JOURNAL OF CHILD AND ADOLESCENT PSYCHOPHARMACOLOGY
Volume 27, Number 2, 2017

Pp. 125–131
DOI: 10.1089/cap.2016.0107

Using a Patient-Centered Outcome Measure to Test Methylphenidate Versus Placebo in Children with Autism Spectrum Disorder

Lawrence Scahill, MSN, PhD,[1,2] Karen Bearss, PhD,[1,2] Rena Sarhangian, MD,[3,*] Christopher J. McDougle, MD,[4,5] L. Eugene Arnold, MD, MEd,[6] Michael G. Aman, PhD,[6] James T. McCracken, MD,[7] Elaine Tierney, MD,[8] Scott Gillespie, MS,[1,2] Valentina Postorino, PhD,[1,2] and Benedetto Vitiello, MD[9]

Abstract

Objectives: Parent rating scales are commonly used to evaluate change in clinical trials. Despite advantages, these measures may not capture parental impression of the child's most salient problems. We examine the use of parent target problems (PTPs) in a randomized trial of methylphenidate (MPH) in children with autism spectrum disorder and symptoms of attention-deficit/hyperactivity disorder.

Methods: This multisite, 4-week, randomized crossover trial compared three dose levels (low, medium, and high) of MPH with placebo. At baseline, the independent evaluator (IE) asked parents to nominate the child's two biggest problems. For each problem, the IE and parent coconstructed a brief narrative of the behavior and the impact on family life. The IE and parents reviewed and revised the narratives at subsequent visits. A panel of four judges, blind to treatment condition, independently reviewed the narratives to rate change from baseline on a 9-point scale: 1, normal; 2, markedly improved; 3, definitely improved; 4, equivocally improved; 5, no change; 6, possibly worse; 7, definitely worse; 8, markedly worse; 9, disastrously worse. The mean of the four raters was compared with primary and key secondary ratings from the original study.

Results: Two PTPs were recorded at baseline for 60 participants. The inter-rater reliability of the four judges across all PTPs and time points was excellent (intraclass correlation = 0.95). On the primary outcome measure (Aberrant Behavior Checklist Hyperactivity subscale), the medium and high-dose levels were superior to placebo. On the mean PTP rating, only the high dose was superior to placebo. We also compared PTP cutoff scores 3.0 (definitely improved), 3.25, and 3.5 with the rate of positive response on the Improvement item of the Clinical Global Impressions scale in the original study. Sensitivities ranged from 68% to 88%.

Conclusions: The parent target problem method offers a systematic way to identify and track patient-centered outcomes.

Keywords: patient-centered outcomes, methylphenidate, autism spectrum disorder, double-blind method

[1]School of Medicine, Emory University, Atlanta, Georgia.
[2]Marcus Autism Center, Atlanta, Georgia.
[3]David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California.
[4]Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts.
[5]Lurie Center for Autism, Boston, Massachusetts.
[6]Nisonger Center, Ohio State University, Columbus, Ohio.
[7]Division of Child Psychiatry, University of California at Los Angeles, Los Angeles, California.
[8]Kennedy Krieger Institute, Baltimore, Maryland.
[9]National Institute of Mental Health, Bethesda, Maryland.
*Division of Pediatric Hospital Medicine, Department of Pediatrics, Mattel Children's Hospital UCLA, University of California, Los Angeles (UCLA), Los Angeles, California.

Funding: This study was supported by funds under contracts N01MH80011 (Dr. Aman), N01MH70001 (Dr. McDougle), N01MH70010 (Dr. McCracken), and N01MH70009 (Dr. Scahill) from the National Institute of Mental Health, Bethesda, Md, by grants M01 RR00750 for Indiana University, Indianapolis, M01RR00052 for Johns Hopkins University, Baltimore, MD, M01 RR00034 for The Ohio State University, Columbus, and M01 RR06022 for Yale University, New Haven, Connecticut, from the General Clinical Research Centers, National Center for Research Resources, National Institutes of Health, Bethesda, by grants K23 MH068627 (Dr. Posey) and K24 MH001805 (Dr. McCracken) from the National Institute of Mental Health, and by the Korczak Foundation, Amsterdam, the Netherlands (Dr. Scahill). The study was also supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1 TR000454 (Emory University) and the Marcus Foundation.

Introduction

In the Fourth Edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), pervasive developmental disorders (PDDs) were defined by the early childhood onset of impaired social interaction, delayed communication, repetitive behavior, and restricted interests (American Psychiatric Association 2000). Detailed reviews of available epidemiological studies estimate the worldwide prevalence between 6.2 and 7.6 per 1000 children for PDDs and further estimate that 30% to 40% of affected children have intellectual disability (Elsabbagh et al. 2012; Hahler and Elsabbagh 2014). The three most common PDDs (autistic disorder, PDD-not otherwise specified (NOS), and Asperger's disorder) are now subsumed under the single diagnostic category of autism spectrum disorder (ASD) in *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) (American Psychiatric Association 2013).

Children with ASD may also have behavioral problems including tantrums, aggression and self-injury (Carroll et al. 2014), anxiety (Hallett et al. 2013), and hyperactivity with impulsiveness and distractibility (Lecavalier et al. 2006; Kaat et al. 2013). In contrast to DSM-IV, DSM-5 supports the concurrent diagnosis of attention-deficit/hyperactivity disorder (ADHD) in children with ASD (American Psychiatric Association 2000, 2013). Indeed, ADHD appears to be relatively common in children with ASD (Simonoff et al. 2008) and appears similar in clinical presentation when compared with children with ADHD without ASD (Gadow et al. 2005). Data from a U.S. national database indicate that 15% of children with ASD are being treated with stimulants and another 8% are using α_2 agonists presumably for the treatment of ADHD symptoms in many cases (Rosenberg et al. 2010).

To date, only a few studies have evaluated the efficacy and safety of standard medication treatments of ADHD in children with ASD in samples more than 60 subjects. These include immediate release methylphenidate (MPH; Research Units on Pediatric Psychopharmacology [RUPP] Autism Network 2005), extended-release guanfacine (Scahill et al. 2015), and two trials of atomoxetine (Harfterkamp et al. 2012; Handen et al. 2015). In general, these studies support the efficacy of these compounds in children with ASD and the triad of hyperactivity, distractibility, and impulsiveness. However, the positive response rate is lower than the reports in children with ADHD that is uncomplicated by ASD (Steele et al. 2006). Furthermore, the rate of adverse effects in children with ASD is predictably higher for each of these compounds.

These four trials used the parent-rated Hyperactivity subscale of the Aberrant Behavior Checklist (ABC) and/or DSM-IV-based scales such as the ADHD-Rating Scale or the Swanson Nolan and Pelham (SNAP) as primary or key secondary outcomes (Swanson 1992; Du Paul et al. 1998). These reliable and valid outcome measures have been used for many years and have stood the test of time. Among other advantages, these ratings have known distributions that can define a severity threshold for study entry and identify a meaningful change score (Scahill and Lord 2004). However, these standardized rating scales may not capture problems that are most salient to parents of children with ASD (Arnold et al. 2003). Even when showing improvement with treatment, standardized scales may fail to reflect change that is important to the family (Weisz et al. 2011).

In response to these potential disadvantages, other approaches have been proposed to capture patient-centered outcomes (Arnold et al. 2003; Hawley and Weisz 2003; Weisz et al. 2011; McGuire et al. 2014). The purpose of the current analysis is to examine the efficacy of MPH based on parent target problems (PTPs). The narratives for these PTPs were recorded during the course of the RUPP Autism Network MPH trial by independent evaluators (IEs) who were blind to treatment.

Methods

Design

This was a five-site, crossover trial of three doses of MPH and placebo [see RUPP Autism Network (2005) for details]. In brief, after a 1-week, open-label, test dose period, subjects who tolerated the study medication were assigned to receive placebo, low-, medium-, or high-dose MPH in random order (1 week each) under double-blind conditions. The dose levels of active MPH were $\sim$0.125 mg/kg (low dose), 0.25 mg/kg (medium dose), and 0.5 mg/kg (high dose) for the morning and noon doses, and half that dose at 4 p.m. (RUPP Autism Network 2005). For example, the medium dose level for a 20 kg child would be 5 mg in the morning, 5 mg at noon, and 2.5 mg at 4 p.m. Subjects who showed a positive response to at least one dose level of MPH were invited to participate in an 8-week, open-label, extension phase. In this report, we focus on the 4-week randomized, crossover phase.

There were two exceptions to the completely randomized design: (1) subjects who could not tolerate the highest dose level of MPH in the test-dose phase received the medium dose for 2 weeks during the double-blind, crossover phase and (2) the high dose could not follow the placebo, to avoid the potential adverse effects caused by the abrupt dose change. Each participant was followed by two blinded clinicians: a treating clinician who monitored dose and adverse effects and an IE who asked about behavioral problems and coconstructed the PTPs with the parent, but did not discuss dose or adverse effects.

Setting and Subjects

The study was approved by the institutional review board at each site (University of California at Los Angeles, Indiana University, Kennedy Krieger Institute at Johns Hopkins University, Ohio State University, and Yale University). Study subjects were recruited from clinic registries, referrals to these clinical programs, and outreach to parent support groups. The consent process described study purpose, duration, and expectations. No study data were collected until a parent (or legal guardian) signed the informed consent. Few children were able to provide assent. Protocol adherence was monitored through weekly teleconferences with the principal investigators and coordinators. The coordinating center (Yale University) conducted semiannual site visits. Study data were managed by the Nathan Klin Institute. Subject safety, enrollment, and attrition were monitored by an external monitoring board.

An experienced clinical team at each site conducted the screening, baseline, and follow-up assessments. To be eligible for the study, subjects had to be between 5 and 14 years of age, healthy, medication-free, and have a DSM-IV diagnosis of autistic disorder, Asperger's disorder, or PDD-NOS based on clinical assessment and corroborated by the Autism Diagnostic Interview-Revised (ADI-R; Lord et al. 1994). All subjects had moderate or greater symptoms of ADHD (based on PTPs, ABC Hyperactivity subscale, and the SNAP rating scale). The pretreatment assessment also included routine laboratory tests (blood counts, electrolytes, liver function tests, blood urea nitrogen, creatinine, and urinalysis), medical and psychiatric histories, a physical examination, and vital signs. The subjects were further characterized with the Slosson IQ test (Slosson 1983) and the Vineland Adaptive Behavior Scales (Sparrow et al. 1984).

Measures used in this report

Aberrant Behavior Checklist. The primary outcome measure was the parent-rated ABC Hyperactivity subscale collected at baseline, and after weeks 1, 2, 3, and 4 (Aman et al. 1985). The full ABC comprises 58 items with five subscales derived from factor analysis: Irritability, Social Withdrawal, Stereotypy, Hyperactivity, and Inappropriate Speech. All items are scored from 0 to 3, with higher scores indicating greater severity. The 16-item Hyperactivity subscale includes over activity (7 items), impulsiveness (2 items), inattention (3 items), and noncompliance (4 items).

Clinical Global Impressions scales. *The Clinical Global Impressions-severity* is a 7-point scale from nonsymptomatic (score of 1) to extreme (score of 7). A score of 4 (moderate) is a commonly used benchmark for symptom severity in need of intervention (Guy 1976). *The CGI-Improvement (CGI-I)* is also a 7-point scale designed to measure overall symptomatic change compared with baseline. Scores on the CGI-I range from 1 (very much improved) through 4 (unchanged) to 7 (very much worse). By convention, scores of 2 (much improved) or 1 (very much improved) are used to define positive response to treatment. The CGI-I scale was rated at each postbaseline visit by a blinded IE using all available information, but did not include discussion of adverse events. Adverse events were monitored separately by a blinded treating clinician.

Parent target problem. At baseline, the IE asked parents to nominate the child's two biggest problems. The target problems could be directly related to the focus of the trial (e.g., hyperactivity, distractibility, and impulsiveness), closely related (e.g., noncompliance), or unrelated (e.g., repetitive behavior). For each problem, the IE and parent coconstructed a label (e.g., *hyperactivity*), a brief description of the behavior and the impact on family life. For episodic behaviors (e.g., *meltdowns*), the interviewer obtained an estimate of the frequency, duration, and intensity of the behavior. For more pervasive problems such as hyperactivity, the interviewer established the constancy of the problem (e.g., half of the time, most of the time, and all the time) and documented an emblematic example of the behavior. Because the aim of the interview was to develop a description of the behavior, parents were not asked to judge the severity of the problem (e.g., mild, moderate, or severe; on a scale of 10, etc.) or to speculate on the explanation of the behavior. Examples of target problems:

Problem No. 1. ''*Hyperactivity:* 100% of the time, paces, jumps off furniture, can't sit still for more than 2–3 minutes. Often tries to leave the house or bolts in public places—needs close supervision all the time. Home furniture takes a beating, door has to be locked from the inside, family avoids taking him to public places.''

Problem No. 2. ''*Won't listen, won't follow basic rules:* Near constant, uses the phone without permission, rummages in grandmother's room, and plays with her makeup. When called to come in—runs the other way. Family in turmoil—grandmother now says she can't care for him. Parents are running out of child care options.''

Each week, the blinded IE asked parents to provide an update of the same two problems. This discussion included a review of the baseline narrative. Consistent with the approach at baseline, parents were not asked to judge whether the problem was better or worse—but to describe the current behavior. After this review, a new narrative was created. For example:

Problem No. 1. ''*Hyperactivity:* About half the time. Climbs on furniture—nothing broken lately. Has managed two or three successful trips to the store without bolting. Able to sit through dinner on two occasions this week. Requires close supervision.''

Problem No. 2. ''*Won't listen, won't follow basic rules:* About half the time disobeys and defies adults. Able to entertain himself on the computer—staying out of grandmother's room. Family in turmoil—but manageable.''

Parent target problem rating. After completion of the trial, a panel of four judges (B.V., L.E.A., C.J.Mc.D., and L.S.) rated change from baseline on the following 9-point scale: 1, normal; 2, markedly improved; 3, definitely improved; 4, equivocally improved; 5, no change; 6, possibly worse; 7, definitely worse; 8, markedly worse; and 9, disastrously worse (e.g., had to be hospitalized). To prepare for the independent rating for all subjects, the panel reviewed four baseline PTPs and rated four postbaseline narratives. These scores were reviewed on a teleconference to develop consensus on the approach for each rating. Once the common approach was established, all subsequent time points for all cases were rated independently and blind to treatment condition. In addition to the change rating, judges also rated the quality of each narrative on a 6-point scale (1, excellent description through 6, cannot be rated). If two judges gave a score of ''6''—that narrative was dropped. If only one judge indicated ''cannot be rated''—that panel member was asked to supply a score based on the limited information.

Analysis

Baseline characteristics of the study sample were examined using means for continuous variables and counts for categorical variables. The inter-rater reliability of the PTP ratings was estimated by intraclass correlation (ICC) across the panel of four judges. The ICC for the four raters was excellent at 0.95. Thus, PTP scores were averaged across the four raters and the two target problems yielding one mean PTP score for each subject at each dose. Given the crossover design, mean PTP scores for each dose level of MPH were compared with the mean scores for placebo using an intent-to-treat mixed effects linear model. The fixed effects were dose (four levels), site (five levels), and the site by dose interaction. The random effects were the intercept and the slope of the regression line for each dose of MPH. Site by dose interactions were not detected. Thus, only main effects were considered in the models.

We used a similar mixed model to analyze results for the subgroup of subjects with one or two ADHD-related PTPs (e.g., inattention, hyperactivity, and impulsiveness). For subjects with two ADHD-related PTPs, the mean of the two PTP ratings was averaged for each dose. For subjects with one ADHD-related PTP, only the mean of that PTP rating was used for each dose.

Exploratory analyses compared the sensitivity and specificity of various cut points on the mean PTP score with the ''gold standard'' classification for positive response: a rating of *Much Improved* or V*ery Much Improved* on the CGI-I in the original study. These analyses were conducted on the sample of 58 subjects with at least one ADHD-related target problem. This analysis began with a mean PTP score of ≤3.0 (Definite Improvement) and then iteratively ≤3.25, ≤3.5.

Results

Seventy-two subjects were enrolled in the test dose phase. Of these, 6 subjects exited because of intolerable adverse effects, 66 subjects were randomized, 65 actually entered the crossover trial (RUPP Autism Network 2005). Seven subjects withdrew during the crossover phase because of intolerable adverse effects. However,

four of these subjects had at least one dose of active medication and placebo and were included in the analysis to the extent of their involvement. Two additional subjects were excluded because two panel members reported that PTPs *could not be rated*. Thus, the total sample at baseline included 60 subjects (Table 1). By design, the 16 subjects who did not tolerate the high dose in the test dose phase received two medium doses during the crossover phase. The average of the two medium dose weeks was used in the analysis; the high dose was considered missing for these subjects. Thus, the sample size varies across pairwise comparisons.

Two clinicians (L.S. and K.B.) independently reviewed and classified the baseline PTPs into one of seven mutually exclusive categories. Disagreements were resolved by consensus. At baseline, 58 of 60 participants had at least one ADHD-related PTP. Of the 120 documented baseline PTPs, the most common was hyperactivity ($n=42$), followed by inattention ($n=28$) and impulsiveness ($n=20$). Far less common were defiance, aggression, repetitive behavior, and other (complaints that did not fit with in any other category).

Table 1. Baseline Characteristics of Subjects with Usable Parent Target Problem Ratings ($N=60$)

Variable	*N (%)*
Gender	
Male	53 (88.3)
Female	7 (11.7)
Race and ethnicity	
White/Non-Hispanic	44 (73.3)
Black or African American	9 (15)
Asian	4 (6.7)
Hispanic	3 (5)
Educational placement	
Regular school (public and private)	40 (66.7)
School for children with special needs	10 (17.2)
Prekindergarten	4 (6.9)
Special class with regular school	3 (5.2)
Not in school	1 (1.7)
Missing	2 (3.3)
PDD diagnosis	
Autistic disorder (1)	43 (71.7)
PDD-NOS (2)	12 (20)
Asperger's disorder (3)	5 (8.3)
IQ category	
≥70	24 (44.4)
<70	30 (55.6)
Missing	6 (10)
	Mean (SD)
Age (years)	7.4 (2.1)
Aberrant behavior checklist	
Irritability	16.4 (10.3)
Social withdrawal	12.0 (8.7)
Stereotypic behavior	7.4 (5.9)
Hyperactivity/noncompliance	33.2 (8.9)
Inappropriate speech	5.9 (4.1)
Vineland standard score	
Communication	63.8 (22.2)
Daily living skills	55.4 (19.8)
Socialization	62.0 (16.1)
Adaptive behavior composite	55.8 (22.9)

IQ, intellectual quotient; PDD, pervasive developmental disorder; PDD-NOS, pervasive developmental disorder-not otherwise specified.

Change across all parent target problem ratings

The main effects for the mixed effects linear model across the four doses were $F=2.58$, $p=0.06$ for the PTP scores and $F=6.62$, $p=0.0001$ for the ABC Hyperactivity subscale. Given the crossover design, we proceeded with pairwise comparisons for each dose versus placebo (Table 2). Pairwise comparisons on the ABC Hyperactivity subscale showed significant improvement for the medium and high doses of MPH compared with placebo. [Note: these results on the ABC differ slightly from the results in the original article (RUPP Autism Network 2005). In that report, all three active doses were superior to placebo. These slight differences are likely because of missing or unusable PTPs, resulting in a slightly lower sample size.]

On the PTP ratings, the effect sizes for the PTP ratings were smaller than the parent-rated ABC Hyperactivity subscale for the low and medium doses, but similar for the high dose. Only the high dose was significantly superior to placebo on the mean PTP scores. As sensitivity analyses, pairwise comparisons were repeated to evaluate possible effects of age, sex, and maternal education. These covariates were included in the model one at a time and had no effect on the results.

When the analysis was repeated including only the subjects with at least one ADHD-related PTP, there was a statistically significant main effect ($F=7.41$, $p=0.04$ for PTP scores and $F=5.81$, $p=0.002$ for ABC Hyperactivity subscale). In subsequent pairwise PTP comparisons, the low dose was not significant, the medium dose had a p value of 0.15, and the high dose was significantly superior to placebo ($p<0.01$). As shown in Table 3, the parent-rated ABC Hyperactivity subscale scores were statistically superior to placebo for the medium and high doses in this subgroup with ADHD-related PTPs. Effect sizes on the PTPs were smaller than the low and medium dose levels, but higher than the ABC Hyperactivity subscale for the high dose. Here again, pairwise comparisons adjusting for age, sex, and maternal education had no effect on the results.

Sensitivity and specificity

In the original study, 44 of 58 subjects had at least 1 week with a positive response on the CGI-I. Table 4 shows the sensitivity, specificity, and positive predictive value for mean cutoff scores (≤3.0, ≤3.25, and ≤3.5) on ADHD-related PTPs compared with the rate of positive response on the CGI-I (gold standard). The cutoff score of ≤3.0 (Definite Improvement) has a sensitivity of 68.2% (30 of 44) and a specificity of 76.8%. As shown in Table 4, as the cutoff score moves to ≤3.25 and ≤3.5, sensitivity improves; specificity does not change and positive predictive value increases slightly.

Discussion

This study used a patient-centered outcome measure to evaluate the efficacy of MPH in children with ASD accompanied by any combination of hyperactivity, impulsiveness, and distractibility. The study compared three dose levels of MPH with placebo in a 4-week crossover trial (RUPP Autism Network 2005). At baseline, an IE at each site asked parents to nominate the two most pressing problems for each child participant. These parent target problems (PTPs) were documented in brief narratives that described each problem in behavioral terms. The IEs, who were blind to treatment condition and did not discuss adverse effects, reviewed and revised the PTP narratives with parents weekly during the 4-week crossover trial. When the trial was completed, a panel of four judges, who were also blind to treatment condition, rated change from

Table 2. Pairwise Contrasts of Mean Scores Across All Parent Target Problem Panel Ratings and Parent-Rated Aberrant Behavior Checklist Hyperactivity Subscale After 1 Week on Each Active Dose Versus 1 Week on Placebo in the 4-Week Crossover Trial

	PTP rating[a]	*ABC hyperactivity*	p		*Effect size*[b]	
Dose	*Mean (SE)*	*Mean (SE)*[c]	*PTP*	*ABC*	*PTP*	*ABC*
Placebo ($n=59$)	4.37 (0.13)	25.97 (1.43)				
Low ($n=60$)	4.29 (0.14)	23.35 (1.51)	0.63	0.14	0.09	0.26
Medium ($n=56$)[d]	4.12 (0.13)	21.22 (1.41)	0.11	<0.01	0.27	0.46
High ($n=43$)[e]	3.99 (0.16)	21.44 (1.49)	0.04	0.01	0.40	0.44

[a]Means are averaged over two PTPs, unless a subject had only one PTP.
[b]Effect sizes calculated by subtracting model-based means and dividing by pooled SD for placebo; Placebo pooled SD for PTP = 0.94; Placebo pooled SD for ABC = 10.27.
[c]ABC Hyperactivity subscale scores are not exactly as reported in RUPP Autism Network (2005) because of slight differences in sample size.
[d]Medium dose scores for subjects randomized to receive medium dose twice were averaged.
[e]Change in sample size because of missing data and modifications in random assignment to dose (see text).
ABC, Aberrant Behavior Checklist; PTPs, parent target problems; SE, standard error.

baseline on a 9-point scale (no symptoms, 1 through no change, 5 to disastrously worse, 9) based only on the PTP narratives. The ratings by the panel of judges were highly reliable (ICC = 0.95).

When all available PTP ratings were considered, only the high dose of MPH was significantly better than placebo. By contrast, on the parent-rated ABC Hyperactivity subscale, the medium dose (0.25 mg/kg for the morning dose) and the high dose (0.5 mg/kg for the morning dose) were superior to placebo. Fifty-eight of 60 randomized subjects with available PTP data had one or more ADHD-related target problems. Here again, only the high dose was superior to placebo on the mean PTP ratings. And, once again, both the medium and high doses showed significant improvement on the ABC Hyperactivity subscale compared with placebo.

On the parent-rated ABC Hyperactivity subscale, the medium dose was statistically significant, but the effect sizes were small to medium (Tables 2 and 3). Although highly reliable, the ratings of the more narrowly defined PTPs may lack the precision to detect small to medium effect sizes in a brief four-level crossover trial. By contrast, in our placebo-controlled trial of risperidone, there was remarkable convergence between the mean PTP ratings and original ratings on the parent-rated ABC Irritability subscale (Arnold et al. 2003). The risperidone trial used a parallel design and the effect size was large.

Using the CGI-I rating of Much Improved or Very Much Improved by the IE in the original study as the gold standard, PTP mean scores of ≤3.0 (Definitely Improved), ≤3.25, and ≤3.5 showed 68.2%, 81.8%, and 88.6% sensitivity, respectively. The positive predictive values were more than 90% for all three cutoffs. As expected, incrementally more permissive cutoffs on the mean PTP score resulted in higher sensitivity. However, there was no change in specificity across these cutoffs.

Visual inspection of the data showed the same number of true negatives and false positive cases at each threshold. These results on the CGI-I also suggest a limit on the precision of the PTP measure in a multiple-dose crossover study. Indeed, there are potentially important differences between the CGI-I and PTP ratings. The CGI-I uses all available information to make a *global* judgment. By definition, PTPs are focused on two parent-nominated problems. Thus, the lack of incremental change in specificity could reflect the differences in the patient-centered outcome versus a clinician-rated global measure.

Several limitations warrant consideration when interpreting the results of this study. First, there were limitations in the original design. The study did not use a fully randomized design. For example, the random assignment was adjusted such that the high dose would not follow placebo. Also the number of subjects who received the high dose was smaller than the low and medium doses based on tolerability during the test dose period. These design decisions were driven by safety concerns and were appropriate given the vulnerability of children with ASD to stimulant-induced

Table 3. Pairwise Contrasts of Scores on Attention-Deficit/Hyperactivity Disorder-Related Parent Target Problem Panel Ratings and Parent-Rated Aberrant Behavior Checklist Hyperactivity Subscale After 1 Week on Each Active Dose Versus 1 Week on Placebo in the 4-Week Crossover Trial

	PTP rating	*ABC hyperactivity*	p		*ES*[a]	
Dose	*Mean*[b] *(SE)*	*Mean (SE)*	*PTP*	*ABC-H*	*PTP*	*ABC-H*
Placebo ($n=57$)	4.36 (0.15)	25.37 (1.47)				
Low ($n=58$)	4.28 (0.15)	23.38 (1.58)	0.63	0.26	0.08	0.20
Medium ($n=54$)[c]	4.11 (0.14)	21.31 (1.46)	0.12	0.01	0.24	0.40
High ($n=41$)[d]	3.92 (0.18)	20.73 (1.48)	0.02	0.01	0.42	0.46

[a]Effect sizes calculated by subtracting model-based means and dividing by pooled SD for placebo; Placebo pooled SD for PTP = 1.06; Placebo pooled SD for ABC = 10.07.
[b]Means are averaged over two ADHD-related PTPs, unless a subject had only one PTP.
[c]Medium dose scores for subjects randomized to receive medium dose twice were averaged.
[d]Change in sample size because of missing data and modifications in random assignment to dose (see text).
ABC, Aberrant Behavior Checklist; ABC-H, Aberrant Behavior Checklist Hyperactivity; ADHD, attention-deficit/hyperactivity disorder; PTPs, parent target problems; SE, standard error.

Table 4. Sensitivity, Specificity, and Positive Predictive Value at Different Cut Points on the Mean Attention-Deficit/Hyperactivity Disorder-Related PTP Score Compared with Rate of Positive Response on CGI-I in the Trial (N = 58)

PTP cutoff score	*CGI-I positive response*[a]	*CGI-I negative response*[b]	*Sensitivity (%)*	*Specificity (%)*	*PPV (%)*
≤3.5	39 (T+)	3 (F+)	88.6	78.6	92.9
>3.5	5 (F–)	11 (T–)			
≤3.25	36	3	81.8	78.6	92.3
>3.25	8	11			
≤3.0	30	3	68.2	78.6	90.9
>3.0	14	11			

Sensitivity = T+ ÷ (T+) + (F–); Specificity = T– ÷ (T–) + (F+); Positive predictive value = T+ ÷ (T+) + (F+).
[a]At least 1 week *Much Improved* or *Very Much Improved* on the CGI-I.
[b]All other scores on CGI-I.
CGI-I, Clinical Global Impressions-Improvement; PPV, positive predictive value; PTPs, parent target problems; SE, standard error.

adverse effects. Second, 1 week on each dose versus placebo may not have been sufficient to show the full effects of each active dose. As already suggested, the 1 week duration at each dose level posed a challenge of precision for the PTP ratings.

Conclusions

The Food and Drug Administration (2009) as well as other federal agencies has emphasized the importance of patient-reported outcomes. For children with ASD, we rely on parents. The parent target problem method offers a systematic way to identify and track patient-centered problems. In clinical trials, PTPs can provide complementary information to parent- and clinician-rated measures for IEs to determine overall improvement on the CGI-I in real time (Bearss et al. 2015; Scahill et al. 2015). Results of this study are consistent with previous studies that have shown that PTP ratings by a panel of judges are highly reliable (Arnold et al. 2003; McGuire et al. 2014). The method permits quantification and analysis of problems at the level of the individual.

Clinical Significance

In child mental health clinical practice, the initial evaluation invariably includes inquiry about the chief complaint, that is, the reason for seeking treatment. Done well, exploration of primary problems may promote therapeutic alliance and provide a sound basis for setting goals of treatment (Warnick et al. 2014). PTPs offer a semi-structured method of coconstructing narratives for the child's two most pressing clinical problems. The PTPs may be used in combination with rating scales to guide clinical decision making. In this study, for example, the results of the PTP ratings suggest that the lowest dose level of MPH (0.125 mg/kg for the morning dose) was moving in the right direction—but fell short of the ''definitely improved'' designation. This impression was consistent with the parent-rated ABC Hyperactivity subscale for the low dose. In a clinical setting, this picture would likely call for a dose increase. The PTP ratings were also consistent with the parent-rated Hyperactivity subscale scores for the high dose.

Results on the ABC Hyperactivity subscale and the mean PTP score, however, were not entirely consistent on the medium dose. This divergence in the PTP and ABC Hyperactivity subscale results for the medium dose could lead to different conclusions. For example, a prescriber in clinical practice may decide, based on the improvement in the parent-rated scale, to continue with the medium dose. Based on modest improvement on the PTP, the prescriber might decide to increase the dose. Given the evidence that higher doses of MPH are associated with a higher rate of adverse effects (RUPP Autism Network 2005), the clinician might also decide to wait before raising the dose—despite less than definite improvement on the PTP. In clinical settings where rating scales are not routinely used to track progress, the description of the PTPs in behavioral terms provides an individualized baseline that can be compared and contrasted with subsequent narratives. The semi-structured review of the child's behavior could be more informative than open-ended discussion alone.

Disclosures

Dr. Lawrence Scahill has served as a consultant for Neuren, Coronado, Roche, Supernus, MedAdvante, and Bracket. Dr. L. Eugene Arnold has received research funding from Curemark, Forest, Lilly, Neuropharm, Novartis, Noven, Shire, Supernus, and Young Living (as well as NIH and Autism Speaks), and has consulted with or been on advisory boards for Arbor, Gowlings, Ironshore, Neuropharm, Novartis, Noven, Organon, Otsuka, Pfizer, Roche, Seaside Therapeutics, Sigma Tau, Shire, Tris Pharma, and Waypoint, and received travel support from Noven. Dr. Michael G. Aman has received research contracts, consulted with, served on advisory boards, or done investigator training for CogState, Inc.; Confluence Pharmaceutica; CogState Clinical Trials, Ltd.; Coronado Biosciences; Forest Research; Hoffman-La Roche; Lumos Pharma, MedAvante, Inc.; Novartis; Pfizer; ProPhase LLC; and Supernus Pharmaceuticals. Dr. James T. McCracken reports receiving consultant income from Roche, Dart Neuroscience, Think Now, Inc., and Brightech International; he has received study drug from Shire and AstraZeneca. Dr. Benedetto Vitiello has received salary support from NIH and consultant fees from the American Physician Institute for Advanced Professional Studies. No other authors have potential conflicts of interest to disclose.

References

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 89:485–491, 1985.

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision. Washington DC: American Psychiatric Association; 2000.

American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Washington DC: American Psychiatric Association; 2013.

Arnold LE, Vitiello B, McDougle C, Scahill L, Shah B, Gonzalez NM, Chuang S, Davies M, Hollway J, Aman MG, Cronin P, Koenig

K, Kohn AE, McMachon DJ, Tierney E: Parent-defined target symptoms respond to risperidone in RUPP autism study: Customer approach to clinical trials. J Am Acad Child Adolescent Psychiatry 42:1443–1450, 2003.

Bearss K, Johnson C, Smith T, Lecavalier L, Swiezy N, Aman M, McAdam DB, Butter E, Stillitano C, Minshawi N, Sukhodolsky D, Mruzek DW, Turner K, Neal T, Hallett V, Mulick JA, Green B, Handen B, Deng Y, Dziura J, Scahill L: Effect of parent training versus parent education on behavioral problems in children with autism spectrum disorder: a randomized clinical trial. JAMA 313:1524–1533, 2015.

Carroll D, Hallett V, McDougle CJ, Aman MG, McCracken JT, Tierney E, Arnold LE, Sukhodolsky DG, Lecavalier L, Handen B, Swiezy N, Johnson C, Bearss K, Vitiello B, Scahill L: Examination of aggression and self-injury in children with autism spectrum disorders and serious behavioral problems. Child Adolesc Psychiatric Clinics N Am 23:57–72, 2014.

Du Paul GJ, Power TJ, Anastopoulos AD, Reid R: ADHD Rating Scale—IV: Checklists, Norms, and Clinical Interpretation. New York, NY, Guilford Press, 1998, pp. 1–79.

Elsabbagh M, Divan G, Koh Y-J, Kim YS, Kauchali S, Marcín C, Montiel-Nava C, Patel V, Paula CS, Wang C, Yasamy MT, Fombonne E: Global prevalence of autism and other pervasive developmental disorders. Autism Res 5:160–179, 2012.

Food and Drug Administration: Guidance for industry: Patient-reported outcome measures: use in medical product development to support labeling claims, 2009. Available at www.fda.gov/downloads/drugs/guidances/ucm193282.pdf accessed November 7, 2016.

Gadow KD, De Vincent C, Pomeroy J, Azizian A: Comparison of DSM-IV symptoms in elementary school-aged ADHD symptoms in children with PDD versus clinic and community samples. Autism 9:392–415, 2005.

Guy W: ECDEU Assessment Manual for Psychopharmacology. Washington (DC), US Department of Health, Education, and Welfare, 1976.

Hahler EM, Elsabbagh M: Autism: A global perspective. Curr Dev Disord Rep 2:58–64, 2014.

Hallett V, Lecavalier L, Sukhodolsky DG, Cipriano N, Aman MG, McCracken JT, McDougle CJ, Tierney E, King BH, Hollander E, Sikich L, Bregman J, Anagnostou E, Donnelly C, Katsovich L, Dukes K, Vitiello B, Gadow K, Scahill L: Exploring the manifestations of anxiety in children with autism spectrum disorders. J Autism Dev Disord 43:2341–2352, 2013.

Handen BL, Aman MG, Arnold LE, Hyman SL, Tumuluru MV, Lecavalier L, Corbett-Dick P, Pan X, Hollway JA, Buchan-Page KA, Silverman LB, Brown NV, Rice Jr., RR, Hellings J, Mruzek DW, McAuliffe-Bellin S, Hurt EA, Ryan MM, Levato L, Smith T: Atomoxetine, parent training, and their combination in children with autism spectrum and ADHD. J Am Acad Child Adolesc Psychiatry 54:905–915, 2015.

Harfterkamp M, van de Loo-Neus G, Minderaa RB, van der Gaag RJ, Escobar R, Schacht A, Pamulapati S, Buitelaar JK, Hoekstra PJ: A randomized double-blind study of atomoxetine versus placebo for attention-deficit/hyperactivity disorder symptoms in children with autism spectrum disorder. J Am Acad Child Adolesc Psychiatry 2012; 51:733–741.

Hawley KM, Weisz JR: Child, parent, and therapist (dis)agreement on target problems in outpatient therapy: The therapist's dilemma and its implications. Consult Clin Psychol 71:62–70, 2003.

Kaat AJ, Lecavalier L, Aman MG Validity of the aberrant behavior checklist in children with autism spectrum disorder. J Autism Dev Disord 44:1103–1116, 2014.

Lecavalier L: Behavioral and emotional problems in young people with pervasive developmental disorders: Relative prevalence, effects of subject characteristics, and empirical classification. J Autism Dev Disord 2006; 36:1101–1114.

McGuire JF, Sukhodolsky DG, Bearss K, Grantz H, Pachler M, Lombroso PJ, Scahill L: Individualized assessments in treatment research: An examination of parent-nominated target problems in the treatment of disruptive behaviors in youth with Tourette syndrome. Child Psychiatry Hum Dev 45:686–694, 2014.

Research Units on Pediatric Psychopharmacology Autism Network: Randomized, controlled, crossover trial of methylphenidate in pervasive developmental disorders with hyperactivity. Arch Gen Psychiatry 62:1266–1274, 2005.

Rosenberg RE, Mandell DS, Farmer JE, Law JK, Marvin AR, Law PA: Psychotropic medication use among children with autism spectrum disorders enrolled in a national registry, 2007–2008. J Autism Dev Disord 40:342–351, 2010.

Scahill L, Lord C: Subject selection and characterization in clinical trial in children with autism. CNS Spectr 9:22–32, 2004.

Scahill L, McCracken JT, King BH, Rockhill C, Shah B, Politte L, Sanders R, Minjarez M, Cowen J, Mullett J, Page C, Ward D, Deng Y, Loo S, Dziura J, McDougle CJ, and Research Units on Pediatric Psychopharmacology Autism Network: Extended-release guanfacine for hyperactivity in children with autism spectrum disorder. Am J Psychiatry 172:1197–1206, 2015.

Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G: Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. J Am Acad Child Adolesc Psychiatry 47:921–929, 2008.

Slosson RL: Slosson Intelligence Test. East Aurora, NY: Slosson Educational Publications, 1983.

Sparrow SS, Balla DA, Cicchetti DV: Vineland Adaptive Behavior scales: Survey form Manual. Circle Pines: American Guidance Service, 1984.

Steele M, Jensen PS, Quinn DM: Remission versus response as the goal of therapy in ADHD: A new standard for the field? Clin Ther 28:1892–1902, 2006.

Swanson JM: School-Based Assessments and Interventions for ADD Students. Irvine, CA: KC Publishing, 1992.

Warnick E, Bearss K, Weersing VR, Scahill L, Woolston J: Shifting the treatment model: Impact on engagement in outpatient therapy. Adm Policy Mental Health 41:93–103, 2014.

Weisz JR, Chorpita BF, Frye A, Ng MY, Lau N, Bearman SK, Ugueto AM, Langer DA, Hoagwood KE; Research Network on Youth Mental Health: Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. J Consult Clin Psychol 79:369–380, 2011.

Address correspondence to:
Lawrence Scahill, MSN, PhD
Marcus Autism Center
1920 Briarcliff Road
Atlanta, GA 30329

E-mail: lawrence.scahill@emory.edu