

Open camera or QR reader and  
scan code to access this article  
and other resources online.



# A Characterization of the Clinical Global Impression Scale Thresholds in the Treatment of Adolescent Depression Across Multiple Rating Scales

Carl Y. Zhang,<sup>1</sup> Jennifer L. Vande Voort, MD,<sup>2</sup> Deniz Yuruk, MD,<sup>2</sup> Jeffrey A. Mills, PhD,<sup>3</sup>  
Graham J. Emslie, MD,<sup>4,5</sup> Betsy D. Kennard, PsyD,<sup>4</sup> Taryn Mayes, MS,<sup>4</sup> Madhukar Trivedi, MD,<sup>4</sup>  
William V. Bobo, MD, MPH,<sup>6</sup> Jeffrey R. Strawn, MD,<sup>7</sup> Arjun P. Athreya, MS, PhD,<sup>1,\*</sup> and Paul E. Croarkin, DO, MS<sup>2,\*</sup>

## Abstract

**Introduction:** The Clinical Global Impressions-Improvement (CGI-I) scale is widely used in clinical research to assess symptoms and functioning in the context of treatment. The correlates of the CGI-I with efficacy scales for adolescent major depressive disorder are poorly understood. This study focused on benchmarking CGI-I scores with changes in the Children's Depression Rating Scale-Revised (CDRS-R) and the Quick Inventory of Depressive Symptomatology-Adolescent (17-item) Self-Report (QIDS-A17-SR).

**Methods:** We examined three datasets with the clinician-rated CDRS-R to ascertain equivalent percent changes in total scores and CGI-I ratings. Exploratory analyses examined corresponding percentage changes in the QIDS-A17-SR and the CGI-I ratings. The CGI-I was the reference scale for nonparametric equipercentile linking with the Equate package in R.

**Results:** CGI-I scores of 1 mapped to ≥78%–95% change in CDRS-R scores at 4–6 weeks across three datasets. CGI-I scores of 2 mapped to 56%–94% change in CDRS-R scores at 4–6 weeks across three studies. CGI-I scores of 3 mapped to 30%–68% changes in CDRS-R scores at 4–6 weeks across three studies. CGI-I scores of 4 mapped to a range of 29%–44% at 4–6 weeks across three studies. There was no significant difference ( $p \geq 0.6$ ) between treatment groups in both the Treatment of Adolescents with Depression and Treatment of Resistant Depression in Adolescents studies, for each CGI-I score ( $= 1$ , or  $= 2$  or  $= 3$ , or  $\geq 4$ ), associated mapping of total depression severity score, or associated percent change from baseline for corresponding follow-up visits. There was no significant sex difference ( $p > 0.2$ ) in CGI-I linkages to CDRS-R total or percentage changes.

**Conclusions:** These findings establish clear relationships among CGI-I scores and the CDRS-R and the QIDS-A17-SR. These benchmarks have utility for clinical trial study design, inter-rater reliability training, and clinical implementation.

**Keywords:** AMOD, equipercentile linking, response, remission, TADS, TORDIA

<sup>1</sup>Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota, USA.

<sup>2</sup>Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota, USA.

<sup>3</sup>Department of Economics, University of Cincinnati, Cincinnati, Ohio, USA.

<sup>4</sup>Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas, USA.

<sup>5</sup>Children's Health, Children's Medical Center, Dallas, Texas, USA.

<sup>6</sup>Department of Psychiatry and Psychology, Mayo Clinic, Jacksonville, Florida, USA.

<sup>7</sup>Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati, Cincinnati, Ohio, USA.

**Funding:** This work was, in part, supported by the Mayo Clinic's Summer Undergraduate Research Fellowship, Mayo Clinic Foundation for Medical Education and Research, the National Science Foundation under award number 2041339, and the National Institutes Health under award numbers R01MH113700, R01MH124655, and R01AA027486. The sources of funding for this study had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the article; and in the decision to publish the results.

\*These authors contributed equally to this work.

## Introduction

THE CLINICAL GLOBAL IMPRESSION SCALE (CGI) is an archetypal rating scale in clinical trials for the global assessment of transdiagnostic pathology (Guy 1976). This clinician-rated, seven-point scale is used to assess the overall impairment and severity of a psychiatric diagnosis with the Clinical Global Impression Severity scale (CGI-S), while the Clinical Global Impression-Improvement scale (CGI-I) measures illness improvement relative to a pretreatment baseline. Despite decades of widespread use in clinical research and practice, the psychometric properties of the CGI are not well understood in the context of child and adolescent psychopharmacology research. There is some controversy about the utility of the CGI in clinical research and practice (Kadouri et al. 2007; de Beurs et al. 2019).

Reliable, valid, and efficient outcome measures that assess symptom severity and change in major depressive disorder (MDD) or treatment-resistant depression (TRD) are an ongoing unmet need for the practice of child and adolescent psychopharmacology. An ideal assessment tool would be acceptable for both clinical and research environments. This need is particularly important for the treatment of adolescents with MDD. Prior work demonstrates the limitations of adapting widely used assessments from clinical practice to clinical research (Na et al. 2018; Nandakumar et al. 2019).

Furthermore, the clinical adaptation of standard clinical research assessments such as the Children's Depression Rating Scale-Revised (CDRS-R) is implausible (Poznanski et al. 1984; Richardson et al. 2010). While definitions of remission are clear, there are historical (Emslie et al. 1997) and contemporary (Athreya et al. 2022) challenges in defining response to treatment with the CDRS-R. Response is typically characterized based on a percentage change improvement in CDRS-R symptoms.

Prior studies of adults with MDD characterized thresholds of clinically meaningful improvement in the 17-item Hamilton Depression Rating Scale (HDRS-17) with CGI-I scores through equipercentile linking (Hamilton 1960; Bobo et al. 2016). The strength of this approach is that it links a symptom severity measure consistently used as a primary outcome (HDRS-17) with a more practical, secondary outcome measure (CGI-I) from clinical trials for MDD. The findings validated standard consensus definition of categorical outcomes and advanced the understanding of both scales for future research efforts and clinical practice considerations (Bobo et al. 2016). The CGI scales are traditional secondary outcome measures in child and adolescent psychopharmacology research, but surprisingly, few studies have characterized, validated, or used CGI measures as a primary outcome (Mayes et al. 2010; Strawn et al. 2017).

This study sought to characterize the CGI-I scale in the context of standard primary outcome measures for clinical trials for MDD in adolescents. We examined existing datasets from the Treatment for Adolescents with Depression Study (TADS) (TADS 2009), Adolescent Management of Depression (AMOD) (Vande Voort et al. 2021), and the Treatment of SSRI-Resistant Depression in Adolescents (TORDIA) (Brent et al. 2008) studies to benchmark CGI-I scores with changes in CDRS-R and the Quick Inventory of Depressive Symptomatology-Adolescent (17-item) Self-Report (QIDS-A17-SR) (Bernstein et al. 2010).

There are various options to operationalize outcomes with the CDRS-R to include using posttest scores, change scores, residual change scores, and percentage change scores. Clinical definitions of antidepressant response with the CDRS-R are often oper-

ationalized based on percentage change thresholds from the baseline score (as a relative change). This approach does have limitations as prior work has demonstrated 100×logarithmic ratio of two measurements is more informative of relative change (Törnqvist et al. 1985; Vickers et al. 2001). Other work suggests that follow-up clinical scores and percentage change scores are superior to change scores in describing clinical outcomes (Austevoll et al. 2019).

In this study, considering that CGI-I measures relative change in depressive severity, corresponding relative change in the CDRS-R and QID-A17-SR was operationalized as percentage changes rather than raw changes with a broad goal of informing definitions of response for future research. We hypothesized that the percentage changes in the CDRS-R continuous outcome measures corresponding with CGI-I thresholds would be consistent among studies. It was also anticipated that there would be no sex difference among CDRS-R and CGI-I correlates. It was further anticipated that similar CGI-I benchmarks could be identified with the QIDS-A17-SR.

## Materials and Methods

### *Rating scales and clinical outcomes*

The CDRS-R, a 17-item clinician-administered rating scale, was originally developed for assessing depressive symptoms in children 6–12 years of age (Mayes et al. 2010). The CDRS-R total scores range from 17 to 113. A score of  $\geq 40$  is consistent with active depression, a score  $\leq 28$  has been used to define remission (minimal or no symptoms) in response to treatment, and an improvement in CDRS-R total score  $\geq 50\%$  from baseline is defined as response (Tao et al. 2009).

The QIDS-A17-SR is a 17-item self-report rating scale, which yields total scores from 0 to 27 (Bernstein et al. 2010). In response to treatment, QIDS-A17-SR  $\leq 5$  defines remission, and an improvement in QIDS-A17-SR total score  $\geq 50\%$  from baseline corresponds to response. In addition, QIDS-A17-SR total scores of ranges 6 to 10, 11 to 15, 16 to 20, and  $\geq 21$  are generally considered to represent mild, moderate, severe, and very severe depressive symptom severity, respectively (Bernstein et al. 2010). The QIDS-A17-SR was only available in the AMOD dataset.

The CGI-I is a one-item seven-point scale wherein a clinician within the context of their clinical experience makes a global judgment about the improvement in disease severity from initiation of treatment (Guy 1976; Busner and Targum 2007). The CGI-I ratings are in response to “rate the overall improvement in patient since the beginning of treatment,” 1=very much improved, 2=much improved, 3=minimally improved, 4=no change, 5=minimally worse, 6=much worse, and 7=very much worse.

The CDRS-R and CGI-I measures were completed by blinded, independent raters in all three studies.

### *Data sources*

The data sources are described by the depression rating scales (CDRS-R or QIDS-A17-SR) and relevant characteristics are tabulated in Table 1. All studies considered in this work have been previously published and were approved by Institutional Review Boards (IRBs) of performance sites. All participants and their parents gave written informed assent and consent in accordance with local IRB regulations (Brent et al. 2008; TADS 2009; Vande Voort et al. 2021). Data from participants who completed treatment in each study (TADS, AMOD, and TORDIA) were analyzed for

TABLE 1. SAMPLE CHARACTERISTICS

Variables	Study datasets		
	TADS	AMOD	TORDIA
Total (N)	353	146	262
Mean age in years (standard deviation)	14.53 (1.59)	15.38 (1.49)	15.76 (1.5)
Sex (male, female)	M: 149 (42%); F: 204 (58%)	M: 31 (21%); F: 115 (79%)	M: 76 (29%); F: 186 (71%)
Race, <i>n</i> (%)			
White	266 (75.4%)	126 (86.3%)	229 (87.4%)
Black	40 (11.3%)	1 (0.7%)	7 (2.7%)
Hispanic	31 (8.8%)	0 (0%)	0 (0%)
Asian	3 (0.8%)	6 (4.1%)	6 (2.3%)
American Indian	2 (0.6%)	0 (0%)	2 (0.8%)
Multiple	11 (3.1%)	0 (0%)	17 (6.5%)
Other	0 (0%)	12 (8.2%)	0 (0%)
Unknown	0 (0%)	1 (0.7%)	1 (0.3%)
Treatment (s)	Fluoxetine: 92; CBT: 85; Fluoxetine with CBT: 88; Placebo: 88	Assignments based on clinical judgment or pharmacogenetic testing	SSRI only: 60; SSRI+CBT: 41; SNRI only: 81; SNRI+CBT: 80
Rating Scale(s)	CDRS-R	CDRS-R; QIDS-SR	CDRS-R
Baseline depression severity (median)	45:98 (59)	CDRS-R: 41:82 (58) QIDS-SR: 5:24 (15)	40:102 (61)
Time points for assessing treatment response	6 and 12 weeks	4 and 8 weeks	4, 8, and 12 weeks

AMOD, Adolescent Management of Depression; CBT, Cognitive Behavior Therapy; CDRS-R, Children's Depression Rating Scale-Revised; SNRI, Selective Norepinephrine Reuptake Inhibitor; SSRI, Selective Serotonin Reuptake Inhibitor; TADS, Treatment for Adolescents with Depression Study; TORDIA, Treatment of SSRI-Resistant Depression in Adolescents.

this study, resulting in smaller sample sizes than what have been documented in each parent study.

Exploratory analyses to study the linking of CGI-I scores to depression symptom severity on a self-reported scale (e.g., QIDS-A17-SR) were conducted on data from AMOD study. The QIDS-A17-SR and CGI-I were assessed at baseline and 4 and 8 weeks.

### Statistical analyses

Equipercntile links scores (e.g., total score or percent change) on a given scale (e.g., CDRS-R or QIDS-A17-SR) to the scores on a reference scale (CGI-I) that has the same percentile rank, under the assumption that the two scales are correlated. In this study, CGI-I was the reference scale as it is rated based on the improvement in disease state as a global opinion based on experience and overall assessment of the patient. Equipercntile linking is a nonparametric method in that it does not assume any specific underlying distribution of data (i.e., total scores, percent change or CGI-I scores) and has shown to tolerate measurement artifacts [i.e., rating drift or inter-rater variability (Kolen and Brennan 2014; Bobo et al. 2016)]. The analyses proceeded in two steps. Data for this study were analyzed using the Equate package in R (Albano 2016).

First, we generated Spearman correlations between percent change and absolute total scores on CDRS-R or QIDS-A17-SR scales with CGI-I scores. Spearman correlations were used as it checks for monotonic relationships between two variables. Second, if the correlations were significant (Spearman correlation with  $p < 0.05$ ), then equipercntile linkage was used to link percent change of total scores on CDRS-R or QIDS-A17-SR to CGI-I scores derived at each time point in the clinical trials. For each time point (e.g., 4, 6, 8, or 12 weeks) and for a given depression symptom

severity rating scale (CDRS-R or QIDS-A17-SR), a CGI-I score was mapped to a corresponding (1) percent change in total depression score from baseline and (2) total depression score.

The mappings (Figs. 1 and 2) were then plotted using average smoothing technique, with dashed lines mapping CGI-I (y-axis) to corresponding scores on CDRS-R or QIDS-A17-SR (x-axis). We also report the mapping of scores CGI-I = 1, 2, 3, and  $\geq 4$  (due to limited sample sizes with CGI-I = 5, 6, or 7) to ranges of percent change and absolute total scores on CDRS-R or QIDS-A17-SR.

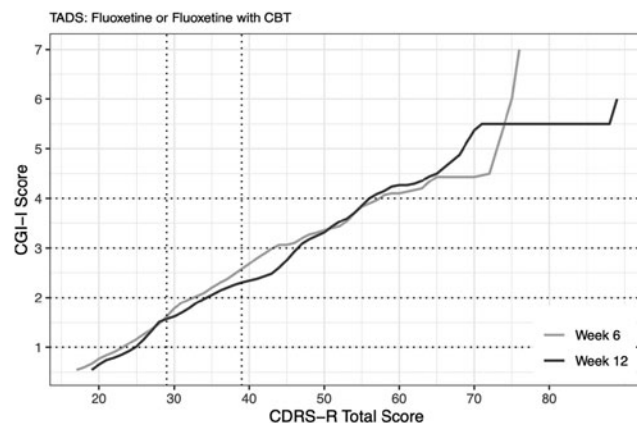
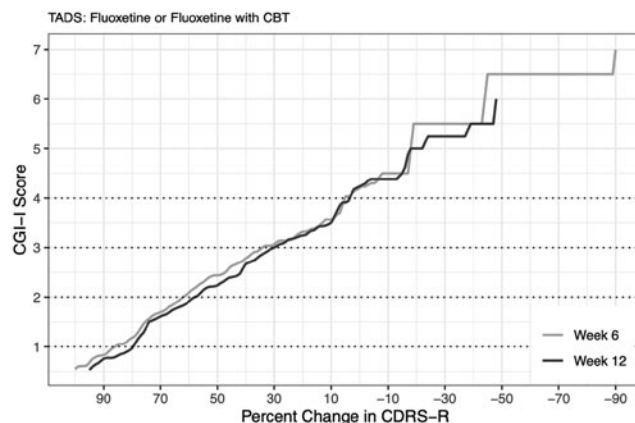
We also conducted additional sensitivity analyses. First, equipercntile linking was compared across treatment arms. For each study with multiple treatment arms with CDRS-R data and each CGI-I value (=1, =2, =3, and  $\geq 4$ ), Kolmogorov-Smirnov tests (KS-tests) were used to compare differences in percent change in depression severity or the total depression severity score itself at each follow-up time point, respectively.

The equipercntile linking among patients with MDD and TRD was compared after 12 weeks of pharmacotherapy. As both TADS and TORDIA were studies with acute-phase endpoints of 12 weeks, if the range of baseline depression severity was comparable ( $p$ -value of KS-test  $> 0.05$ ), then KS-tests were used to compare differences in percent change in depression severity or the total depression severity score.

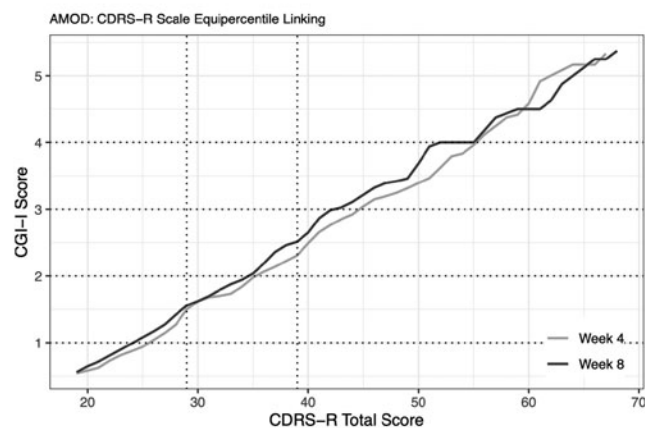
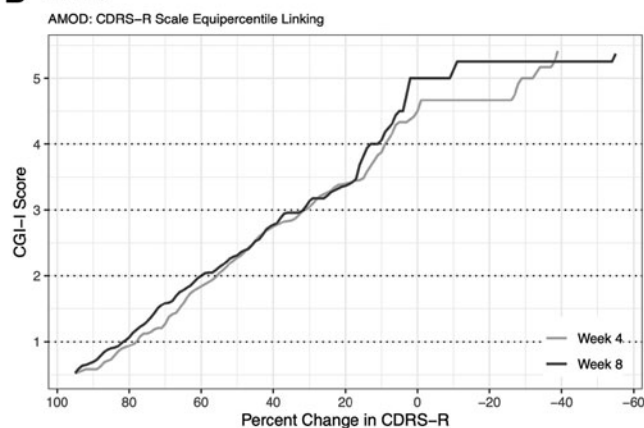
As both TADS and TORDIA studies had sufficient sex representation (sex with low samples  $> 1/3$ rd of study size), with each study, KS-tests were used to compare differences in percent change in depression severity or the total depression severity score at follow-up time points.

To avoid effects of high or low depression severity at baseline, we repeated the equipercntile linkage of CGI-I and total depression severity score and associated percent change from baseline

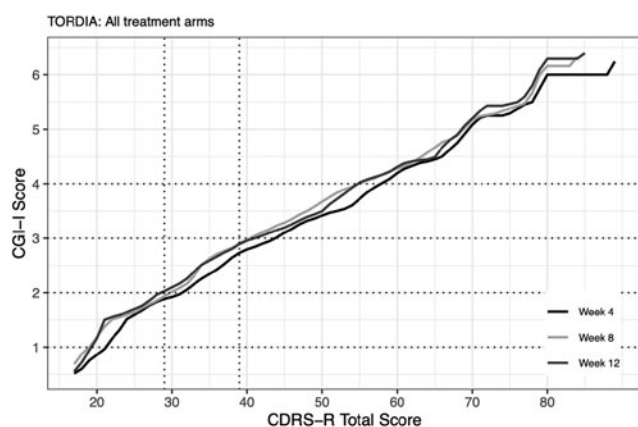
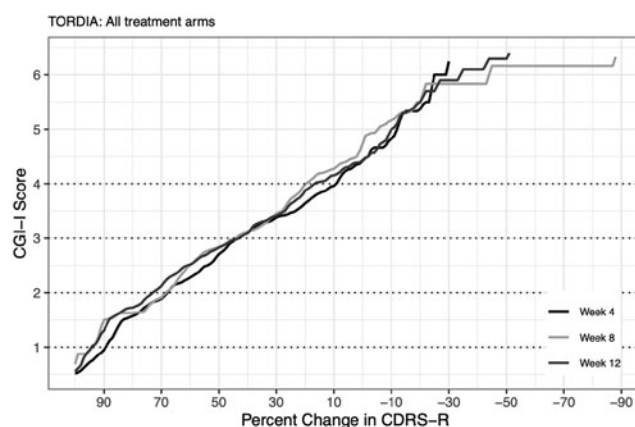
### A TADS (fluoxetine or fluoxetine with CBT)



### B AMOD

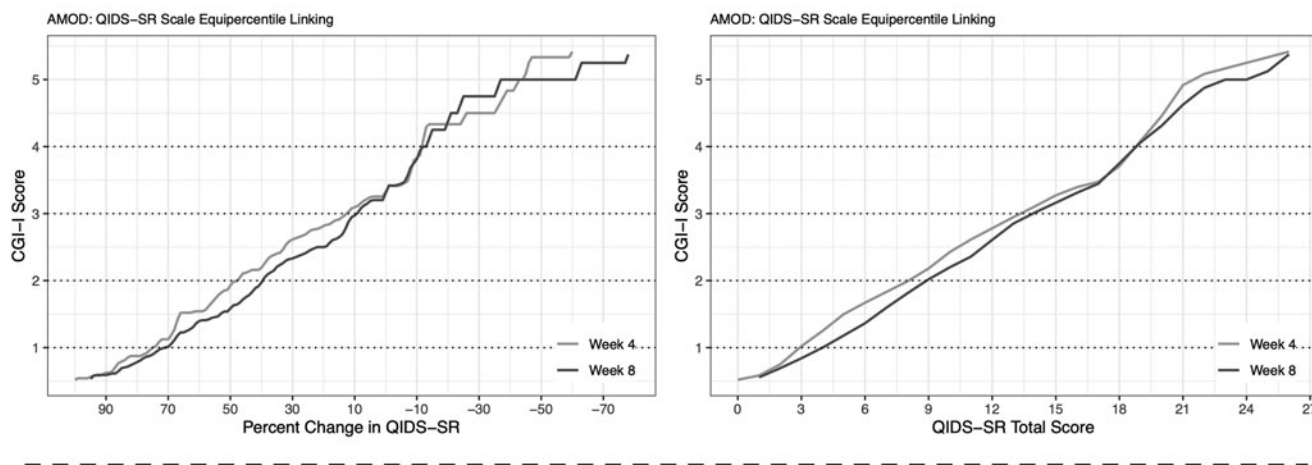


### C TORDIA (all arms)

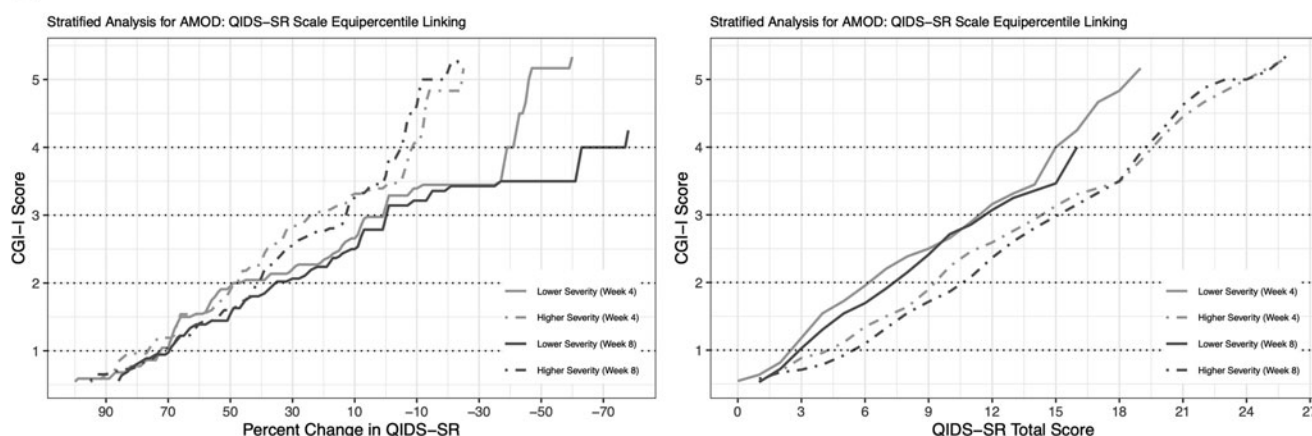


**FIG. 1.** Equipercntile linkage based on concordance between CGI-I score and percent change in CDRS-R total score from baseline, and CDRS-R total score at follow-up time points in (A) TADS (patients receiving fluoxetine or fluoxetine+CBT), (B) AMOD, and (C) TORDIA, including patients who received CBT+SSRI, CBT+SNRI, SNRI, and SSRI. AMOD, Adolescent Management of Depression; CBT, cognitive behavioral therapy; CDRS-R, Children's Depression Rating Scale-Revised; CGI-I, Clinical Global Impression-Improvement; TADS, Treatment for Adolescents with Depression Study; TORDIA, Treatment of SSRI-Resistant Depression in Adolescents.

## A AMOD: Unstratified analyses



## B AMOD: Stratified analyses



**FIG. 2.** Equipercetile linking plotted as lines based on concordance between CGI-I (CGI-I and percent change in QIDS-A17-SR total score from baseline) and CDRS-R total score at follow-up time points in AMOD study, wherein in (A), patients are not stratified, and in (B), patients are stratified by median split. AMOD, Adolescent Management of Depression; CDRS-R, Children's Depression Rating Scale-Revised; CGI-I, Clinical Global Impression-Improvement; QIDS-A17-SR, Quick Inventory of Depressive Symptomatology-Adolescent (17-item) Self-Report.

based on two depression severity groups split by median depression severity scores (CDRS-R or QIDS-A17-SR) at baseline (higher severity group: patients with baseline depression severity  $\geq$  median depression severity of cohort; and conversely for lower severity group).

Finally, the equipercetile linking of CGI-I to CDRS-R = 29 and 39 at follow-up visits was examined. A CDRS-R total score of  $\leq 28$  or  $\geq 40$  is considered remission or in active depression during follow-up visits; we sought the linking of CDRS-R total score of 29 and 39 to a score on CGI-I scale.

## Results

For each time point after baseline and in each rating scale, the Spearman correlations between CGI-I and percent change in total depression severity scores or total depression severity scores were significant ( $p \leq 0.05$ , Supplementary Table S1). Given these significant correlations between measures of both rating scales and CGI-I, subsequent equipercetile linkage analyses were conducted.

### Equipercetile linking of CGI-I with CDRS-R

In TADS, the following are the equipercetile linkages for CGI-I = 1, 2, 3, and  $\geq 4$  (Fig. 1A and "A" in Table 2): a CGI-I of 1 mapped to  $\geq 80\%$  (at 6 weeks) and  $\geq 87\%$  (at 12 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of  $\leq 24$  (at 6 weeks) and  $\leq 23$  (at 12 weeks).

A CGI-I of 2 mapped to 58%–79% (at 6 weeks) and 63%–86% (at 12 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of 25–34 (at 6 weeks) and 24–32 (at 12 weeks). A CGI-I of 3 mapped to 30%–57% (at 6 weeks) and 35%–62% (at 12 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of 35–46 (at 6 weeks) and 33–43 (at 12 weeks). A CGI-I of  $\geq 4$  mapped to  $\leq 29\%$  (at 6 weeks) and  $\leq 34\%$  (at 12 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of  $\geq 47$  (at 6 weeks) and  $\geq 44$  (at 12 weeks).

In AMOD, the following are the equipercetile linkages for CGI-I = 1, 2, 3, and  $\geq 4$  (Fig. 1B and "A" in Table 2): a CGI-I of 1 mapped to  $\geq 78\%$  (at 4 weeks) and  $\geq 82\%$  (at 8 weeks) change in

TABLE 2. EQUIPERCENTILE LINKING CGI-I WITH CDRS-R AND QIDS-A17-SR SCORES

Study/treatment arms	CGI-I	Linking to CDRS-R total score range	Linking to percent change in CDRS-R total score	Linking to CDRS-R total score range	Linking to percent change in CDRS-R total score	Linking to CDRS-R total score range	Linking to percent change in CDRS-R total score
A. CDRS-R							
		6 Weeks		12 Weeks			
TADS (FLX, or CBT with FLX)	1	≤24	≥80	≤23	≥87		
	2	25–34	58–79	24–32	63–86		
	3	35–46	30–57	33–43	35–62		
	≥4	≥47	≤29	≥44	≤34		
		4 Weeks		8 Weeks			
AMOD (All patients)	1	≤26	≥78	≤24	≥82		
	2	27–35	56–77	25–34	60–81		
	3	36–44	31–55	35–42	33–59		
	≥4	≥45	≤30	≥43	≤32		
		4 Weeks		8 Weeks		12 Weeks	
TORDIA (All patients)	1	≤18	≥95	≤19	≥95	≤21	≥90
	2	19–29	69–94	20–28	73–94	22–31	68–89
	3	30–40	45–68	29–41	43–72	32–43	44–67
	≥4	≥41	≤44	≥42	≤42	≥44	≤43
Study/treatment arms	CGI-I	Linking to QIDS-SR total score range	Linking to percent change in QIDS-SR total score	Linking to QIDS-SR total score range	Linking to percent change in QIDS-SR total score		
B. QIDS-A17-SR							
		4 Weeks		8 Weeks			
AMOD	1	≤3	≥74	≤4	≥71		
	2	4–8	48–73	5–8	40–70		
	3	9–13	13–47	9–13	10–39		
	≥4	≥14	≤12	≥14	≤9		

AMOD, Adolescent Management of Depression; CBT, Cognitive Behavior Therapy; CGI-I, Clinical Global Severity-Improvement; CDRS-R, Children Depression Rating Scale-Revised; FLX, Fluoxetine; QIDS-A17-SR, Quick Inventory of Depressive Symptomatology-Adolescent (17-item) Self-Reported; SNRI, Selective Norepinephrine Reuptake Inhibitor; SSRI, Selective Serotonin Reuptake Inhibitor; TADS, Treatment for Adolescents with Depression Study; TORDIA, Treatment of SSRI-Resistant Depression in Adolescents.

CDRS-R total score from baseline and CDRS-R total scores of ≤26 (at 4 weeks) and ≤24 (at 8 weeks). A CGI-I of 2 mapped to 56%–77% (at 4 weeks) and 60%–81% (at 8 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of 27–35 (at 4 weeks) and 25–34 (at 8 weeks).

A CGI-I of 3 mapped to 31%–55% (at 4 weeks) and 33%–59% (at 8 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of 36–44 (at 4 weeks) and 35–42 (at 8 weeks). A CGI-I of ≥4 mapped to ≤30% (at 4 weeks) and ≤32% (at 8 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of ≥45 (at 4 weeks) and ≥43 (at 8 weeks).

In TORDIA (in patients across all arms), the following are the equipercetile linkages for CGI-I=1, 2, 3, and ≥4 (Fig. 1C and “A” in Table 2). A CGI-I of 1 mapped to ≥95% (at 4 weeks), ≥95% (at 8 weeks), and ≥90% (at 12 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of ≤18 (at 4 weeks), ≤19 (at 8 weeks), and ≤21 (at 12 weeks). A CGI-I of 2 mapped to 69%–94% (at 4 weeks), 73%–94% (at 8 weeks), and 68%–89% (at 12 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of 19–29 (at 4 weeks), 20–28 (at 8 weeks), and 22–31 (at 12 weeks).

A CGI-I of 3 mapped to 45%–68% (at 4 weeks), 43%–72% (at 8 weeks), and 44%–67% (at 12 weeks) change in CDRS-R total score

from baseline and CDRS-R total scores of 30–40 (at 4 weeks), 29–41 (at 8 weeks), and 32–43 (at 12 weeks). A CGI-I of ≥4 mapped to ≤44% (at 4 weeks), ≤42% (at 8 weeks), and ≤43% (at 12 weeks) change in CDRS-R total score from baseline and CDRS-R total scores of ≥41 (at 4 weeks), ≥42 (at 8 weeks), and ≥44 (at 12 weeks).

#### *Within-arm (in TADS) linkages of CGI-I and corresponding CDRS-R percent change and range of scores*

There were no significant differences ( $p \geq 0.6$ ) between treatment groups in both TADS and TORDIA studies and for each CGI-I score (=1, or=2, or=3, or=4) and associated mapping of total depression severity score or associated percent change from baseline for corresponding follow-up visits.

#### *Comparing equipercetile linking of CGI-I to CDRS-R between MDD (no TRD) and TRD patients after 12 weeks of pharmacotherapy*

Distributions of baseline CDRS-R total scores between TADS and TORDIA patients were not statistically different (KS-test,  $p=0.12$ ). For all CGI-I=1, 2, 3, and ≥4, the respective ranges of CDRS-R depression severity at baseline and 12 weeks were not

statistically different ( $p \geq 0.15$ ) between TADS and TORDIA patients. For CGI-I=1 and  $\geq 4$ , the distribution of linked percent change in CDRS-R total score from baseline was not statistically significant ( $p \geq 0.49$ ). For CGI-I=2 and 3, the associated percent change in CDRS-R total score from baseline was statistically different ( $p \leq 0.02$ ). The median percent change was 33% (TADS) and 51% (TORDIA) for CGI-I=2, and 23% (TADS) and 35% (TORDIA) for CGI-I=3.

### *Equipercetile linking of CGI-I with QIDS-A17-SR*

The following are equipercetile linkages for CGI-I=1, 2, 3, and  $\geq 4$  ("B" in Table 2) in AMOD study. A CGI-I of 1 mapped to  $\geq 74\%$  (at 4 weeks) and  $\geq 71\%$  (at 8 weeks) change in QIDS-A17-SR total score from baseline and QIDS-A17-SR total scores of  $\leq 3$  (at 4 weeks) and  $\leq 4$  (at 8 weeks). CGI-I of 2 mapped to 48%–73% (at 4 weeks) and 40%–70% (at 8 weeks) change in QIDS-A17-SR total score from baseline and QIDS-A17-SR total scores of 4–8 (at 4 weeks) and 5–8 (at 8 weeks).

CGI-I of 3 mapped to 13%–47% (at 4 weeks) and 10%–39% (at 8 weeks) change in QIDS-A17-SR total score from baseline and QIDS-A17-SR total scores of 9–13 (at 4 and 8 weeks). CGI-I of  $\geq 4$  mapped to  $\leq 12\%$  (at 4 weeks) and  $\leq 9\%$  (at 8 weeks) change in QIDS-A17-SR total score from baseline and QIDS-A17-SR total scores of  $\geq 14$  (at 4 and 8 weeks).

### *Equipercetile linkage of CGI-I and depression severity rating scales in patients stratified by median split*

Across all three studies and two rating scales, there were no significant differences in linkages derived between CGI-I corresponding ranges of total scores or percent change in total scores of CDRS-R/QIDS-A17-SR from baseline (Fig. 2 and Supplementary Tables S2 and S3). Just as in the unstratified analyses, TORDIA subjects stratified by median split in baseline depression severity had increased percent change in CDRS-R total scores at 12 weeks in comparison with TADS subjects for CGI-I=2 or 3 linkages, with no such difference for CGI-I=1 or  $\geq 4$ .

### *Sex differences in CGI-I linkage to CDRS-R total scores or associated percent change*

In both TADS and TORDIA and within sex-stratified groups and respective follow-up time points, there was no statistical difference ( $p > 0.2$ ) in CGI-I linkages to ranges of CDRS-R total scores or associated percent change from baseline.

### *Equipercetile linking for CDRS-R total score of 29 and 39 to CGI-I at follow-up visits*

Across studies wherein MDD subjects received pharmacotherapy (TADS and AMOD), CDRS-R of 29 and 39 linked to CGI-I=2 and 3, respectively. For this range of depression severity, majority of subjects in TADS (patients receiving fluoxetine or fluoxetine with CBT: 88% at week 6, 77.3% at week 12), AMOD (76.6% at week 4, 81.4% at week 8) achieved CGI-I  $\leq 2$ . In TORDIA subjects, CDRS-R of 29 and 39 linked to CGI-I=2 and 3, respectively, at week 4 and 12, and CGI-I=3 and 3, respectively at week 8. For this range of depression severity in TORDIA subjects, 33.73%, 51.18%, and 61.63% achieved CGI-I  $\leq 2$  at week 4, week 8 and week 12 respectively.

## **Discussion**

This is the first and largest study to describe CGI-I thresholds in the context of changes in CDRS-R scores. Demonstrating how clinician-rated outcomes such as the CDRS-R corresponds with the CGI-I has clinical utility. This work also presents insights for planning and executing clinical trials for adolescents with depression as well as everyday utility for clinical practice. The CGI-I score appears to have consistent correlations with a continuous symptom severity measure that is a standard primary outcome measure for clinical research. The CGI is typically a standard secondary outcome measure in clinical research studies.

These findings provide a framework for reconsidering CGI-I as a primary outcome measure. This could enhance the reliability of ratings within studies and broadly would provide a strategic translation for clinical practice. Given that the CGI-I can be rated quickly, it could be acceptable for busy clinicians. The CDRS-R is not routinely used in clinical practice in light of the time burden for clinicians. Broad use of the CGI-I in clinical practice and prospective clinical studies could advance clinical care, research, and related translations.

The examination of heterogenous data sets and samples was the strength of this study. It is noteworthy that the mapping of CGI-I and percentage change in the CDRS-R scores are consistent across the TADS and AMOD studies, despite difference in study design, recruitment strategies, treatments, and sample characteristics. The AMOD study was, by design, an effectiveness study positioned within a clinical practice. The demonstrated reliability of the CGI-I mapping is reassuring and underscores the translational opportunities in implementation of the CGI-I scale. The consistent absence of sex differences also increases the utility of our findings.

There was broad consensus that CDRS  $>28$  and  $<40$  maps to CGI-I  $\leq 2$  in adolescents with MDD. However, this was not the case for adolescents with TRD. Although CGI-I to CDRS-R ranges at 12 weeks were consistent for TADS versus TORDIA, CGI-I=2 and 3 had different ranges of % improvements, despite patients having comparable baseline depression severity. In CDRS-R,  $>28$  and  $<40$  maps to CGI-I  $\leq 2$  for  $>75\%$  of patients in AMOD and TADS, but  $\leq 61\%$  in TORDIA across 4, 8, and 12 weeks. This suggests that adolescents with TRD require greater percent change in CDRS-R symptoms for CGI-I benchmarks of improvement.

It is important to emphasize that although linkages did not diverge among treatment arms of TADS and TORDIA, differences in tolerability were not considered in treatment response. Tolerability, side effects, pharmacodynamic features, and pharmacokinetics are infrequently considered prospectively in clinical pharmacology trials for adolescents (Dobson et al. 2019). Recent modeling studies demonstrate the importance of broader considerations for dosing antidepressants and defining meaningful changes in symptom severity (Poweleit et al. 2019; Strawn et al. 2019). Alternatively, future studies should consider and examine use of the CGI efficacy index (Guy 1976; Busner and Targum 2007; Busner et al. 2009).

These findings are also instructive in broadening the understanding of self-report measures such as the QIDS-A17-SR. Based on CGI-I linking to range of QIDS-A17-SR definitions at 4 or 8 weeks, CGI-I=1 mapped to remission or near absence of depressive symptoms, CGI-I=2 mapped to mild depressive symptom severity, CGI-I=3 mapped to moderate depressive symptom severity, and CGI-I  $\geq 4$  mapped to severe or very severe symptoms severity (Bernstein et al. 2010).

The findings from the stratified analyses did not identify consistent differences in CGI-I linkages among patients with high or

low symptom severity based on the CDRS-R or QIDS-A17-SR at baseline. The TORDIA subjects did have a greater percentage change in CDRS-R total scores at 12 weeks in comparison to TADS for CGI-I = 2 or 3 linkages, but not 1 or  $\geq 4$ . This is aligned with our hypothesis and suggests that raters and patients designate a greater symptom severity change for minimal and moderate improvement in the context of TRD.

Prior work focused on adolescents in treatment for depression demonstrates that a significant improvement in depressive symptoms (around 50%) at 4 weeks is a good standard of response and frequently demonstrated in patients who have remission of depressive symptoms with acute treatment (Tao et al. 2009). This also suggests that based on clinical efficacy measures of symptom severity, dose adjustments should be considered earlier during treatment than what is typical in clinical practice (Shippee et al. 2018).

This study has a number of limitations to consider in the interpretation of findings and planning for future studies. The study used data from completers only and did not assess the effects of early dropouts or noncompleters of therapy. The datasets had variability in patient samples, methodology, treatments, time points, and rater characteristics. This study did not consider tolerability or side effects. Quality of life, functional impairment, and specific symptom measurements were not considered. Given the public health impact of suicidality, individual measures of suicidality would be worth considering in future studies. The QIDS-A17-SR rating scale was only available in one study (AMOD), limiting the generalizability of associated findings in this rating scale across other treatment settings.

Improvement of the CDRS-R and QIDS-A17-SR was operationalized as percentage change with the intent of informing research focused on future definitions of response. This approach had limitations as it attempted to equate two distinct constructs (a single assessment of improvement after treatment with a percentage change between pretest and posttest severity scores). This approach likely had statistical inefficiencies (Vickers et al. 2001; Zhang et al. 2014). The use of raw change scores might have been a more valid approach. The sample size and homogenous demographics were not sufficient for sensitivity analyses by race.

Finally, depression in children and adolescents is heterogenous and improvements related to nonspecific factors may not be adequately captured by CGI-I ratings. In clinical trials, CGI-I measures are often completed in conjunction with CDRS-R scores and expectedly highly correlated. However, if the CGI-I was used alone as a primary outcome, without access to assessments of changes in specific symptoms, ratings could be more difficult to assess.

## Conclusions

In summary, this study demonstrates well-defined and consistent relationships among CGI-I scores and continuous measures of depressive symptom severity change assessed with the CDRS-R and QIDS-A17-SR. Future studies could examine the predictive validity of the CGI-I, utility as a primary outcome measure in clinical trials for MDD in adolescents, and implementation in clinical practice.

## Clinical Significance

These findings provide a compelling opportunity to advance the treatment of depression in adolescents. The CGI-I is a clinician-friendly, efficient assessment that could be easily implemented as a standard assessment tool at each patient contact. The linkage

findings suggest that the CGI-I benchmarks correspond with standard research assessments for depressive symptom severity. Implementation of the CGI-I has the prospect of advancing clinical care of adolescents, providing opportunities for practice-based research, and catalyzing the translation of research findings.

## Authors' Contributions

A.P.A., J.A.M., J.R.S., and P.E.C. conceptualized and designed the study, C.Z. and A.P.A. developed the analyses, C.Z., A.P.A., and P.E.C. developed the initial draft of the article. C.Z., J.V.V., D.Y., J.A.M., G.J.E., B.D.K., T.M., M.T., W.V.B., J.R.S., A.P.A., and P.E.C. edited and assisted with revisions of subsequent versions of the article. All authors read and approved the final version of the article.

## Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Mayo Clinic Foundation for Medical Education and Research, National Science Foundation, or the National Institutes of Health.

## Disclosures

J.L.V.V. was a co-primary investigator on an investigator-initiated study that had grant-in-kind support for supplies and genotyping from Assurex Health. J.A.M. receives research support from the Yung Family Foundation. G.J.E. receives research support from Duke University, Forest Research Institute (partner of Merck KGaA; formerly known as Forest Laboratories), and Janssen Pharmaceuticals, Research & Development.

G.J.E. is a consultant for Lundbeck, Neuronetics, and Otsuka. B.D.K. receives royalties from Guilford Press, Inc., M.T. has provided consulting services to Acadia Pharmaceuticals, Inc., Alkermes, Inc., Allergan, Inc., Alto Neuroscience, Inc., Applied Clinical Intelligence, Axsome Therapeutics, Boehringer Ingelheim, Engage Health Media, GH Research Limited, GreenLight VitalSign6, Inc., Health Care Global Village, Janssen, Merck Sharp & Dohme Corp., Myriad Neuroscience, Navitor Pharmaceutical, Inc., Neurocrine Biosciences, Inc., Orexo US, Inc., Otsuka, Perception Neuroscience, Pharmerit International, SAGE Therapeutics, Signant Health, and Titan Pharmaceuticals, Inc. He has received grant/research funding from NIMH, NIDA, Patient-Centered Outcomes Research Institute (PCORI), and Cancer Prevention Research Institute of Texas (CPRIT).

In addition, he has received editorial compensation from Oxford University Press. W.V.B.'s research has been supported by the NIMH, AHRQ, NSF, Mayo Foundation for Medical Education and Research, and the Myocarditis Foundation; he has contributed chapters to UpToDate on the treatment of bipolar disorders. J.R.S. has received research support from Edgemont, Shire, Forest Research Institute, Otsuka, the Yung Family Foundation, and the National Institutes of Health (NICHD, NIMH, and NIEHS). He receives royalties from Springer Publishing for two texts and has received material support from Myriad genetics and honoraria from CMEology and Neuroscience Educational Institute. He provides consultation to the U.S. Food and Drug Administration as a Special Government Employee. A.A. receives research support from the Mayo Foundation for Medical Education and Research and NSF.

P.E.C. has received research grant support from Mayo Foundation for Education and Research, Neuronetics, Inc.; NeoSync,



Inc.; NSF, NIMH, and Pfizer, Inc. He has received grant-in-kind (equipment support for research studies) from Assurex; MagVenture, Inc; and Neuronetics, Inc. He has served as a consultant for Engrail Therapeutics, Myriad Neuroscience, Procter & Gamble, and Sunovion. The other authors have no disclosure or potential conflict of interests.

## Supplementary Material

Supplementary Table S1  
Supplementary Table S2  
Supplementary Table S3

## References

- Albano AD: Equate: Observed-score linking and equating in R. *Appl Psychol Meas* 40:361–362, 2016.
- Athreya AP, Vande Voort JL, Shekunov J, Rackley SJ, Leffler JM, McKean AJ, Romanowicz M, Kennard BD, Emslie GJ, Mayes T, Trivedi M, Wang L, Weinshilboum RM, Bobo WV, Croarkin PE: Evidence for machine learning guided early prediction of acute outcomes in the treatment of depressed children and adolescents with antidepressants. *J Child Psychol Psychiatry* 2022. [Epub ahead of print]; DOI: 10.1111/jcpp.13580.
- Austevoll IM, Gjestad R, Grotle M, Solberg T, Brox JJ, Hermansen E, Rekeland F, Indrekvam K, Storheim K, Hellum C: Follow-up score, change score or percentage change score for determining clinical important outcome following surgery? An observational study from the Norwegian registry for Spine surgery evaluating patient reported outcome measures in lumbar spinal stenosis and lumbar degenerative spondylolisthesis. *BMC Musculoskelet Disord* 20:31, 2019.
- Bernstein IH, Rush AJ, Trivedi MH, Hughes CW, Macleod L, Witte BP, Jain S, Mayes TL, Emslie GJ: Psychometric properties of the Quick Inventory of Depressive Symptomatology in adolescents. *Int J Methods Psychiatr Res* 19:185–194, 2010.
- Bobo WV, Anglerio GC, Jenkins G, Hall-Flavin DK, Weinshilboum R, Biernacka JM: Validation of the 17-item Hamilton Depression Rating Scale definition of response for adults with major depressive disorder using equipercentile linking to Clinical Global Impression scale ratings: Analysis of Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) data. *Hum Psychopharmacol* 31:185–192, 2016.
- Brent D, Emslie G, Clarke G, Wagner KD, Asarnow JR, Keller M, Vitiello B, Ritz L, Iyengar S, Abebe K, Birmaher B, Ryan N, Kennard B, Hughes C, DeBar L, McCracken J, Strober M, Suddath R, Spirito A, Leonard H, Melhem N, Porta G, Onorato M, Zelazny J: Switching to another SSRI or to venlafaxine with or without cognitive behavioral therapy for adolescents with SSRI-resistant depression: The TORDIA randomized controlled trial. *JAMA* 299:901–913, 2008.
- Busner J, Targum SD: The clinical global impressions scale: Applying a research tool in clinical practice. *Psychiatry (Edmont)* 4:28–37, 2007.
- Busner J, Targum SD, Miller DS: The clinical global impressions scale: Errors in understanding and use. *Compr Psychiatry* 50:257–262, 2009.
- de Beurs E, Carlier IVE, van Hemert AM: Approaches to denote treatment outcome: Clinical significance and clinical global impression compared. *Int J Methods Psychiatr Res* 28:e1797, 2019.
- Dobson ET, Bloch MH, Strawn JR: Efficacy and tolerability of pharmacotherapy for pediatric anxiety disorders: A network meta-analysis. *J Clin Psychiatry* 80:17r12064, 2019.
- Emslie GJ, Rush AJ, Weinberg WA, Kowatch RA, Hughes CW, Carmody T, Rintelmann J: A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Arch Gen Psychiatry* 54:1031–1037, 1997.
- Guy W: ECDEU Assessment Manual for Psychopharmacology. US Department of Health, Education, and Welfare, Public Health Service Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs, 1976.
- Hamilton M: A Rating Scale for Depression. *J Neurol Neurosurg Psychiatry* 23:56–62, 1960.
- Kadouri A, Corruble E, Falissard B: The improved Clinical Global Impression Scale (iCGI): Development and validation in depression. *BMC Psychiatry* 7:7, 2007.
- Kolen MJ, Brennan RL (eds): Test Equating, Scaling, and Linking. New York, Springer-Verlag Nature, 2014.
- Mayes TL, Bernstein IH, Haley CL, Kennard BD, Emslie GJ: Psychometric properties of the Children's Depression Rating Scale-Revised in adolescents. *J Child Adolesc Psychopharmacol* 20:513–516, 2010.
- Na PJ, Yaramala SR, Kim JA, Kim H, Goes FS, Zandi PP, Vande Voort JL, Sutor B, Croarkin P, Bobo WV: The PHQ-9 Item 9 based screening for suicide risk: A validation study of the Patient Health Questionnaire (PHQ)-9 Item 9 with the Columbia Suicide Severity Rating Scale (C-SSRS). *J Affect Disord* 232:34–40, 2018.
- Nandakumar AL, Vande Voort JL, Nakonezny PA, Orth SS, Romanowicz M, Sonmez AI, Ward JA, Rackley SJ, Huxshal JE, Croarkin PE: Psychometric Properties of the Patient Health Questionnaire-9 modified for major depressive disorder in adolescents. *J Child Adolesc Psychopharmacol* 29:34–40, 2019.
- Poweleit EA, Aldrich SL, Martin LJ, Hahn D, Strawn JR, Ramsey LB: Pharmacogenetics of sertraline tolerability and response in pediatric anxiety and depressive disorders. *J Child Adolesc Psychopharmacol* 29:348–361, 2019.
- Poznanski EO, Grossman JA, Buchsbaum Y, Banegas M, Freeman L, Gibbons R: Preliminary studies of the reliability and validity of the children's depression rating scale. *J Am Acad Child Psychiatry* 23:191–197, 1984.
- Richardson LP, McCauley E, Grossman DC, McCarty CA, Richards J, Russo JE, Rockhill C, Katon W, et al.: Evaluation of the Patient Health Questionnaire-9 Item for detecting major depression among adolescents. *Pediatrics* 126:1117–1123, 2010.
- Shippee ND, Mattson A, Brennan R, Huxsahl J, Billings ML, Williams MD: Effectiveness in regular practice of collaborative care for depression among adolescents: A retrospective cohort study. *Psychiatr Serv* 69:536–541, 2018.
- Strawn JR, Dobson ET, Mill JA, Cornwall GJ, Sakolsky D, Birmaher B, Compton SN, Piacentini J, McCracken JT, Ginsburg GS, Kendall PC, Walkup JT, Albano AM, Rynn MA: Placebo response in pediatric anxiety disorders: Results from the child/adolescent anxiety multimodal study. *J Child Adolesc Psychopharmacol* 27:501–508, 2017.
- Strawn JR, Mills JA, Croarkin PE: Switching selective serotonin reuptake inhibitors in adolescents with selective serotonin reuptake inhibitor-resistant major depressive disorder: Balancing tolerability and efficacy. *J Child Adolesc Psychopharmacol* 29:250–255, 2019.
- Tao R, Emslie GJ, Mayes T, Nakonezny PN, Kennard B, Hughes C: Early prediction of acute antidepressant treatment response and remission in pediatric major depressive disorder. *J Am Acad Child Adolesc Psychiatry* 48:71–78, 2009.
- Törnqvist L, Vartia P, Vartia YO: How should relative changes be measured? *Am Statist* 39:43–46, 1985.
- Treatment for Adolescents with Depression Study (TADS): The Treatment for Adolescents with Depression Study (TADS): Out-

- comes over 1 year of naturalistic follow-up. *Am J Psychiatry* 166: 1141–1149, 2009.
- Vande Voort JL, Orth SS, Shekunov J, Romanowicz M, Geske JR, Ward JA, Leibman NI, Frye MA, Croarkin PE: A randomized controlled trial of combinatorial pharmacogenetics testing in adolescent depression. *J Am Acad Child Adolesc Psychiatry* 61:46–55, 2022.
- Vickers AJ: The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: A simulation study. *BMC Med Res Methodol* 1:1–4, 2001.
- Zhang S, Paul J, Nantha-Aree M, Buckley N, Shahzad U, Cheng J, DeBeer J, Winemaker M, Wismer D, Punthakee D, Avram V, Thabane L: Empirical comparison of four baseline covariate ad-

justment methods in analysis of continuous outcomes in randomized controlled trials. *Clin Epidemiol* 6:227–235, 2014.

Address correspondence to:  
Paul E. Croarkin, DO, MS  
Department of Psychiatry and Psychology  
Mayo Clinic  
200 First Street SW  
Rochester, MN 55905  
USA

E-mail: croarkin.paul@mayo.edu