PT in the study, the sessions considered most helpful were Behavioral Principles, Daily Schedules, Planned Ignoring, and Reinforcement.

The predictor/mediator analyses correlating Vineland gains from baseline to 24 weeks with the HSQ and irritability changes from 24 weeks to follow-up were generally noncontributory. The rather flat correlation slopes were mostly parallel between COMB and MED, with p values nonsignificant both for general prediction regardless of original treatment group and for interaction of main effect with group. The exception was the mediating effect of improvement in Vineland daily living skills on later changes in the HSQ (Fig. 2). Those assigned to MED alone had a positive correlation slope such that the more they improved in daily living skills from baseline to 24 weeks, the more their HSQ severity score increased from week 24 to follow-up (showing deterioration). However, those assigned to COMB had a negative slope such that the more improvement they showed in daily living skills from baseline to week 24, the more their HSQ score decreased from week 24 to follow-up (showing improvement). The interaction on this primary outcome was significant ($p=0.02$). The same Vineland variable showed a marginal trend of mediation ($p=0.08$) on the ABC Irritability Subscale (the planned secondary exploratory analysis). An additional post-hoc exploratory analysis showed a significant ($p<0.02$) mediating effect on the ABC hyperactivity/noncompliance subscale with the same pattern as found for the HSQ; however, the significant interaction for this outcome was due to the MED-alone group deteriorating more 24 weeks to follow-up if they improved more on Vineland baseline to 24 weeks while the COMB group remained unaffected by preceding Vineland improvement. Further, there was a general predicting effect ($p=0.005$) of improvement in Vineland being followed by worsening of hyperactivity/noncompliance, driven by the correlation of the MED-alone group.

DISCUSSION

This one-year follow-up of a 24-week study, with 70% retention at 17–18 months, found that that all participants continued to be improved from baseline (possibly due to maturation or continued effectiveness of medication). However, the advantage of adding parent training in behavior management techniques to highly effective FDA-approved medication for severe tantrums/irritability, self-injury, and aggression in ASD attenuated by half from its significant post-treatment advantage during the intervening year. While the loss of power from attrition cannot completely explain the loss of significance, it certainly contributes to it, as illustrated by the fact that the reduced sample size no longer yielded a significant difference even at the 24-week posttest. However, the raw data and relative effect sizes also suggest a real loss of clinical significance from 24 weeks to follow-up. Possibly this illustrates a need for systematic planned booster PT sessions. Another possible factor in loss of significance may have been the slight (nonsignificant) retention bias for those who did not show as much COMB advantage at week 24 as the whole sample: The difference in change-score improvement on the HSQ mean severity at 24 weeks was 0.60 ($p=0.006$) for the whole sample, compared to 0.52 ($p=0.14$) at 24 weeks for those retained at follow-up.

Perhaps the most encouraging finding is the clinically significant improvement from baseline for all participants.

The graph of the primary outcome (HSQ mean) included in the original report (Aman et al, 2009) showed that in the first month the MED-alone group fared modestly better, but by the end of 6-month treatment the Comb group caught up and significantly, albeit modestly, surpassed them. Those results were interpreted as showing slight detracting from risperidone benefit from behavioral “extinction burst” early in the PT of the Comb group, with later success of the PT. [EDITOR: THIS IS THE SECOND TIME THIS GRAPH IS REFERRED TO, THIS TIME AT REQUEST OF A REVIEWER. SHOULD THE GRAPH BE REPRODUCED FOR THE CONVENIENCE OF THE READERS?] The significant advantage of Comb over med-alone at 24 weeks was small (effect size $d=0.34$) but highly significant ($p=0.006$) and clinically meaningful in that it was added to the very large proven effect of the antipsychotic agent, which left little variance for showing further improvement. Unfortunately, the power lost by attrition during the follow-up period made statistical comparisons uninformative for the follow-up, but inspection of Fig. 1 shows the expected slight regression for both groups, more for the Comb group, which had improved the most. The advantage of the Comb group, although attenuated, is still apparent a year after the end of study treatment; but its near-negligible effect size ($d=0.12$) may not be clinically meaningful in many cases. The other two notable outcomes, which showed a medium clinically significant effect at 24 weeks even in the reduced follow-up sample, the irritability subscale ($d=0.46$, $p=0.038$) and hyperactivity/noncompliance subscale ($d=0.51$, $p=0.02$), attenuated to practically no difference ($d=0.03$ & $d=0.01$ respectively).

A common bane of follow-up studies is biased attrition, which may be relevant to the loss of the 24-week significance for the retained subsample. Inspection of Table 1 discloses only 2 significant differences between the retained and lost participants. 1) In contrast to the other two diagnoses, all 8 of the Asperger’s participants were retained whereas only 20 of 35 PDD-NOS participants were retained. However, the majority of both retained and lost participants had autistic disorder, making this an unlikely explanation. Nevertheless, it is conceivable that participants with Asperger’s received less benefit from added PT or that those with PDD-NOS received more benefit. 2) Those taking risperidone at the end of the study were retained at a significantly higher rate (76% vs. 38% for taking risperidone at week 24, 85% vs. 57% for taking it at last observation). The design was that risperidone nonresponders would switch to different medication while continuing in the study. Thus it appears that risperidone responders were preferentially retained in the follow-up. It is possible that risperidone responders showed less advantage from adding PT than did nonresponders; e.g., a poor response to risperidone may predict a poor response to other drugs, leaving more room to show improvement from PT, whereas a good risperidone response would leave less room to show additional improvement. This would be consistent with the 24-week effect size of the retained subsample being $d=0.52$ compared to $d=0.6$ for the whole sample. It is impossible to tell whether retention of the whole sample might have retained the significant Comb advantage through follow-up.

Another possibility to explain the attenuated effect at follow-up could be that the PT itself had little specific effect and that the advantage it showed was due to such nonspecific variables as the therapist-parent relationship, the added attention for the whole family, the expectation of improvement, and the need to find improvement after investing so much effort in learning the behavioral principles (the parents were raters of the primary outcome, the HSQ). Once the therapist contact and attention to the family were withdrawn, regression could commence. This speculation relates to a core issue in evaluating behavioral interventions, which are usually impossible to blind, in contrast to studies of ingestibles¹¹.

The most significant follow-up finding concerned service utilization. A clinically and statistically significant difference emerged in the use of behavior management techniques, as expected, but also in the seeking of further PT after the study. It is a bit surprising that all of those who had already received PT would seek more but less than half of those who did not receive it would obtain PT even though it was recommended at the end of the main study (families were given referral information, including phone numbers). PT is usually considered a finite package, and one might have expected that those who already had it would not need more. Conversely, we assumed that those who did not already receive PT as part of the study would want it, given that they had volunteered for a trial in which there was a 3/5 chance of getting it. It appears that the parents were influenced by their experience and tended to continue the treatment with which they were familiar.

An explanation for this paradoxical service utilization pattern may come from the parental commitment required for parent-mediated behavioral treatment. Parent Training is more time-intensive than medication. Parents are asked to modify their own behaviors and practices to support behavior change in their children. Most benefits are seen when the interventions are applied with consistency. This takes a great deal of coordination and work beyond the time of visits, which may not be appealing to all, especially after the child's problems have been ameliorated for several months by medication. For many, administering a drug is much more convenient and palatable. In some cases, parent trainers have to "sell" the behavioral techniques. Perhaps those assigned to PT were "sold" by the experience and the results they saw; hence their seeking of post-study PT was a tacit recognition of a need for booster sessions. Conversely, those who had not experienced PT were less eager to engage in such a labor-intensive undertaking, for which they no longer saw a pressing need. Yet, despite the fact that many families in the COMB group continued to seek post-study parent training, the differences between treatment groups at follow-up narrowed.

Consistent with the tendency to continue the familiar, about 2/3 of both treatment groups, both of whom received antipsychotic medication in the study, continued with it at follow-up. For those still taking risperidone, the COMB group, which had a significantly lower optimal dose (1.98 vs. 2.26 mg/day, $p=0.04$) at 24 weeks for the whole sample, no longer had a lower dose at follow-up: 2.30 vs. 2.33 mg/day). The loss of this significant effect seems more due to increases in the COMB group than decreases in the MED group after study end.

The results of the predictor/mediator analyses suggest that gains in daily living skills may be important for later compliance and control of irritability, but favorably so only in the context of parents trained to apply behavior modification. There was a significant interaction such that children in the COMB group fared better at follow-up the more gains they showed in daily living skills at 24 weeks, but the opposite pertained for the MED group: the more their daily living skill gains at 24 weeks, the worse their follow-up outcome. Remember that the Vineland daily living skills improvement was measured from baseline to 24 weeks and the change in HSQ (and ABC irritability and hyperactivity/noncompliance subscales) was from week 24 to follow-up, so the Vineland improvement preceded the HSQ change. Although suggestive, this does not prove causality. A previous paper¹² in this journal reported a significant gain in Vineland age-equivalent scores baseline-to-24-weeks paralleling the significant HSQ improvement in showing superiority of the COMB treatment and suggested that HSQ improvement preceded Vineland improvement. The follow-up budget did not allow for a follow-up Vineland, which might have been enlightening. Perhaps the positive reinforcement that trained parents continued to provide "kept the momentum going." In contrast, the gains in daily living skills with MED alone may have come too easily and raised the parental expectations to an impractical level, with those who had the most gains the first 24 weeks regressing to the mean in the absence of systematic parental reinforcement. However, we must note that the significance level for this exploratory

finding was only $p=0.02$, which would not have survived correction for multiple testing among the eight predictor/mediator models tested.

Further, a post-hoc exploratory analysis with the secondary outcome of ABC hyperactivity/noncompliance was only partially compatible with the two planned analyses. It did show the same direction of significant mediating interaction as the HSQ, but also showed a general prediction of hyperactivity/noncompliance the opposite of the expected direction: for the whole sample, the greater the baseline to 24 wk Vineland gains, the worse the 24 wk to follow-up change in hyperactivity/noncompliance (driven by the MED-only group). Potential explanations for this finding would be the same as those above for the MED group, including regression to the mean: Because Vineland and other clinical outcomes were somewhat collinear in the baseline-to-24-week change¹¹, Vineland improvement could be a marker of hyperactivity/noncompliance improvement at 24 weeks; and we might expect those with the most improvement on both at 24 weeks to deteriorate the most between 24 weeks and follow-up. Because we do not have follow-up Vineland scores, we cannot test that speculation. Nevertheless, parental exposure to PT prior to 24 weeks appeared to prevent the “regression to the mean” on hyperactivity/noncompliance experienced by the MED-alone group, analogous to what happened with the HSQ and marginally with ABC irritability: With parent training, either greater improvement on the Vineland at 24 weeks was followed by greater improvement on the behavioral variables or there was no effect, whereas without PT, greater Vineland improvement was consistently associated with worse later behavioral outcomes.

The attenuation of primary outcome at follow-up after a highly significant finding at study treatment end is not unusual and has been reported for other disorders. A good example with the more common attention-deficit/hyperactivity disorder (ADHD) comes from the multi-site Multimodal Treatment Study of ADHD (the MTA)¹³. The MTA randomized 589 children with ADHD to four treatments, MED-alone, behavioral-alone, the Combination (COMB), and community treatment as usual for 14 months. At treatment end, the two groups with study-managed medication improved more than the other two groups. Although the MTA reported continued statistical significance at the 10-month follow-up of the 14-month treatment study (at the 24-month point from original baseline), the effect size attenuated by half.¹⁴ Statistical significance was retained by sheer power of sample size (over 500), but the effect size attenuation was comparable to that reported here. By the 36-month MTA follow-up, the effect was completely lost.¹⁵

The question arises whether the contrast of such follow-up results with the original positive study results exemplify the concern raised by Prasad & Ionnadis in a *JAMA* Viewpoint.^{16,17} They note that an “evidence-based” standard of care often flows from a single trial, becomes established, and is difficult to modify when later well-controlled replication attempts demonstrate more harm than good (e.g. more cost, risk, and discomfort than benefit). In the case of parent training added to antipsychotic agent, is the original study^{1,15} an evidence base for a standard of care that merely imposes more cost and nuisance for no long-term benefit? Should the follow-up results be taken to negate the positive results previously reported? We do not think so, for the following reasons: 1. The follow-up is not a controlled study, certainly not an attempted replication.. 2. The children and families benefited during the treatment phase of the study and even show a residual tendency of benefit a year later; the follow-up does not negate that. 3. We don’t know what would have happened if study treatment were continued during the follow-up.

Another similarity to the MTA is the tendency to continue the treatment received in the study: 86% of those assigned to medication in the MTA continued medication during at least part of the 10-month follow-up¹⁴. This was in contrast to only 44% taking medication of

those originally assigned to behavioral treatment without medication, even though for most of them the end-of-study recommendations included a trial of medication. The MTA group reached the same conclusion we do, that families tend to continue what is familiar, and this appears to be the biggest influence in treatment choice. Whether this phenomenon is good, bad, or indifferent cannot be answered by the results of this follow-up.

The main study limitation is the 70% retention rate, which could have biased the sample. In fact, those retained were more likely to have autistic disorder or Asperger's rather than PDD-NOS, to have been taking risperidone at the end of the main study, and to have had more positive items on the HSQ, suggesting slightly more baseline problems. However, diagnosis did not moderate treatment response¹⁸; there is no reason to suspect that the higher percent taking risperidone at study end would affect the difference between COMB and MED; and the slightly higher rate of baseline problems should, if anything, bias in the opposite direction of the slightly lower effect size found. In any event, the results are at least representative of 70% of those initially enrolled, and the attrition should not vitiate the treatment-group comparisons other than reducing power for statistical tests.

The average of 11 sessions of parent training in the original study added a significant increment of improvement to FDA-approved medication for irritability/noncompliance in autism while decreasing the optimal dose required. However, that benefit attenuated by more than half over the following year, as did the medication dose difference. This suggests a need for continued monitoring, troubleshooting, and periodic recalibration of parents' skills to maintain the gains. The fact that so few of the original MED-only group obtained the recommended PT on their own suggests that clinicians need to take an active role in arranging psychosocial treatment for families.

Acknowledgments

The original study was supported by National Institute of Mental Health through the following RUPP grants: Ohio State University, U10MH66768 (Michael G. Aman, PI); Indiana University, U10MH66766 (Christopher J. McDougle, PI); and Yale, U10MH66764 (Laurence Scahill, PI). It was further supported by Award UL1RR025755 from the National Center For Research Resources. This follow-up study was supported by a grant from Autism Speaks to L. Eugene Arnold. The content is solely the responsibility of the authors and does not necessarily represent official views of the National Institutes of Health, National Center for Research Resources, or Autism Speaks. We gratefully acknowledge Ms. Madison Spychalski's efficient assistance in manuscript formatting and editing of details.

References

1. Aman MG, McDougle CJ, Scahill L, et al. Research Units on Pediatric Psychopharmacology Autism Network. Medication and Parent Training in Children with Pervasive Developmental Disorders and Serious Behavior Problems: Results from a Randomized Clinical Trial. *J Am Acad Child & Adolesc Psychiatry*. 2009; 48:1143–1154. [PubMed: 19858761]
2. Johnson CR, Handen BL, Butter E, et al. Development of a parent training program for children with pervasive developmental disorders. *Behavioral Interventions*. 2007; 22:201–221.
3. Autism Network RUPP. Parent training for children with pervasive developmental disorders: a multisite feasibility trial. *Behavioral Interventions*. 2007; 22:179–199.
4. Barkley, RA.; Edwards, GH.; Robin, AL. *Defiant Teens: A Clinician's Manual for Assessment and Intervention*. New York: Guilford Press; 1999.
5. Chowdhury M, Aman MG, Scahill L, et al. The Home Situations Questionnaire–PDD Version: Factor structure and Psychometric Properties. *Journal of Intellectual Disabilities Research*. 2010; 54:281–291.
6. Aman MG, Singh NN, Stewart AW, Field CJ. The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. *Am J Ment Defic*. 1985a; 89:485–491. [PubMed: 3993694]

7. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951; 16:297–334.
8. Aman MG, Singh NN, Stewart AW, Field CJ. The psychometric characteristics of the Aberrant Behavior Checklist. *Am J Ment Defic*. 1985b; 89:492–502. [PubMed: 3158201]
9. Aman MG, Singh NN, Turbott SH. Reliability of the Aberrant Behavior Checklist and the effect of variations in instructions. *Am J Ment Defic*. 1987; 92:237–240. [PubMed: 3434595]
10. Carter AS, Volkmar FR, Sparrow SS, et al. The Vineland Adaptive Behavior Scales: supplementary norms for individuals with autism. *J Autism Dev Disord*. 1998; 28:287–302. [PubMed: 9711485]
11. Arnold, LE. Political zeal as theological science. Review of *Psychotherapy as Religion: the Civil Divine in America*. In: Epstein, William M., editor. *J Child & Adolesc Psychiatry*. Vol. 46. 2007. p. 1-4.
12. Scahill L, McDougle CJ, Aman MG, Johnson C, Handen B, Bearss K, Dziura J, Butter E, Swiezy NB, Arnold LE, Stigler KA, Sukhodolsky DG, Lecavalier L, Pozdol SL, Nikolov R, Hollway JA, Korzekwa P, Gavaletz A, Kohn AE, Koenig K, Grinnon S, Mulick JA, Yu S, Vitiello B. for the Research Units on Pediatric Psychopharmacology Autism Network. Effects of risperidone and parent training on adaptive functioning in children with a pervasive developmental disorders and serious behavioral problems. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2012; 51(2):136–146. [PubMed: 22265360]
13. The MTA Cooperative Group. A 14-Month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*. 1999; 56:1073–1086. [PubMed: 10591283]
14. The MTA Cooperative Group. The NIMH MTA Follow-up: 24-month Outcomes of Treatment Strategies for Attention-Deficit/Hyperactivity Disorder (ADHD). *Pediatrics*. 2004; 113:754–761. [PubMed: 15060224]
15. Jensen PS, Arnold LE, Swanson J, Vitiello B, Abikoff HB, Greenhill LL, Hechtman L, Hinshaw SP, Pelham WE, Wells KC, Conners CK, Elliott GR, Epstein J, Hoza B, Molina BSG, Newcorn JH, Severe JB, Wigal T, Gibbons RD, Hur K. Follow-up of the NIMH MTA study at 36 months after randomization. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2007; 46(8):988–1001.
16. Prasad V, Cifu A, Ioannidis JPA. Reversals of Established medical practices: Evidence to abandon ship. *JAMA*. 2012; 307 (1):37–38. [PubMed: 22215160]
17. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005; 294(2):218–228. [PubMed: 16014596]
18. Farmer CA, Lecavalier L, Yu S, Arnold L, McDougle C, Scahill L, Handen B, Johnson C, Stigler K, Bearss K, Swiezy N, Aman MG. Predictors and moderators of response to behavioral treatment in children with autism spectrum disorders and serious behavioral problems. *Journal of Autism and Developmental Disorders*. 2012; 42:1037–1044. [PubMed: 21822762]

Fig. 1. Mean HSQ noncompliance score at baseline, end of study treatment (week 24) and at 1-year follow-up for the 87 with follow-up. Lower score is better. Separation at both end of study Tx (24 weeks) and 1-year follow-up was not significant at this subgroup size even though the end-of-study separation was significant for the whole sample.

Fig. 2.

Prediction and mediation of Home Situations Questionnaire (HSQ) follow-up change (change from 24 weeks to follow-up) by 24-week Vineland daily living skill changes (improvement from baseline to 24 weeks). MED = medication alone (individual points represented by o, slope by solid line); COMB = medication + parent training (individual points represented by +, slope by dotted line). The lower the HSQ, the better; the higher the Vineland, the better. For COMB, the better the change in daily living skills at 24 weeks, the better the FU outcome. For MED, the better the change in daily living skills at 24 weeks, the worse the FU outcome. Interaction $p=0.02$

Table 1

Comparisons between subjects with follow-up and subjects without follow-up.

N=124	No Follow-up n=37 (30%)	With Follow-up n=87 (70%)	p –value
Male	33 (89%)	72 (83%)	0.43
Combination treatment group	24 (65%)	51 (59%)	0.45
White	25 (68%)	68 (78%)	0.26
Site: IU/OSU/YALE	17/12/8 (46%/32%/22%)	24/37/26 (28%/42%/30%)	0.14
Diagnosis: Autistic Disorder/Asperger's Disorder/PDD-NOS	22/0/15 (59%/0%/41%)	59/8/20 (68%/9%/23%)	0.04
Age			
Mean (SD)	7.17 (2.72)	7.54 (2.33)	0.45
Median (range)	6.35 (4.01–13.80)	7.30 (4.07–13.46)	
Taking RIS at the end of RUPP main study:			
First row: taking at week 24	14 (38%)	66 (76%)	<0.0001
2nd row: taking at LOCF	21 (57%)	74 (85%)	0.001
Dose of RIS: Mean daily mg (sd) (range) at 24 weeks	2.00 (0.50–3.25)	2.13 (0.75–3.50)	0.30
HSQ Mean(SD) at Baseline:	3.94 (1.60)	4.38 (1.57)	0.87
HSQ Mean(SD) at Main Study End (24 weeks)	1.50 (1.61)	1.51 (1.31)	0.97
HSQ Yes at Baseline, out of 25 possible yeses	17.24 (4.49)	19.32 (3.94)	0.011
HSQ Yes at EOS, out of 25 possible yeses	9.89 (6.5)	11.07 (6.9)	0.38
ABC Irritability Subscale at baseline	29.7 (7.4)	29.38 (6.30)	0.81
ABC Irritability Subscale at main study end	12.84 (9.18)	12.95 (7.96)	0.94
ABC Lethargy Subscale at baseline	15.65 (8.23)	16.07 (9.05)	0.8085
ABC Lethargy Subscale at Main Study End	5.51 (6.79)	5.30 (6.05)	0.862
ABC Stereotypy Subscale at baseline	8.62 (5.70)	8.82 (5.41)	0.8573
ABC Stereotypy Subscale at Main Study End	4.78 (5.13)	4.40 (4.89)	0.6958
ABC Hyperactivity Subscale 4 at baseline	36.97 (9.02)	35.07 (8.10)	0.2495
ABC Hyperactivity Subscale 4 at EOS	17.92 (12.47)	18.05 (10.82)	0.9546
ABC Inappropriate Speech Subscale at baseline	6.41 (3.72)	5.82 (3.68)	0.4164
ABC Inappropriate Speech Subscale at EOS	3.54 (3.47)	2.69 (2.99)	0.142

Note: IU = Indiana University; OSU = Ohio State University; PDD-NOS = pervasive developmental disorder not otherwise specified; RIS = Risperidone; RUPP = Research Units on Pediatric psychopharmacology (main study, ending at 24 weeks; LOCF = last observation carried forward

from week 16 or later; HSQ = Home Situations Questionnaire, developmental disabilities adaptation; ABC = Aberrant Behavior Checklist; BL = baseline; EOS = end of main study at 24 weeks.

Table 2

Comparison of the two randomized treatment groups on symptom scores at key time points and comparison of change in scores over time. Lower score is better on all scales.

Measure	Treatment Group	Baseline Mean (sd)	Mean Diff (sd) at Baseline p-value	End of RUPP Main Study (24Weeks) Mean (sd)	Mean Diff (sd) at 24 Weeks p-value	End of Follow-Up Mean (sd)	Mean Diff (sd) At Follow-up p-value	Change BL to 24wk Mean Diff (sd) p-value ES d	Change BL to FU Mean Diff (sd) p-value ES d
HSQ mean	MED	4.41 (1.51)	0.052 (1.58) p = 0.88	1.85 (1.32)	0.57 (1.28) p = 0.0431	2.12 (1.87)	0.28 (1.64) p = 0.43	0.52 (1.63) p = 0.14 d = 0.32	0.23 (1.86) p = 0.57 d = 0.12
	COMB	4.35 (1.63)		1.28 (1.26)		1.84 (1.46)			
HSQ Yeses	MED	19.67 (3.05)	0.59 (3.96) p = 0.47	12.00 (6.80)	1.59 (6.89) p = 0.29	13.67 (7.04)	0.98 (6.40) p = 0.48	1.00 (6.79) p = 0.50 d = 0.15	0.39 (6.85) p = 0.79 d = 0.06
	COMB	19.08 (4.48)		10.41 (6.96)		12.69(5.91)			
ABC I Irritability Subscale	MED	29.86 (6.40)	0.82 (6.32) p = 0.55	15.64 (8.47)	4.58 (7.68) p = 0.0075	15.25 (3.36)	1.15 (9.50) p = 0.58	3.76 (8.19) p = 0.038 d = 0.46	0.33 (9.93) p = 0.88 d = 0.03
	COMB	29.04 (6.26)		11.06 (7.07)		14.10 (3.60)			
ABC II Lethargy Subscale	MED	18.42 (8.20)	4.00 (8.88) p = 0.0413	6.81 (6.97)	2.57 (5.95) p = 0.0645	7.39 (6.83)	2.74 (5.93) p = 0.0365	-1.43 (7.17) p = 0.36 d = 0.20	-1.26 (8.00) p = 0.47 d = 0.16
	COMB	14.41 (9.33)		4.24 (5.11)		4.65 (5.21)			
ABC III Stereotypy Subscale	.MED	10.50 (5.44)	2.87 (5.25) p = 0.0139	6.58 (5.68)	3.72 (4.55) p = 0.0010	5.61(5.31)	1.55 (4.42) p = 0.14	0.85(3.96) -p = 0.33 d = 0.21	-1.32 (4.95) p = 0.22 d = 0.27
	COMB	7.63 (5.14)		2.86 (3.56)		4.06 (3.67)			
ABC IV Hyperactivity Subscale	MED	36.08 (6.36)	1.73 (8.10) p = 0.30	21.83 (10.96)	6.46 (10.39) p = 0.0054	18.94 (11.42)	1.57(11.63) p = 0.54	4.73(9.25) p = 0.02 d = 0.51	-0.16 (10.67) p = 0.95 d = 0.01
	COMB	34.35 (9.13)		15.37 (9.98)		17.37(11.78)			
ABC V Inappropriate Speech Subscal	MED	5.83(3.90)	0.03(3.69) p = 0.97	2.92 (3.44)	0.49 (3.00) p = 0.46	3.22(3.36)	-0.05 (3.03) p = 0.94	0.46 (3.14) p = 0.51 d = 0.15	-0.08 (3.24) p = 0.91 d = 0.02
	COMB	5.80(3.53)		2.43 (2.64)		3.27 (2.77)			

Note: ABC = Aberrant Behavior Checklist; MED = medication alone; COMB = medication + parent training in behavior management; Tx = treatment; sd = standard deviation; Diff = Difference; FU = follow up; BL = Baseline; wk = week; ES = effect size *d*.

Table 3

shows by randomized group and for the whole sample the number and percent who continued with the medication they left the study on, the number and percent who were taking risperidone at follow-up, number and percent taking antipsychotic, mean doses of risperidone for those still taking risperidone, and mean doses in risperidone equivalents for those who were taking another antipsychotic at follow-up, as well as percent using the methods learned in parent training.

	All patients with FU N=87	MED-only N=36	COMB N=51	p-value
# (%) of patients who continued with the medication taken at EOS	58 (67%)	23 (64%)	35 (69%)	0.65
# (%) who changed or added Medication	53 (61%)	19 (53%)	34 (67%)	0.25
# (%) who changed dose of study drug from what was taken at EOS	36 (41%)	19 (53%)	24 (47%)	0.67
# of patients who were taking Risperidone at follow-up	53 (61%)	19 (53%)	34 (67%)	0.25
Risperidone	2.13 (0.75–3.50)	2.50 (1.25–3.50)	2.00 (0.75, – 3.00)	0.01
At 24 weeks				
Mean dose (sd)	2.31 (0.80)	2.33 (1.03)	2.30 (0.81)	0.9
At FU				
# (%) taking antipsychotic other than Risperidone at FU	21 (24%)	9 (25%)	11 (22%)	0.9
Doses in RIS equivalents for those who were taking antipsychotics other than Risperidone at FU Mean (sd), median (range)	N=20 2.83 (1.34) 3.00 (0.50–4.5)	N=9 2.39 (0.90) 2.50 (1.25, – 4)	N=11 3.20 (1.56) 3.00 (0.50–6)	0.19
Seek help in learning behavior management during follow-up	65 (75%)	14 (39%)	51 (100%)	<0.0001
Currently using Behavior Management Principles	67 (77%)	19 (53%)	48 (94%)	<0.0001
# (%) using Behavior Management Principles (BMP)				
Some of the time	12 (13.8%)	3 (8.3)	9 (17.6%)	0.28
Much of the time	16 (18.4%)	4 (11.1%)	12 (23.5%)	0.22
Most of the time	18 (20.7%)	5 (13.9)	13 (25.5%)	0.28
At least 2 adults use BMP, and most of the time	21 (24.1%)	7 (19.4)	14 (27.5%)	0.50
# (%) who received treatments other than medication at follow-up	54 (62%)	25 (69%)	29 (57%)	0.33
# (%) of parents assigned to receive PT in Behavior Management (N=51) who found each session among the most helpful				
Behavioral Principles	N/A	N/A	19 (37%)	
Bedtime & Sleep Problems	N/A	N/A	3 (6%)	
Crisis Management	N/A	N/A	11 (22%)	
Contingency Contracting	N/A	N/A	2 (4%)	
Compliance Training	N/A	N/A	13 (25%)	
Daily Schedule	N/A	N/A	23 (45%)	
Feeding Disorders	N/A	N/A	1 (2%)	

	All patients with FU N=87	MED-only N=36	COMB N=51	p-value
Functional Communication Training	N/A	N/A	10 (20%)	
Generalization & Maintenance	N/A	N/A	6 (12%)	
Imitation Skills	N/A	N/A	1 (2%)	
Planned Ignoring	N/A	N/A	22 (43%)	
Prevention Strategies	N/A	N/A	13 (25%)	
Reinforcement	N/A	N/A	19 (37%)	
Teaching Skills	N/A	N/A	10 (20%)	
Time Out	N/A	N/A	12 (24%)	
Toilet Training	N/A	N/A	4 (8%)	

Note: FU = Follow up; MED = medication alone; COMB = medication + parent training (PT) in behavior management; EOS = end of study; RIS = Risperidone; n.s. = not significant.