Citation for published version (APA):
A double-blind randomized placebo-controlled trial of citalopram adjunctive to stimulant medication in youth with chronic severe irritability

Kenneth Towbin, MD†, Pablo Vidal-Ribas, PhD†, Melissa A. Brotman, PhD, Andrew Pickles, PhD, Katherine V. Miller, BA, Ariela Kaiser, BS, Aria D. Vitale, BA, Chana Engel, MSN, PMHNP-BC, Gerald P. Overman, PharmD, Mollie Davis, MSW, LCSW-C, Beth Lee, MSN, RN-BC, Cheri McNeil, PsyD, Wanda Wheeler, MSW, Catherine H. Yokum, MSW, Catherine T. Haring, MD, Alexandra Roule, BA, Caroline G. Wambach, BS, Banafsheh Sharif-Askary, BS, Daniel S. Pine, MD, Ellen Leibenluft, MD, Argyris Stringaris, MD, PhD, FRCPsych.

† Contributed equally.

Dr. Towbin, Dr. Vidal-Ribas, Ms. Miller, Ms. Kaiser, Ms. Vitale, Ms. Engel, and Dr. Stringaris are with Mood Brain and Development Unit, Emotion and Development Branch, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA.

Dr. Vidal-Ribas is also with Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, UK.

Ms. Davis, Dr. McNeil, Ms. Wheeler, Ms. Yokum and Dr. Leibenluft are with Section on Mood Dysregulation and Neuroscience, Emotion and Development Branch, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA.

Dr. Brotman is with Neuroscience and Novel Therapeutics, Emotion and Development Branch, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA.

Prof. Pickles is with Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King’s College London, London, UK.

Dr. Overman is with Clinical Center Pharmacy Department, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA.

Ms. Lee is with University of Arizona, College of Nursing, Arizona, USA.
Dr. Haring is with Department of Otolaryngology- Head and Neck Surgery at the University of Michigan, Ann Arbor, Michigan, USA.

Ms. Roule is with Pennsylvania State University, State College, Pennsylvania, USA.

Ms. Wambach is with Georgetown University School of Medicine, Washington, DC, USA.

Ms. Sharif-Askary is with Duke University School of Medicine, Durham, North Carolina, USA.

Dr. Pine is with Section on Development and Affective Neuroscience, Emotion and Development Branch, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA.

**Corresponding author:**

Argyris Stringaris, MD, PhD, FRCpsych.
Mood, Brain and Development Unit
Emotion and Development Branch
National Institute of Mental Health
National Institutes of Health
9000 Rockville Pike, Building 15k, Room 206
Bethesda, MD, 20892
Phone: (301) 443-8019
E-mail: argyris.stringaris@nih.gov

**Acknowledgements:**

This study represents independent research funded by Intramural Research Program of the National Institute of Mental Health (NIH) with analysis supported by the UK National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London. The views expressed are those of the author(s) and not necessarily those of the UK NHS, NIHR or Department of Health and Social Care of United Kingdom. Authors have no financial or other conflicts of interest to report. The authors wish to acknowledge the skill, dedication and effort of the 1 SW Nursing Staff from the NIH.
Clinical Center, Bethesda, Maryland, without whom this study would not have been possible.
We gratefully acknowledge the children and families who participated in this trial and made this research possible.

**Keywords:**
Irritability, disruptive mood dysregulation disorder, citalopram, methylphenidate, RCT

**Facebook:**
A new study in #JAACAP uses a double-blind randomized controlled design to examine the effects of citalopram vs placebo in children and adolescents with severe #irritability unresponsive to stimulant medication. The authors find that severe irritability symptoms improved in significantly more participants who received add-on citalopram (35%) than those receiving add-on placebo (6%).

**Twitter:**
New study @JAACAP using double-blind randomized controlled design finds citalopram might be effective in reducing severe #irritability in children and adolescents unresponsive to stimulant medication @NIMHgov
A double-blind randomized placebo-controlled trial of citalopram adjunctive to stimulant medication in youth with chronic severe irritability

Abstract

Objective: Despite the clinical importance of chronic and severe irritability, there is a paucity of controlled trials for its pharmacological treatment. Here, we examine the effects of adding citalopram (CTP) to methylphenidate (MPH) in the treatment of chronic severe irritability in youth using a double-blind randomized placebo-controlled design.

Methods: After a lead-in phase of open treatment with stimulant, 53 youth meeting criteria for severe mood dysregulation (SMD) were randomly assigned to receive CTP or placebo (PBO) for 8 weeks. Forty-nine participants – 48 of them (98%) meeting disruptive mood dysregulation disorder (DMDD) criteria - were included in the intent-to-treat analysis. The primary outcome measure was the proportion of response based on improvements of irritability at the 8th week of the trial.

Results: At the end of the trial, a significantly higher proportion of response was seen in those participants randomly assigned to CTP+MPH compared to PBO+MPH (35% CTP+MPH vs. 6% PBO+MPH; OR=11.70, 95%CI 2.00, 68.16, p=0.006). However, there were no differences in functional impairment between groups at the end of the trial. No differences were found in any adverse effect between treatment groups, and no trial participant exhibited hypomanic or manic symptoms.
Conclusion: Adjunctive CTP might be efficacious in the treatment of chronic severe irritability in youth resistant to stimulant treatment alone. This trial was registered on ClinicalTrials.gov (Identifier: NCT00794040).
Introduction

Chronic irritability is amongst the most common reasons for referral to child psychiatric services\(^1,2\) and is associated with substantial concurrent and future impairment.\(^3-6\) As a result, the American Psychiatric Association recently introduced the diagnosis of Disruptive Mood Dysregulation Disorder (DMDD) into the DSM-5 to capture youth with severe, chronic irritability.\(^7\) Yet, there is a paucity of clinical trials for DMDD.\(^2,8\) In this paper, we present the results of a randomized controlled trial of citalopram (CTP) versus placebo (PBO) as add ons to open-label stimulant optimization for the treatment of irritability in severe mood dysregulation (SMD), which became the basis for DSM-5's DMDD.

Despite the increasing recognition of the importance of irritability, rigorous testing of possible treatments is still at a nascent state. In terms of psychological therapy, there is encouraging preliminary evidence of efficacy from a small number of trials,\(^9,10\) but further work is needed.\(^11-13\) Since pharmacological treatments have a clear role in most common psychiatric disorders in youth, including anxiety,\(^14\) depression,\(^15,16\) and ADHD,\(^17,18\) and may work synergistically with psychological treatments,\(^14\) it is also imperative to identify pharmacological agents for the treatment of children experiencing chronic irritability.\(^2\)

While prior literature suggested a therapeutic effect of lithium on aggression,\(^19,20\) a small trial (N=24)\(^21\) found lithium to be ineffective for the treatment of severe irritability in children and adolescents with SMD.\(^22\) There has been no RCT of antipsychotic medication specifically
targeting irritability in children other than those with autism spectrum disorder (ASD). However, an open trial employed risperidone in youth with SMD and showed a reduction in irritability.\textsuperscript{23}

Further evidence comes from treatment trials with stimulant in youth with ADHD. A post hoc analysis of the Multimodal Treatment Study of Children with ADHD\textsuperscript{24} suggested that stimulant treatment of irritability in those with ADHD may be effective. Waxmonsky and colleagues,\textsuperscript{25} in a retrospective analysis of data from a cross-over, placebo-controlled trial of methylphenidate (MPH), reported a significant reduction of irritability/aggression in ADHD. Recently, evidence from several open-label trials with stimulant in youth with ADHD and comorbid SMD/DMDD has emerged. Waxmonsky et al.\textsuperscript{26} found that stimulant combined with PMT and CBT might be efficacious to treat irritability in this population. The same group has also shown that stimulant alone might decrease behavioral and mood symptoms in youth with ADHD and comorbid DMDD.\textsuperscript{27} And a recent small open-label trial (N=22) has demonstrated reductions in irritability with medium to large effect sizes in these children following stimulant monotherapy.\textsuperscript{28} Nevertheless, a significant proportion (~50\%) of those with aggression and ADHD remain refractory to stimulant medication even when combined with parent training/behavioral treatments.\textsuperscript{25,29}

Focusing on the related construct of aggression, the Treatment of Severe Childhood Aggression (TOSCA) study group found add-on risperidone and placebo were no different to each other when combined with parent-training plus stimulant medication in children with ADHD;\textsuperscript{30} similar results were obtained for the 12-week follow up of this study.\textsuperscript{31} In terms of other mood-stabilizing medication, Blader and colleagues reported that add-on valproate reduced aggression
for a small (n=14) cohort of children with ADHD who did not respond to optimized stimulant plus family education.\textsuperscript{32}

Thus far, there has been no RCT of serotonin reuptake inhibitors (SRIs) specifically targeting chronic irritability. Indirect evidence derived from adult samples indicates that SRIs may be efficacious in the treatment of irritability in depressed adults,\textsuperscript{33} as well as patients with intermittent explosive disorder\textsuperscript{34} and those with premenstrual dysphoria.\textsuperscript{35} A review on the effect of antidepressants on irritability in young people\textsuperscript{36} identified two uncontrolled studies of SRIs that reported on irritability as an outcome;\textsuperscript{37,38} both indicated improvement of irritability with SRI treatment. In addition, a recent metaanalysis has shown that irritability may be a specific and robust predictor of future anxiety and depressive disorders.\textsuperscript{3} Moreover, findings from genetically-informative studies suggest shared pathophysiological mechanisms among irritability, anxiety and depression\textsuperscript{39,40}, thus indicating that agents effective for these disorders could also be useful for irritability. Taken together, these data make SRIs a promising candidate for the treatment of children with severe chronic irritability. Yet, this remains to be tested.

To address this gap in the literature, we conducted an RCT of CTP, an SRI, plus MPH against PBO plus MPH in youth who were originally recruited because they fulfilled criteria for SMD. Prior to randomization, children took part in an open-label stimulant optimization lead-in phase. This lead-in phase was motivated by the observation that a majority of youth with DMDD also suffer from ADHD\textsuperscript{41} and that stimulants appear efficacious for irritability in ADHD.\textsuperscript{24} Thus, our RCT was designed to provide a rigorous test of SRI effects over and above stimulant optimization in youth with severe irritability.
Methods

Subjects

This study was conducted at the National Institute of Mental Health Division of Intramural Research Programs (NIMH DIRP) from November 2008 until January 2018, and was approved by the National Institute of Health’s Combined Neuro-Science Institutional Review Board (CNS-IRB). Prior to participation, the study was explained to parents and patients. All children gave written assent and parents gave written informed consent. Subjects (ages 7–17 years) were recruited through advertisements that were placed in local parenting magazines, on support groups’ websites, and distributed to psychiatrists nationwide. Also, information about the study was provided in talks to local practitioners, and advocacy/parent groups. The study was registered on ClinicalTrials.gov (Identifier: NCT00794040).

Screening

All participants met criteria for SMD, which were designed to capture those youth with non-episodic irritability symptoms who were frequently diagnosed with bipolar disorder in clinical settings. Based on almost a decade of research on SMD, DMDD was introduced in DSM-5 after this study began. All but one randomized participant (n=48, 98%) in this study met criteria for DMDD; one participant did not meet DMDD criteria because onset of severe irritability was
after 10 years old, but before age 12. Inclusion criteria for the trial were: (1) irritability operationally defined as markedly increased reactivity to negative emotional stimuli manifest verbally or behaviorally at least three times weekly; (2) abnormal mood (anger or sadness), present at least half of the day most days; (3) three hyperarousal symptoms (insomnia, agitation, distractibility, racing thoughts or flight of ideas, pressured speech, intrusiveness); (4) symptoms cause severe impairment in at least one setting (home, school, or peers) and at least mild impairment in a second setting; (5) symptom onset before age 12 and currently present for at least 12 months without any symptom-free period greater than 2 months; (6) failing treatment as defined by current CGAS score <60, outpatient treatment provider agrees that the child’s response to treatment is no more than minimal 7) On the basis of record review and interviews with child and parent, the research team agrees that the child’s response to his/her current treatment is no more than minimal. Having a score of > 12 on the irritability subscale of the Aberrant Behavior Checklist (ABC) was initially an inclusion criteria but we stopped collecting the ABC for several reasons (see Supplement 1, available online). All the patients had histories of failed response to treatment either pharmacological, psychological or both. None of patients were medication naive at the time of enrollment. For all patients, their providers in the community endorsed their patient’s participation in the study and acknowledged that the current treatment was only minimally successful and other options should be explored.

Exclusion criteria for the trial were: (1) presence of cardinal bipolar symptoms of elevated, expansive mood, grandiosity, inflated self-esteem, or episodically decreased need for sleep; (2) distinct episodes of hypo/manic symptoms greater than 1 day; (3) current Major Depressive Disorder; (4) Autism Spectrum Disorder; (5) psychosis; (6) Post-Traumatic Stress Disorder; (7)
substance abuse within 3 months; (8) medical illnesses that require medications, are chronic, unstable, could cause SMD symptoms or are contraindications to treatment with SRI or stimulant (e.g. liver, seizure, renal, platelet disorder), or require medications that are contraindicated with SRIs or MPH; (9) intelligence quotient (IQ) <70; (10) pregnant, lactating, or sexually active without barrier method of contraception; (11) failed adequate trial (defined as four weeks of consecutive treatment with no less than 20 mg citalopram or 10 mg escitalopram) or severe ill effects while on citalopram or escitalopram; (12) history of hypersensitivity or severe adverse reaction to methylphenidate or serious adverse reactions (psychosis, severely increased activation compared to baseline) to methylphenidate or amphetamines.

Co-morbid ADHD, ODD, or anxiety disorders were not exclusionary. Full-scale intelligence quotient (FSIQ) was measured using the Wechsler Abbreviated Scale of Intelligence (WASI) for all subjects.43

Following a telephone interview to screen for relevant inclusion or exclusion criteria, record review, and consultation with the child’s treating clinician, candidates for the study were invited to the NIMH IRP (n= 311). On-site screening included the Kiddie Schedule for Affective Disorders Present and Lifetime Version (K-SADS-PL)44 with an additional supplement for SMD, designed in collaboration with Joan Kaufman, Ph.D., to ascertain whether children met criteria for this syndrome. All diagnostic measures were administered to the parent and child individually by trained post-graduate level clinicians with established interrater reliability (k=0.9, including distinguishing SMD subjects from those with distinct hypo/manic episodes).22,44
Diagnoses were based on best-estimate procedures\textsuperscript{45} generated in a consensus conference led by two psychiatrists with extensive experience evaluating children with mood disorders.

\textit{Enrollment}

See \textbf{Figure S1} for trial design overview and \textbf{Figure 1} for the CONSORT diagram. After on-site screening, subjects who met inclusion criteria were offered enrollment in the RCT. To maximize safety and observational data collection while they were tapered off medication and began open treatment with MPH, patients received inpatient care on a research child psychiatric unit at the NIH Clinical Center. Upon admission to the hospital, participants were assessed as described below (\textit{Baseline at admission}) and then gradually tapered off psychotropic medications (\textit{Phase I}). They were monitored closely for clinical deterioration and side effects; weekend passes with family were offered weekly or every other weekend depending on proximity to NIH. Medication tapering (Median=3 weeks) was individually tailored and prioritized medications with longer half-lives (e.g., atypical neuroleptics, antidepressants) before those with short half-lives (e.g., psychostimulants). Care in the hospital for all patients included weekday on-site school, weekday rounds meeting with a psychiatrist or nurse practitioner, weekly meetings with parent(s), individual and group activities, and milieu treatment that are common in inpatient pediatric psychiatry units. Concomitant therapy and any other psychological treatment were not allowed during participation in the protocol. During medication taper, 12 participants were discharged for meeting exclusion criteria (see Supplement 1, available online) (\textbf{Figure 1}).

\textit{Treatment of ADHD symptoms:}
After completing up to two medication-free weeks (Phase II), participants completed up to 5 weeks of open treatment with MPH to achieve optimal control of ADHD symptoms (Phase III).

For details on dosage, see Supplement 1, available online.

Blinded, randomized add-on citalopram or placebo

Prior to randomization, those who no longer met SMD criteria due to decreased irritability (n = 11) or met other exclusion criteria (n=1) were discharged (see Supplement 1, available online) (Figure 1). Those who continued to meet SMD criteria while on MPH (n = 53) had a second set of baseline measures collected (Baseline at randomization) and then were randomized to add-on treatment with either CTP or PBO for an 8-week double-blind RCT (Phase IV) (Figure S1).

Treatment allocation was randomized by the Pharmaceutical Development Service of the National Institutes of Health Clinical Center Pharmacy Department using a random numbers table. It was randomized in alternating blocks of six and four in a 1:1 ratio.

During the randomized phase, every prescriber and everyone who came into contact with the children was blind to their treatment assignment. Commercial preparations of CTP were obtained from local distributors and investigational (blinded) capsules were compounded and supplied by the NIH Clinical Center Pharmaceutical Development Section.

The CTP/PBO dosing schedule began with one capsule (5 mg) daily; at 5-day intervals the dose could increase by one-capsule. After roughly 3 weeks of the 8-week trial (Mean=3.1 weeks, SD=1.5, with no difference between groups p=0.899), when patients were receiving four capsules (equivalent to 20 mg/d), and clinicians decided it was safe to do so, they were
discharged from inpatient care and returned home. Once home, they received study medication and were monitored with weekly clinical ratings, as described below, by blind raters who were not prescribers. These ratings were done by telephone, alternating with weekly outpatient visits. Based on clinical judgment of minimal side effects and ongoing symptoms, weekly increments were permitted until a maximum dose of 8 capsules per day (equivalent to 40 mg) was achieved. The average dose achieved was 28.33 mg/day (range 5-40 mg/day) or 29.23 mg/day (range 20-40 mg/day) if one participant who withdrew four days after randomization is omitted. During the RCT, lorazepam was available as a PRN medication for agitation; however, no SMD subject received lorazepam during the RCT.

Side Effects:

Side effects were ascertained by checklists as described in the Supplement 1, available online. Information on suicidality and manic symptoms was collected by asking directly to participants and parents.

Assessment

Two graduate-level, highly-experienced clinicians completed weekly the Clinical Global Impression’s improvement (CGI-I) and severity (CGI-S) scales\textsuperscript{46} independently of each other, as described in the Supplement 1, available online, as well as the Children’s Global Assessment of Severity (CGAS),\textsuperscript{47} the Pediatric Anxiety Rating Scale (PARS),\textsuperscript{48} and on the Children’s Depression Rating Scale (CDRS).\textsuperscript{49} Finally, ratings were reached by consensus in a case
conference with a senior child and adolescent psychiatrist (KT); all participants were blind to the
treatment condition.

Our primary categorical clinical outcome was treatment response defined as a CGI-I score of 2
[much improved], or 1 [symptom free], consistent with the original protocol and common
practice. Sensitivity analyses were also conducted using the CGI-S, as in other studies (see
Supplement 1, available online, for detailed information).

The ABC was initially designated as a primary outcome measure; however, it proved
inappropriate for the population we studied as it showed minimal variability (see Supplement 1,
available online, for detailed information). Thus, we stopped collecting it early on. Secondary
outcomes were functional impairment (measured with CGAS), improvement in anxiety
symptoms (measured with PARS), and improvement in depressive symptoms (measured with
CDRS).

Statistical analysis

A statistical analysis plan was written after end of data collection, but prior to the group-label
unblinding of the analyst and statistician (P.V-R. and A.P.). All children who completed at least
one post-randomization assessment were included in intent-to-treat analyses.
Additionally, to examine the effect of stimulant optimization before randomization on irritability severity, we compared changes in CGI-S between baseline at admission and baseline at randomization with a paired t-test, in all randomized subjects.

Primary and secondary outcomes after randomization were analyzed using multilevel models (MLM) estimated by maximum-likelihood enabling the inclusion of participants incompletely observed under the missing at random assumption. This is the recommended analytic approach for repeated measures data\(^5^1\) over other approaches like last observation carried forward (LOCF). However, a sensitivity analyses was carried out using the LOCF method (see Supplement 2, available online).

For our primary categorical outcome (i.e. treatment response as defined by CGI-I score < 3), an efficient estimate was obtained by fitting a growth curve model for the repeated binary response fitted by maximum likelihood in the Stata program \texttt{gsem}, with binomial family and logit link function. The model included a random intercept. The fixed part of the model included treatment group, number of weeks as measure of time, and the group by week interaction (a test for a quadratic term was also carried out). The post-estimation \texttt{lincom} was used to estimate the group difference at the 8th week and its 95\% confidence interval (CI). To assist in interpretation, this conditional or subject-specific effect estimate (and its 95\% CI) was translated to an approximate, but more easily understood, marginal mean estimate.\(^5^2\) The estimate from the group by week interaction in the model provided a measure of difference in response rate between groups across the 8 weeks of the trial. Estimated proportions of response were plotted with
estimates provided by command margins. See Supplement 1, available online, for analysis of continuous outcomes.

Frequencies of the most common adverse effects are reported if present in more than one subject in either study group. These were compared using 2-sided Fisher’s exact test.

## Results

### Participant Sample Derivation and Characteristics

As shown in the CONSORT diagram depicted in Figure 1, of the 53 participants eligible for randomization, 25 were allocated to receive adjunctive citalopram (CTP) and 28 to receive adjunctive placebo (PBO). All subjects completed at least one post-randomization assessment. However, the first four participants recruited – 2 from each group- were excluded from the analysis because of technical problems with data collection. Of the remaining 49 participants included in the intent-to-treat analysis, 41 completed the trial. One participant was withdrawn by the experimenters at week 7 due to increased levels on Liver Function Test (LFT); specifically, alanine transaminase (ALT) and aspartate transaminase (AST) were elevated to three times the upper limit of normal, but these normalized after cessation of methylphenidate, while the patient continued on open citalopram. Another 7 participants withdrew assent before week 8. There were no differences between groups in the number of weeks in inpatient care before participants were discharged to home (CTP+MPH: Mean=3.1 SD=1.1, PBO+MPH: Mean=3.1 SD=1.8 t(45)=−0.13, p=0.899). Table 1 contains the demographic and clinical characteristics of the 49...
participants included in the intent-to-treat analysis. All the participants in the trial had at least one co-morbid diagnosis.

Response to Open Label Stimulant Optimization

Severity of irritability decreased from admission to randomization (ES=0.60, 95%CI 0.30-0.89, \( p<0.001 \)) in the sample included in the intent-to-treat analysis (Figure 2, depicted as the black dotted line). However, of the randomized participants, only one was a responder based on CGI-I score. This is because, as per protocol, those who responded and no longer met irritability threshold criteria (\( n=11 \)) were withdrawn prior to randomization. As expected, stimulant medication had a strong effect on hyperarousal symptoms (ES=1.10, \( p<0.001 \)) given the overlap with ADHD symptoms, but only a moderate effect on temper outbursts (ES=0.54, \( p=0.001 \)), and no effect on mood between outbursts (ES=-0.08, \( p>0.05 \)). We present the detailed results and relevant graphs for separate SMD symptom domains in Supplement 2, available online.

Response to Citalopram versus Placebo

For our primary outcome (i.e., rates of response to treatment as measured with CGI-I), estimated proportions of response differed at the 8th week of treatment between citalopram and placebo groups (35% CTP+MPH vs. 6% PBO+MPH, difference of 29%, \( p=0.005 \); OR=11.70, 95%CI 2.00, 68.16, \( p=0.006 \)) (Table 2), with a number of patients needed to treat (NNT) of three (3). In addition, the group by week interaction for the estimated proportions of response across the 8
weeks of trial was also significant (b=0.43, 95%CI 0.12, 0.74, p=0.006) (Table 2). Figure 3 depicts the estimated proportion of response for all participants included in the analyses against the observed proportion of responses for individuals with available data.

The difference in irritability severity - as measured with the CGI-S - at the 8th week of trial between groups was a non-significant trend (b=-0.62, 95%CI -1.32, 0.09, p=0.085, ES=-1.11 95%CI -2.38, 0.15). However, a significant group by week interaction for the CGI-S emerged across the 8 weeks of the trial (b=-0.11, 95%CI -0.21, -0.002, p=0.046) (Table 2 and Figure 2). This difference was primarily driven by changes in the severity of temper outbursts (Figure S2) and, to a lesser extent, irritable mood between outbursts (Figure S3). No difference was observable for hyperarousal symptoms (Figure S4)

Groups did not differ in functional impairment as measured with the CGAS at the 8th week of trial (b=4.72, 95%CI -1.30, 10.74 p=0.124, ES=1.36 95%CI -0.37, 3.09), and the group by week interaction was not significant (b=0.75, 95%CI -0.17, 1.66, p=0.109) (Table 2).

Groups did not differ in severity of depressive symptoms as measured with the CDRS at the 8th week of the trial (b=0.02, 95%CI -4.76, 4.80, p=0.993, ES=0.00 95%CI -0.62, 0.62), and the group by week interaction was not significant (b=-0.15, 95%CI -0.87, 0.57, p=0.680) (Table 2). However, the CDRS irritability item resulted in a non-significant trend at the 8th week (b=0.77, 95%CI -0.02, 1.56, p=0.057, ES=0.50 95%CI -0.02, 1.02). Similarly, groups did not differ in severity of anxiety symptoms as measured with the PARS at the 8th week of the trial (b=-1.02,
95%CI -4.23, 2.19, \( p = 0.534 \), ES=-0.19 95%CI -0.79, 0.41), and the group by week interaction was not significant (\( b=0.02, 95\% CI -0.40, 0.45, p=0.909 \)).

Distribution of lowest level residuals was close to normal, so no transformations were required. Results and estimates of the sensitivity analyses, which employed last observation carried forward method, were very similar to the results described above (see Supplement 2, available online). Finally, number of weeks in inpatient care did not have an impact in the outcomes.

**Tolerability and Adverse Effects**

**Table S1** shows rates of adverse effects in each of the treatment groups. No differences were found in any adverse effect or total number of adverse effects between treatment groups (see Supplement 3, available online).

**Discussion**

This is the first RCT of an SRI for the treatment of irritability in youth. Specifically, we report the findings of an 8-week, flexible-dose, double-blind, placebo-controlled trial of CTP as an add-on to open-label MPH in children with severe irritability. Our results provide some support for the efficacy of CTP+MPH in the treatment of irritability in children with severe irritability who are already receiving methylphenidate. However, it is important to note that there were no differences in functional impairment between groups at the end of the trial. Of note, while
participants were initially recruited because they fulfilled SMD criteria, all but one participant also fulfilled DMDD criteria.

As expected, open-label treatment with MPH led to a significant reduction in SMD symptoms, specifically hyperarousal and temper outburst, with effect sizes consistent with what has previously been reported. The RCT’s primary outcomes are suggestive of citalopram’s efficacy. That is, at week 8 a significantly higher percentage of patients on citalopram improved, compared to those who received placebo; this was also true when taking all time points into account. At week 8, the reduction in the continuous score of irritability severity did not reach significance (p=0.085), although the week by group interaction across the 8 weeks did (p=0.046). The differentiation between citalopram and placebo appeared to emerge around week 5 (for observed data, between-group differences were: week 5, p=0.046; week 6, p=0.021; week 7, p=0.007). This time course is consistent with the timing of action of SRIs. Importantly, the confidence intervals of the effect sizes vary from relatively small to large, indicating considerable uncertainty about the magnitude of the effect.

We note that the magnitude of the response to SRIs in this study was relatively low compared to that in anxiety and depression trials. In addition, the placebo response rate was also extremely low (i.e., about 5%). Together, these two observations explain the very favorable number needed to treat (NNT) of 3. However, whilst the response rate to active medication may appear low in comparison to effects observed in non-add-on trials, it is consistent with what is observed in add-on trials. For example, the remission rate in STAR*D for add-on treatment was 30% and in CO-MED it was 38%. Similarly, in a meta-analytic study, add-on treatment for ADHD
showed smaller effect sizes than monotherapies.\textsuperscript{57} Thus, it is unclear whether the relatively low response rate to CTP+MPH treatment (and placebo) in irritability in this trial, compared to those for depression or anxiety, is due to differences in the target phenotype or in trial design.

At the end of the trial, there were no differences in the CGAS scores between groups, with participants still showing moderate functional impairment after completion (or severe in at least one area, range 40-50). This suggests that improvements in irritability appear not to have translated into reductions in overall impairment with siblings, peers, and at school within the time scale of the trial. This interpretation would be in keeping with what has been described previously, in that it is common to see a dissociation between symptoms and impairment levels.\textsuperscript{58} However, it is important to note that the CGAS focuses on the lowest level of function during the period under consideration and thus may be suboptimal as an outcome measure\textsuperscript{47} for disorders like DMDD. Since the functioning of a chronically irritable child during the previous week could include even short bursts of dysregulation, the CGAS might be expected to lag behind a measure of modal functioning.

We also observed no significant differences in CDRS scores. We note that depressive symptoms were low throughout the trial for most individuals, which is not surprising given that individuals with ongoing MDD were ineligible. In addition, no significant differences were observed for anxiety symptoms as measured with PARS. PARS scores were in the mid-range and barely changed in either group during the trial. It is well known that irritability frequently overlaps with depression and anxiety in patients and that they share etiological underpinnings.\textsuperscript{39,40} Our results...
suggest that improvements in irritability with SRI treatment may be seen in the absence of change in depression and anxiety.

The rate of side effects did not differ between CTP+MPH and PBO+MPH. However, the absolute level of side effects was high in both groups. This may be a consequence of directly asking about side effects and the detailed observations that trained nursing staff offered during the inpatient phase of the study. In addition, nurses and parents were instructed to rate what they observed without reference to changes from baseline irritability or time of day, and thus even low levels of severity were reported in the checklist. Only one patient was withdrawn by experimenters due to adverse effects; specifically, at week 7 due to elevated liver function tests which normalized with methylphenidate discontinuation. We observed no hypomania or mania in any participant at any phase of the trial. This is consistent with previous findings from studies showing that chronic irritability is not a presage of bipolar disorder type I or type II and that chronically irritable youth are not at elevated risk for manic switches. However, the number of participants exposed to citalopram in this trial was too small to draw firm conclusions about the risk for medication-induced mania in this population.

Our study has several strengths. It used a rigorous double-blind, placebo-controlled approach in a well-phenotyped sample of youth to investigate the effects of a CTP+MPH in DMDD. It also employed an initial open-label stimulant optimization phase which resembles common clinical practice and reflects a rational approach to treatment of irritability in the presence of ADHD. The design of the study should also be placed in the context of where the field was when it began a
decade ago. The design maximized human subjects’ protections by using hospitalization for safety and standardizing treatment and environmental care prior to randomization.

Our study also has limitations. First, the sample size was smaller than initially planned. In the 10-year duration of the study, we pursued nation-wide recruitment using local resources, sending letters to every practicing child and adolescent psychiatrist in the American Academy of Child and Adolescent Psychiatry, and giving local talks and presentations across the country. While the initial response to the study was positive, over time, the response declined steadily and a corresponding increase in the use of SRIs - as well as MPH - for chronic irritability took place in clinical practice. Yet, our experience in recruitment was not unique. There are several potential explanations for this, including the increasing familiarization of community practitioners with prescribing for youth with chronic irritability, especially after the publication of studies showing that chronic irritability does not predict bipolar disorder. Nonetheless, we note that at the current sample size, the study had a power of approximately 75% to detect a difference at $p<0.05$. Clearly, further study with a larger sample would be warranted. A second limitation of this study is that it did not use a broad array of dimensional measures of irritability or other psychopathology. The study preceded the psychometric development of the Affective Reactivity Index (ARI), or the Multidimensional Assessment of Preschool Disruptive Behavior (MAP-DB) scales, which would have been useful adjuncts. To date, though, there are still no validated self- or parent-report scales for assessing change, as opposed to trait measurement, of irritability. Future trials should use more comprehensive multi-source, multi-method approaches to the measurement of the clinical phenotype. A third limitation is that the sample in this study may not be generalized to all children with DMDD, since we initially recruited a population with
SMD, which requires hyper-arousal symptoms (e.g., insomnia, agitation, distractibility) and might represent a more severe subgroup of the DMDD population. Furthermore, there are limitations of generalizing from subjects coming to research at NIMH to other populations because of the design of the study. Indeed, the length of the study, the compulsory inpatient interval, and requiring that patients come off all psychotropic medication, were all obstacles to participation and limit the generalizability of our findings. Finally, the design of this trial required a period of hospitalization in an inpatient unit, which some studies have linked to initial improvement of irritability-related behaviors\(^65,66\) and might also explain the discordance between different informants.\(^67\) However, all participants were subjected to this potential phenomenon, and therefore it does not explain the distinct treatment effects seen after randomization.

In summary, we find that, among those unresponsive to stimulant alone, severe irritability symptoms improved in significantly more children who received add-on citalopram than those receiving add-on placebo. However, our results do not show evidence of effect on impairment in those patients.

**Lay summary:**

After evaluating and treating children who had severe chronic irritability using optimal doses of methylphenidate, the authors randomly assigned participants to receive add-on citalopram or placebo. All treaters, patients and family were blind as to which agent patients received. After 8 weeks on study medication, 35% of those who were prescribed add-on citalopram, compared to 6% of participants who were prescribed add-on placebo, responded to the treatment. No episodes of hypo-mania or mania were noted during the trial.
Clinical guidance:

A combination of stimulants plus SSRIs may be beneficial to target the condition of chronic severe irritability and symptoms of hyperarousal. A stepped approach that starts with optimal stimulant medication is appropriate.

References


Interventions for Anger, Irritability, and Aggression in Children and Adolescents.


15. Walkup JT. Antidepressant Efficacy for Depression in Children and Adolescents:

combination for adolescents with depression: Treatment for Adolescents With

17. Group TMC. A 14-month randomized clinical trial of treatment strategies for attention-
deficit/hyperactivity disorder. Multimodal Treatment Study of Children with ADHD.
*Archives of general psychiatry.* 1999;56(12):1073-1086.

18. Cortese S, Adamo N, Del Giovane C, et al. Comparative efficacy and tolerability of
medications for attention-deficit hyperactivity disorder in children, adolescents, and

with conduct disorder: a double-blind and placebo-controlled study. *Journal of the

20. Malone RP, Delaney MA, Luebbert JF, Cater J, Campbell M. A double-blind placebo-
controlled study of lithium in hospitalized aggressive children and adolescents with


Disorder and Disruptive Mood Dysregulation Disorder Across Home and School. 


<table>
<thead>
<tr>
<th></th>
<th>MPH + CTP n=23 SMD</th>
<th>MPH + PBO n=26 SMD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, female, n (%)</strong></td>
<td>10 (43)</td>
<td>6 (23)</td>
</tr>
<tr>
<td><strong>Age (years), Mean (SD), range</strong></td>
<td>11.4 (2.5), 7-15</td>
<td>11.7 (2.1), 8-14</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>19 (83)</td>
<td>20 (77)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0 (0)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Asian</td>
<td>1 (4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Multiple races</td>
<td>3 (13)</td>
<td>2 (8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>IQ, Mean (SD)</strong></td>
<td>109.3 (14.8)</td>
<td>113 (9.2)</td>
</tr>
<tr>
<td><strong>Co-occurring diagnoses, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any disorder</td>
<td>23 (100)</td>
<td>26 (100)</td>
</tr>
<tr>
<td>ADHD</td>
<td>20 (87)</td>
<td>24 (92)</td>
</tr>
<tr>
<td>ODD</td>
<td>17 (74)</td>
<td>22 (85)</td>
</tr>
<tr>
<td>CD</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Any Anxiety disorder</td>
<td>13 (57)</td>
<td>15 (58)</td>
</tr>
<tr>
<td><strong>Medicated at admission, n (%)</strong></td>
<td>21 (91)</td>
<td>23 (84)</td>
</tr>
<tr>
<td><strong>Number of medications at admission, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>2 (9)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>1</td>
<td>4 (17)</td>
<td>5 (19)</td>
</tr>
<tr>
<td>2</td>
<td>5 (22)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>3 or more</td>
<td>12 (52)</td>
<td>15 (57)</td>
</tr>
<tr>
<td><strong>Type of medication, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atypical antipsychotic</td>
<td>14 (60)</td>
<td>17 (65)</td>
</tr>
<tr>
<td>Typical antipsychotic</td>
<td>0 (0)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Stimulant</td>
<td>8 (35)</td>
<td>9 (35)</td>
</tr>
<tr>
<td>Non-stimulant (Ato, Gua, Clo)</td>
<td>9 (39)</td>
<td>12 (46)</td>
</tr>
<tr>
<td>SRI</td>
<td>7 (30)</td>
<td>7 (27)</td>
</tr>
<tr>
<td>Other antidepressant</td>
<td>3 (13)</td>
<td>3 (12)</td>
</tr>
<tr>
<td>Anti-convulsant</td>
<td>10 (43)</td>
<td>7 (27)</td>
</tr>
<tr>
<td>Lithium</td>
<td>4 (17)</td>
<td>3 (12)</td>
</tr>
<tr>
<td><strong>Prior medication treatments, Mean (SD), range</strong></td>
<td>4.1 (3.5), 0-13</td>
<td>5.3 (3.1), 1-13</td>
</tr>
<tr>
<td><strong>Prior hospitalizations, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>15 (65)</td>
<td>15 (58)</td>
</tr>
<tr>
<td>1</td>
<td>2 (9)</td>
<td>5 (19)</td>
</tr>
<tr>
<td>2 or more</td>
<td>6 (26)</td>
<td>6 (23)</td>
</tr>
<tr>
<td><strong>Educational programs, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mainstream school (MS)</td>
<td>11 (48)</td>
<td>4 (15)</td>
</tr>
<tr>
<td>MS with accommodations</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>In-home school</td>
<td>1 (4)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Special education</td>
<td>10 (43)</td>
<td>20 (77)</td>
</tr>
<tr>
<td><strong>Baseline admission, Mean (SD)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGI-S a</td>
<td>4.4 (0.6)</td>
<td>4.6 (0.5)</td>
</tr>
</tbody>
</table>
### Table 2. Summary of Primary and Secondary Outcomes of the Randomized Controlled Trial with Adjunctive Citalopram vs. Placebo

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Baseline randomization</th>
<th>8th Week of trial</th>
<th>Between group difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CTP</td>
<td>PBO</td>
<td>Group by week interaction</td>
</tr>
<tr>
<td>CGI-I response</td>
<td>% (SE) n</td>
<td>% (SE) n</td>
<td>% (SE) n</td>
</tr>
<tr>
<td>- Estimated</td>
<td>-</td>
<td>-</td>
<td>35 (10) 8</td>
</tr>
<tr>
<td>- Observed</td>
<td>-</td>
<td>-</td>
<td>25 (10) 5</td>
</tr>
<tr>
<td>CGI-S</td>
<td>4.0 (0.4) 23</td>
<td>4.3 (0.4) 26</td>
<td>3.1 (0.3) 26</td>
</tr>
<tr>
<td>CGAS</td>
<td>44.4 (0.2) 22</td>
<td>42.9 (0.2) 26</td>
<td>52.6 (2.3) 22</td>
</tr>
<tr>
<td>CDRS</td>
<td>29.4 (0.5) 16</td>
<td>34.6 (0.7) 18</td>
<td>28.6 (1.8) 16</td>
</tr>
<tr>
<td>PARS</td>
<td>13.3 (1.2) 23</td>
<td>15.8 (1.0) 25</td>
<td>12.0 (1.2) 23</td>
</tr>
</tbody>
</table>

**Note:** CDRS = Children’s Depression Rating Scale; CGAS = Children’s Global Assessment Severity; CGI = Clinical Global Impression; CTP = citalopram; M = Mean; MPH = methylphenidate; PARS = Pediatric Anxiety Rating Scale; PBO = placebo; SE = standard error.

*aBaseline descriptive statistics are based in observed data.*
b Descriptive statistics at 8th week of the trial and p-values of differences in the randomized controlled trial are based on model-based estimates of the intent-to-treat analysis.

c For CGI-I response, both estimated (n=49) and observed data (n=41) are provided.
**Figure 1.** CONSORT Diagram of the Trial.

**Note:** Of the randomized participants (n=53), 4 participants were excluded from analysis because the CGI collected was a different version and thus not comparable. Of the 49 participants included in the analyses, 41 completed the trial; 7 withdrew assent before the 8th week and 1 was ruled out of the study by the experimenters due to high levels of liver function test.

**Figure 2.** Change in Irritability Severity Before and After Randomization in the Sample Included in the Intent-to-treat Analysis.

**Note:** The period between baseline at admission (Adm. Baseline) and baseline at randomization (Rand. Baseline) was variable and included the washout period (Phase I), medication-free period (Phase II), and open-label lead phase with stimulant optimization (Phase III). Before randomization (depicted as the black dotted line), change in irritability severity is shown for the entire sample (n=49). After randomization, both observed and estimated severity - provided by command `margins` in Stata - are displayed by week and treatment group (CTP, n=23; PBO, n=26).

**Figure 3.** Proportion of Treatment Response by Week and Treatment Group.

**Note:** Both observed and estimated proportions of response are displayed. Bars represent 95%CI. Estimated proportions were extracted with command `margins` in Stata and are based on n=49. Observed proportions are based on n=49 at week 1; n=48 at week 2; n=47 at weeks 3, 4; n=43 at weeks 5, 6, 7; and n=41 at week 8.
Potential participants screened on the phone (N=1,264)

Potential participants screened on-site (N=311)

Participants enrolled in trial (N=103)

Phase I: Medication withdrawal (N=103)

Phase II: Medication free period (N=69)

Phase III: Open treatment with methylphenidate (N=69)

Phase IV: Participants randomized (N=53)

Citalopram + Methylphenidate (N=25)
  - Eight weeks.
  - Weekly assessment
  - Withdrew assent (N=3)
  - Analyzed (N=23)
  - Excluded from analyses due to different CGI (N=2)

Placebo + Methylphenidate (N=28)
  - Eight weeks.
  - Weekly assessment
  - Withdrew assent (N=4)
  - Ruled out (N=1)
  - Analyzed (N=26)
  - Excluded from analyses due to different CGI (N=2)

Not eligible, not contactable or declined (N=953)

Not eligible (N=85) Declined (N=113) Not contactable (N=10)

Withdrew assent prior to or during medication taper (N=22)

Not eligible due to other exclusion criteria (N=12)

Withdrew assent after medication taper (N=4)

Not eligible due to low irritability (N=11)

Not eligible due to other exclusion criteria (N=1)
Figure 2
Figure 3
A double-blind randomized placebo-controlled trial of citalopram adjunctive to stimulant medication in youth with chronic severe irritability

Supplemental material

**Figure S1.** Diagram of the Trial Design. **Note:** Any of the patients used Day Treatment.
Supplement 1

Enrollment

Reasons for exclusion before randomization:

In total, of the 103 participants enrolled in the trial, 13 were discharged for meeting the following exclusion criteria: brain abnormalities found during magnetic resonance imaging (MRI) (n=2), presented autism spectrum disorder (ASD) symptoms or diagnosis (n=5; one with possible psychotic symptoms), too much aggression (n=3), conduct disorder (CD) (n=1, with possible psychotic symptoms), posttraumatic stress disorder (PTSD) (n=1), and excessive anxiety about randomization (n=1).

Treatment of ADHD symptoms

After completing up to two medication-free weeks (Phase II), participants completed up to 5 weeks of open treatment with MPH to achieve optimal control of ADHD symptoms (Phase III). Only oral administration of MPH (immediate-release, intermediate release, or extended release forms) was used. Commercial preparations of MPH (in its various forms) were obtained from local distributors. MPH dosing was based on clinical judgement optimizing effects on attention, restlessness, and impulsivity while balancing adverse effects on sleep and appetite. Dose was not fixed by protocol but a dose range was allowed based on efficacy and tolerability. The goal was 0.8 mg/kg/day to 1.2 mg/kg/day not to exceed 80 mg or 2 mg/kg/day. Treatment decisions about MPH drew from reports from nursing staff and classroom teachers at the NIH Clinical Center School which patients attended during their inpatient care. MPH treatment began with immediate release forms twice daily and then were converted to intermediate or extended release forms.
forms of MPH to achieve once-a-day dosing to the greatest degree possible. Morning dosing with immediate plus longer release forms of MPH was permitted.

Assessment

CGI for chronic severe irritability

Children in this cohort typically presented with multiple comorbidities (most commonly, anxiety disorders and ADHD, Table 1). Since the goal of this trial was to determine the impact of stimulant plus SRI treatment on irritability specifically, the primary outcome was a CGI-I scale focused on irritability; a similar strategy has been used in bipolar disorder ¹.

On admission and every week thereafter, clinicians (masters or PhD-level) blind to treatment assignment used a semi-structured interview to ascertain the severity of mood dysregulation symptoms (i.e., temper outbursts, irritable mood between outbursts, hyperarousal). The “or” rule was applied to collate these reports. These were used to generate a CGI-I SMD score relative to the defined baseline (i.e., baseline for the time just prior to admission and baseline before randomization); in this score, hyperarousal symptoms were given minimal weight, so that the score largely reflected temper outbursts and irritable mood between outbursts. To ensure reliability, the correspondence between each rating and the clinical documentation supporting that rating was reviewed by a senior child psychiatrist (KT) who was also blind to treatment assignment.

CGI reliability
The CGI was derived by the consensus of two independent clinician-raters (blind to treatment assignment) who, for the inpatient segment of the study, interviewed the patient and primary nursing staff member (and sometimes parents, since patients were often on pass with their parents on weekends) and at admission baseline and for the outpatient segments, interviewed the parent and child. The interviewers were trained on the KSADS to achieve inter-rater reliability above 0.85. In addition, the clinician raters achieved inter-rater reliability of 0.85 or greater on symptoms of severe mood dysregulation used in the CGI. The domains of these symptoms were: temper outbursts, mood between outbursts and hyper-arousal symptoms. The interviews lasted on average about an hour per case. On the basis of these interviews, the two clinician-raters generated a consensus on the CGI (CGI-S and CGI-I) and these consensus ratings were reviewed by a senior clinician (KT) who was also blind to treatment assignment.

Rationale to stop collecting ABC

The ABC was developed with atypically developing populations in mind. To the best of our knowledge, ours is the first study that applied the ABC-I to a non-ASD or non-intellectually disabled population. We had used it in the absence of another measure, but experienced difficulties with its use in two ways.

First, we found inconsistency across reporters. The study was designed so that the ABC was rated by parents at admission, nursing staff (untrained in reliability) during the patient’s hospitalization, and parents again after discharge (i.e., during the last weeks of the randomized controlled trial). We observed little agreement between these raters, with parents (typically, but
not invariably) scoring children highly, and nurses’ ratings showing a floor effect (see below). Thus, the ABC ratings gave an inconsistent picture across a patient’s time in the study. The lack of variability—floor effect with mean scores of as low as 4—during the patients’ hospitalization (when nurses were rating) was particularly problematic. Importantly, these ABC ratings were at odds with the observations of the blinded trained raters, the children’s experiences, and indeed with information obtained from the nurses themselves by our blinded trained raters. In addition, we observed that there was a very low incidence of at least 6 items that are common in those with ASD and/or intellectual disabilities (e.g. injures self, screams inappropriately, cries over minor hurts, stamps feet/bangs objects, hurts him/herself, physical violence to self).

The second reason for eventually dropping the ABC as a measure was that it did not seem to capture the observed severity of the clinical presentation. We collected the ABC in 25 children before deciding to stop it. Whereas parents tended to provide high ABC ratings, we observed five cases where, at admission, the parent-rated ABC was implausibly below 12. (which, initially, was the threshold for inclusion). Indeed, both the clinical presentation at admission -as assessed by two senior graduate-level clinicians and supervised by a senior child psychiatrist (KT) who were blind to the parents’ ABC-I rating - as well as the patients’ history indicated severe irritability. Whilst beyond our scope to test empirically, we hypothesize that these low parental ratings may result from accommodation, e.g. parents giving in to children’s demands, or reducing parental expectations, in order to avoid children’s temper outbursts. We eventually decided to drop the ABC because of these problems and admit those patients to the study.
Side Effects

Side effects were ascertained by checklists. These were completed by nurses (during inpatient admission) or parents (after discharge), all of whom were blind to treatment assignment, and instructed to rate what they observed without reference to changes from baseline irritability, time of day or severity of the presentation.

Suicidality information was collected by asking directly to participants/parents, and in addition, this information is also collected as part of the CDRS with two items (i.e., suicidal ideation and morbid ideation). Manic symptoms were also asked directly, especially when participants reported changes in sleep, energy, or levels of activity (which are collected in the adverse effect checklist).

Statistical Analysis

For the analysis of continuous outcomes (i.e., severity as measured by CGI-S, functionality as measured by CGAS, anxiety symptoms as measured by PARS, and depressive symptoms as measured by CDRS) a growth curve model was fitted using the \texttt{xtmixed} routine of the Stata statistical program, with restricted maximum likelihood (REML) estimation and an exchangeable covariance matrix for the covariances of the errors of the repeated measures. The fixed part of the model included the baseline values of the outcome variables, treatment group, number of weeks as measure of time, and the interaction of group by week. The post-estimation \texttt{lincom} was used to estimate the group difference at 8th week and its 95% CI. The estimate from the group by week interaction in the model provided a measure of difference in severity between...
groups across the 8 weeks of the trial. Distributional assumptions of our primary outcomes were checked by the use of Q-Q plots of residuals.

**Supplement 2**

*Results*

**Sensitivity analysis using LOCF – Response to Citalopram versus Placebo:**

For our primary outcome (i.e., rates of response to treatment as measured with CGI-I), estimated proportions of response differed at the 8th week of treatment between citalopram and placebo groups (33% CTP vs 5% PBO; OR=12.31, 95%CI 2.20-68.89, \( p=0.004 \)), with 3 being the number of patients needed to treat (NNT). In addition, the week by group interaction for the estimated proportions of response across the 8 weeks of trial was also significant (b=0.48, 95%CI 0.15, 0.80, \( p=0.004 \)).

For the CGI-S, the difference in irritability severity between groups at week 8 was a non-significant trend (b=-0.66, 95%CI -1.34, 0.02, \( p=0.056 \), ES=-1.20 95%CI -2.43-0.03). However, a significant week by treatment group interaction emerged when taking into account the difference in irritability severity across the 8 weeks of the trial (b=-0.11, 95%CI -0.21, -0.01, \( p=0.029 \)).
In terms of functional impairment as measured with CGAS a non-significant trend emerged both at the 8th week of trial \((b=4.82, 95\% CI -0.73, 10.36, p=0.089, ES=0.63, 95\% CI 0.06-1.20)\), and in the week by group interaction \((b=0.77, 95\% CI -0.04, 1.59, p=0.063)\).

Groups did not differ in severity of depressive symptoms as measured with the CDRS at the 8th week of the trial \((b=-0.37, 95\% CI -4.77, 4.03, p=0.869, ES=-0.05, 95\% CI -0.62-0.52)\), and the week by group interaction was not significant \((b=-0.18, 95\% CI -0.84, 0.49, p=0.602)\). Similarly, groups did not differ in severity of anxiety symptoms as measured with the PARS at the 8th week of the trial \((b=-1.14, 95\% CI -4.20, 1.92, p=0.465, ES=-0.21, 95\% CI -0.78-0.36)\), and the week by group interaction was not significant \((b=0.00, 95\% CI -0.40, 0.40, p=1.000)\).

Response to Open Label Stimulant Optimization by SMD symptom domains

Here we report the changes in SMD symptom domains (i.e., temper outbursts, irritable mood between outbursts and hyperarousal symptoms) in the sample included in the intent-to treat analysis between admission and randomization.

Severity of temper outbursts decreased from admission to randomization \((ES=0.54, 95\% CI 0.24-0.82, p=0.0006)\) (Figure S1, depicted as the black dotted line). Severity of irritable mood between outbursts did not significantly change from admission to randomization \((ES=-0.08, 95\% CI -0.37-0.21, p=0.585)\) (Figure S2, depicted as the black dotted line). Severity of hyperarousal symptoms showed the larger decrease between admission and randomization \((ES=1.10, 95\% CI 0.81-1.40, p<0.0001)\) (Figure S3, depicted as the black dotted line).
Response to Citalopram versus Placebo by SMD symptom domains

Here we report the changes in SMD symptom domains (i.e., temper outbursts, irritable mood between outbursts and hyperarousal symptoms) in the sample included in the intent-to treat analysis during the 8-week RCT.

The difference in severity of temper outbursts - as measured with the CGI-S - at the 8th week of trial between groups was a non-significant trend (b=-0.68, 95%CI -1.38, 0.03, p=0.059, ES=-1.12 95%CI -2.29-0.04). However, a significant group by week interaction for the CGI-S emerged across the 8 weeks of the trial (b=-0.11, 95%CI -0.22, -0.01, p=0.039) (Figure S2).

The difference in severity of irritable mood between outbursts - as measured with the CGI-S - at the 8th week of trial between groups was not significant (b=-0.59, 95%CI -1.31, 0.13, p=0.107, ES=-0.85 95%CI -1.87-0.18). However, a significant group by week interaction for the CGI-S emerged across the 8 weeks of the trial (b=-0.11, 95%CI -0.21, -0.01, p=0.035) (Figure S3).

The difference in severity of hyperarousal symptoms - as measured with the CGI-S - at the 8th week of trial between groups was not significant (b=-0.11, 95%CI -0.53, 0.31, p=0.610, ES=-0.16 95%CI -0.79-0.46). However, a significant group by week interaction for the CGI-S emerged across the 8 weeks of the trial (b=-0.07, 95%CI -0.13, -0.01, p=0.030) (Figure S4).
Figure S2. Change in temper outbursts severity before and after randomization in the sample included in the intent-to-treat analysis. The period between baseline at admission (Adm. Baseline) and baseline at randomization (Rand. Baseline) was variable and included the washout period (Phase I), medication-free period (Phase II), and open-label lead phase with stimulant optimization (Phase III). Before randomization (depicted as the black dotted line), change in irritability severity is shown for the entire sample (N=49). After randomization, both observed and estimated severity - provided by command margins in Stata - are displayed by week and treatment group (CTP, N=23; PBO, N=26).
Figure S3. Change in mood between outbursts severity before and after randomization in the sample included in the intent-to-treat analysis. The period between baseline at admission (Adm. Baseline) and baseline at randomization (Rand. Baseline) was variable and included the washout period (Phase I), medication-free period (Phase II), and open-label lead phase with stimulant optimization (Phase III). Before randomization (depicted as the black dotted line), change in irritability severity is shown for the entire sample (N=49). After randomization, both observed and estimated severity - provided by command margins in Stata - are displayed by week and treatment group (CTP, N=23; PBO, N=26).
**Figure S4.** Change in *hyperarousal symptoms* severity before and after randomization in the sample included in the intent-to-treat analysis. The period between baseline at admission (Adm. Baseline) and baseline at randomization (Rand. Baseline) was variable and included the washout period (Phase I), medication-free period (Phase II), and open-label lead phase with stimulant optimization (Phase III). Before randomization (depicted as the black dotted line), change in irritability severity is shown for the entire sample (N=49). After randomization, both observed and estimated severity - provided by command `margins` in Stata - are displayed by week and treatment group (CTP, N=23; PBO, N=26).
Supplement 3

Tolerability and Adverse Effects

In addition, there were no differences in total number of adverse effects reported between groups (CTP+MPH, Mean=14.3 SD=7.1; PBO+MPH, Mean=11.5 SD=6.1; t(47)=1.50, \( p=0.138 \)). Whereas participants were asked directly about suicidal ideation during the trial, only one participant receiving CTP+MPH reported passive suicidal ideation on week 6. However, the participant indicated that this ideation was no longer present at week 7 and 8. No trial participant exhibited hypomanic or manic symptoms. The most common reported effects in the citalopram group were changes in appetite (100%), anger (83%), and aggression along with insomnia and intrusiveness (all three 74%). In the placebo group, the most common reported adverse effects were insomnia (92%), anger (85%), and changes in appetite (81%).

<p>| Table S1. Adverse Effect Rates of Youth Randomly Allocated to Citalopram or Placebo |
|----------------------------------------|----------------------------------------|----------------------------------------|</p>
<table>
<thead>
<tr>
<th></th>
<th>MPH + CTP (n=23)</th>
<th>MPH + PBO (n=26)</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gastrointestinal system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dry Mouth</td>
<td>3</td>
<td>4</td>
<td>1.000</td>
</tr>
<tr>
<td>Drooling</td>
<td>2</td>
<td>0</td>
<td>0.215</td>
</tr>
<tr>
<td>Increased thirst</td>
<td>3</td>
<td>3</td>
<td>1.000</td>
</tr>
<tr>
<td>Trouble swallowing</td>
<td>2</td>
<td>3</td>
<td>1.000</td>
</tr>
<tr>
<td>Nausea</td>
<td>10</td>
<td>8</td>
<td>0.390</td>
</tr>
<tr>
<td>Vomiting</td>
<td>5</td>
<td>3</td>
<td>0.448</td>
</tr>
<tr>
<td>Stomach pains</td>
<td>13</td>
<td>16</td>
<td>0.777</td>
</tr>
<tr>
<td>Bloated abdomen</td>
<td>3</td>
<td>0</td>
<td>0.096</td>
</tr>
<tr>
<td>Changes in stool</td>
<td>8</td>
<td>6</td>
<td>0.528</td>
</tr>
<tr>
<td>Cardiovascular system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change in heart rate fast/slow</td>
<td>1</td>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td>Central nervous system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drowsiness morning/afternoon</td>
<td>7</td>
<td>3</td>
<td>0.157</td>
</tr>
<tr>
<td>Dizziness</td>
<td>5</td>
<td>1</td>
<td>0.086</td>
</tr>
<tr>
<td>General well-being</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condition</td>
<td>Count1</td>
<td>Count2</td>
<td>Count3</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Tiredness/fatigue</td>
<td>12</td>
<td>52</td>
<td>9</td>
</tr>
<tr>
<td>Insomnia</td>
<td>17</td>
<td>74</td>
<td>24</td>
</tr>
<tr>
<td>Early morning awakening</td>
<td>3</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Appetite changes</td>
<td>23</td>
<td>100</td>
<td>21</td>
</tr>
<tr>
<td>Weight changes</td>
<td>13</td>
<td>57</td>
<td>13</td>
</tr>
<tr>
<td>Fever</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td><strong>Genitourinary system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enuresis</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td><strong>Head and neck</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eyes sensitive to light</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Headache</td>
<td>11</td>
<td>48</td>
<td>12</td>
</tr>
<tr>
<td>Ringing or buzzing in ears</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Runny nose</td>
<td>6</td>
<td>26</td>
<td>7</td>
</tr>
<tr>
<td><strong>Hematological changes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy bruising</td>
<td>2</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td><strong>Mental status changes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anger</td>
<td>19</td>
<td>83</td>
<td>22</td>
</tr>
<tr>
<td>Aggression</td>
<td>17</td>
<td>74</td>
<td>12</td>
</tr>
<tr>
<td>Nervousness</td>
<td>15</td>
<td>65</td>
<td>13</td>
</tr>
<tr>
<td>Clingy/separation anxiety</td>
<td>13</td>
<td>57</td>
<td>8</td>
</tr>
<tr>
<td>Loss of interest/apathy</td>
<td>7</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Poor concentration</td>
<td>9</td>
<td>39</td>
<td>8</td>
</tr>
<tr>
<td>Confusion</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Disorganized thinking</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Memory loss</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Speech changes</td>
<td>7</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Intrusiveness</td>
<td>17</td>
<td>74</td>
<td>15</td>
</tr>
<tr>
<td>Paranoia</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td><strong>Musculoskeletal system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restless/Inability to sit still</td>
<td>16</td>
<td>70</td>
<td>19</td>
</tr>
<tr>
<td>Stiffness/Involuntary movements</td>
<td>2</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Extreme stiffness/lack of movement</td>
<td>2</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Muscle twitches</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Motor tics</td>
<td>8</td>
<td>35</td>
<td>11</td>
</tr>
<tr>
<td><strong>Pulmonary system</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty breathing</td>
<td>3</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Severe or chronic cough</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Wheezing</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td><strong>Skin changes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rash/itch</td>
<td>8</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>Increased sweating</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Dry skin</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Hair changes</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Acne</td>
<td>4</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>Picking at skin or nails</td>
<td>5</td>
<td>22</td>
<td>2</td>
</tr>
</tbody>
</table>

**Note:** Table only shows adverse events that were reported by more than one participant.
Adverse events that were not reported or reported only once by one participant included: changes in taste, irregular heartbeat, seizures, fainting, increased sensitivity to heat/cold, nightmares/vivid dreams, increased/decreased urination, menstrual changes, irregular cramping, dark or brown urine, vision changes like blurred or double vision, delusions, hallucinations, trouble keeping balance when standing or walking, slowing of body movements, tremor of hands or fingers, extremity changes (numb/tingling/cold/swelling hands/feet/fingers/toes), leg spasms at night, increased sensitivity to sunlight, blisters, folliculitis, severe dandruff, and psoriasis