



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2021

How Are Discrepant Parent-Child Reports Integrated? A Case of Depressed Adolescents

Baumgartner, Noemi ; Foster, Simon ; Emery, Sophie ; Berger, Gregor ; Walitza, Susanne ; Häberling, Isabelle

DOI: <https://doi.org/10.1089/cap.2020.0116>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-195841>

Journal Article

Published Version

Originally published at:

Baumgartner, Noemi; Foster, Simon; Emery, Sophie; Berger, Gregor; Walitza, Susanne; Häberling, Isabelle (2021). How Are Discrepant Parent-Child Reports Integrated? A Case of Depressed Adolescents. *Journal of Child and Adolescent Psychopharmacology*, 31(4):279-287.

DOI: <https://doi.org/10.1089/cap.2020.0116>

How Are Discrepant Parent–Child Reports Integrated? A Case of Depressed Adolescents

Noemi Baumgartner, PhD,¹ Simon Foster, PhD,¹ Sophie Emery, MSc,¹ Gregor Berger, Prof,¹
Susanne Walitza, Prof,^{1–3} and Isabelle Häberling, PhD¹

Abstract

Objective: Severity ratings of psychopathology in minors are often based on a composite score of the parent’s and child’s reports. However, parent’s and child’s reports often differ substantially, resulting in the integration method affecting the final scores. Nevertheless, effects of integration algorithms are seldom assessed and poorly understood.

Method: The dataset is derived from the Treatment for Adolescents with Depression Study (TADS) and consists of 439 adolescents, 54% female, with a Major Depressive Disorder. The interviewer conducted the clinical interview Children’s Depression Rating Scale-Revised (CDRS-R) with the parent and the adolescent and the TADS manual advised the interviewer to use the higher score as the final rating unless an informant was judged to be unreliable. Polynomial regressions, multivariate analyses, and mixed models were used to analyze the effects of this integration algorithm on the final scores and associated factors.

Results: In 77% of the cases, the interviewer followed the TADS rating rule to use the higher CDRS-R item score. However, the final item scores differed significantly from the rule using the higher value, with the higher score being less often adapted at follow-up assessments and in female patients.

Conclusions: The algorithm used to integrate divergent reports affects study outcomes and might introduce data-specific biases. Judgments of the validity and reliability of informants compromise the objectivity of outcomes in major clinical trials by introducing a subjective bias. Therefore, the agreement between children’s and parent’s reports and the method of integration should routinely be reported in research on pediatric psychopathology. ClinicalTrials.gov NCT00006286.

Keywords: TADS, interviewer, agreement, integration, CDRS-R, clinical trials

Introduction

THE FINDING OF a poor agreement between child’s and parent’s reports in assessing psychopathology has been extensively replicated and counts among the most robust findings in clinical child research (De Los Reyes and Kazdin 2005). Comparatively, less attention has been paid on how the interviewer, be that a clinician or a member of a research team, integrates the two reports into a final diagnosis or rating, and whether the interviewer agrees more with the parent or more with the child (De Los Reyes 2013). Overall, a better agreement has been observed between interviewer–parent for treatment targets (Hawley and Weisz 2003), for functional impairment (Kramer et al. 2004), for social functioning (De Los Reyes et al. 2011), and for diagnosis of anxiety disorders (Grills and Ollendick 2003; Storch et al. 2012; Hamblin et al. 2016)

compared with interviewer–adolescent agreement. In contrast though, the interviewer’s rating was more closely aligned to the child’s reports for other behaviors, such as for family and environment problems (Hawley and Weisz 2003) or illegal behavior (Kramer et al. 2004).

These studies investigated whether the interviewer’s rating agreed more with the child’s or parent’s reports rather than how the clinician integrates discrepant reports. Some authors have proposed models, guidelines, and suggestions for interpreting and integrating discrepant reports (Smith 2007; De Los Reyes et al. 2013). For example, Smith (2007) advises to give more weight to the parent’s report for externalizing problems in younger children and to the child’s report for internalizing problems in older children. Accordingly, due to the internalizing nature of depressive symptoms, in depressed adolescents the interviewer should rely more on the adolescents’ than the

¹Department of Child and Adolescent Psychiatry and Psychotherapy, Psychiatric University Hospital Zurich, University of Zurich, Zurich, Switzerland.

²Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland.

³Center for Integrative Human Physiology Zurich, University of Zurich, Zurich, Switzerland.

Funding: The Treatment for Adolescents with Depression Study study was supported by the American National Institute for Mental Health (RFP-NIH-NIMH 98-DS-0008).

parents' rating. Another strategy, termed the best-estimation method, allows the interviewers to integrate all information and to use their best judgment to dissolve conflicting reports (Klein et al. 2005). Furthermore, the "or" rule describes a way of integration in which a criteria is fulfilled as soon as it is endorsed by either the parent or the child, while according to the "and" rule a symptom criteria is only met when it is endorsed by both parties (Bird et al. 1992; Piacentini et al. 1992). Accordingly, the "or" rule involves adapting the higher score, whereas the "and" rule opts for the lower score.

The use of multiple informants is the standard in clinical practice and research for the evaluation of psychopathology in minors, including depression (Klein et al. 2005). However, given the low correspondence between different informants, prevalence rates and treatment evaluations may depend on the source on which the information is based on. For example, a meta-analysis investigating the effect of psychotherapy in depressed youth observed effect sizes to be three times larger based on the child's compared with the parent's report (Weisz et al. 2006). Moreover, prevalence rates of affective disorders are fourfold increased based on the adolescent's compared with the parent's report (Steinhausen and Winkler Metzke 2003).

In clinical trials, the primary outcome, measuring the effectiveness of an intervention, is often based on a composite score out of the parent's and child's reports (Emslie et al. 1997; March et al. 2004; Compton et al. 2010). A composite score, although, requires the integration of diverging information, with the strategy being used affecting the final scores. The "or" rule, for example, may lead to inflated prevalence rates and false positive identifications (Piacentini et al. 1992). In a recent meta-analysis on the worldwide prevalence rates of attention-deficit/hyperactivity disorder diagnoses, prevalence rates were twice as high when applying the "or" algorithm instead of the best-estimation procedure (Polanczyk et al. 2014).

The semistructured clinical interview Children's Depression Rating Scale-Revised (CDRS-R) (Poznanski and Mokros 1996) is widely used in pediatric depression research, including in the "Treatment for Adolescents with Depression Study (TADS)" (March et al. 2004), one of the largest clinical trials on therapeutic interventions in depressed adolescents. According to the manual and the study guidelines, the use of the "or" algorithm for solving conflicting information is recommended, as long as both informants are judged to be reliable and valid (Poznanski and Mokros 1996). While therefore clear instructions on how to integrate discrepant reports are given, it still requires judging the validity of the informant.

By using the sample of the TADS study (March et al. 2004) we aim to shed light on how an interviewer integrates different reports in clinical trials. First, we investigate in a descriptive way how the integration of the information affects the agreement between interviewer-adolescent and interviewer-parent. Second, we address the hypothesis whether the interviewer is consistently applying the "or" algorithm in case of parent-adolescent disagreement and that the integration of information is affected by gender, age, or assessment time point.

Methods

Our data analysis is drawn from the TADS study. Details on the TADS study rationale and outcome can be found elsewhere (The Treatment for Adolescents with Depression Study Team 2003; March et al. 2004). TADS is a randomized, controlled, multisite clinical trial supported by the American National Institute for Mental Health. Access to the original dataset was obtained from the controlled access datasets distributed from the National Institute of Mental Health (NIMH)-supported National Database for Clinical Trials (NDCT). NDCT is a collaborative informatics system cre-

ated by the NIMH to provide a national resource to support and accelerate discovery related to clinical research in mental health (Dataset identifier: 2145). The trial investigated the effectiveness of cognitive-behavioral therapy (CBT), fluoxetine, their combination, and placebo in depressed adolescents. Institutional Review Boards at the different study sites approved the protocol. Informed consent was obtained from all adolescents and their primary caregivers (The Treatment for Adolescents with Depression Study Team 2003). The present study uses the data of the first 12 weeks.

Participants

Thirteen sites participated in the recruitment of the 439 adolescent outpatients ages between 12 and 17 years. The main inclusion criteria were a diagnosis of a major depressive disorder (MDD) according to Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV; American Psychiatric Association 1994) and a score of ≥ 45 in the CDRS-R. Additional inclusion criteria were an intelligence quotient >80 and no intake of antidepressants before study start. Main exclusion diagnoses were bipolar disorder, severe conduct disorder, substance abuse, pervasive developmental disorder, and thought disorder. Furthermore, adolescents with concomitant psychotropic medication, intolerance to fluoxetine, failed trials with CBT or other medication, a confounding medical condition, or with low levels of English- or with non-English-speaking parents were excluded. Adolescents suffering from suicidality were excluded for ethical considerations and safety issues. The sample included 439 adolescents, 239 being female (54%) with a mean age of 14.6 years (standard deviation = 1.6).

Procedure

Participants were recruited in the community through schools and advertising and at the respective site clinics. Potential and interested patients were screened through telephone and in case of positive results invited to a first visit, in which the study was presented and informed consent was handed out. After signed consent was obtained, the diagnostic interview Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children (Kaufman et al. 1997) was administered to confirm a current diagnosis of MDD according to the DSM-IV.

If the patient met all the eligibility criteria, the patient and the caregiver were invited for the baseline assessment, during which the CDRS-R interview as well as other assessments were conducted (t_0). Subsequently, the patients were randomized to one of the four treatment arms: fluoxetine, CBT, combination of fluoxetine and CBT, and placebo. The CDRS-R was repeated at each study visit that was after 6 weeks (t_1) and 12 weeks (t_2).

An independent evaluator (IE) conducted the CDRS-R interview first with the adolescent and then with the parent and integrated the two reports into a final rating. IEs were blind to treatment arm and the self-report measures of the patients. IEs had at least 6 months of experience in conducting research-related clinical interviews, carrying out the interviews according to study-specific manuals and protocols. The interrater reliability of the CDRS-R ratings were determined in 20% of the sample and was excellent with an intraclass correlation coefficient (ICC) of 0.95 (March et al. 2005).

Measures

The CDRS. The CDRS-R (Poznanski and Mokros 1996) is a clinical interview designed to estimate depression severity in children and adolescents. It assesses 14 typical depressive symptoms

rated on a 1–7 scale, except for appetite and sleep, which are scored on a 1–5 scale. Three additional nonverbal symptoms are included, which are not considered in this study as they are only rated by the interviewer. The interviewer conducts the CDRS-R separately with the child and the parent. After each interview, the interviewer scores the answers of the parent and the child and integrates all the information to reach a composite score for each item. The CDRS-R has been validated in depressed adolescents (Mayes et al. 2010) and the TADS study reported an excellent interrater reliability (March et al. 2005).

Statistics

Several variables were defined from the scores obtained in the CDRS-R. All 14 symptoms were rated by the adolescent, parent, and the interviewer, building the adolescent's, parent's, and interviewer's score, respectively. Adding the scores of all 14 symptoms resulted in the adolescent's, parent's, and interviewer's total score.

In the first part, we were interested in the triadic agreement between parent, interviewer, and child reports. Thus, ICCs using a two-way mixed model for absolute agreement were calculated for each CDRS item of the baseline data. ICCs <0.5, between 0.5 and 0.75, between 0.75 and 0.9, and >0.9 represent poor, moderate, good, and excellent agreement, respectively (Koo and Li 2016). The triadic agreement among the interviewer, parent, and interviewer was illustrated with descriptive statistics and tested for symptom differences with χ^2 -tests.

Multivariate analyses were used to determine whether the interviewers rating aligned more with the parent's or the adolescent's report. Consistent with current recommendations (Laird and Weems 2011; Laird and De Los Reyes 2013), polynomial regressions are used to test whether the agreement between the parent's and the child's report had an influence on the interviewer's rating. Mean centered total CDRS-R total scores of the parent and the adolescent, both squared CDRS-R total scores and the interaction were regressed on the total interviewer's CDRS-R total score. A significant interaction implies that the agreement between parent and adolescent influences the interviewer's rating.

In the second part, we were interested in how the interviewer integrates differential reports. A variable which always consisted of the maximum CDRS-R score of either the parent or the adolescent, was created, representing the consistent application of the "or" algorithm. This variable will subsequently be referred to as the higher score. Multivariate analyses with a 2 (highest score vs. interviewer's rating) \times 2 (gender) \times 3 (assessment points) design controlling for age were conducted to test the differences between the higher score and the interviewer's CDRS-R score.

To further test whether the interviewer did or did not apply the "or" algorithm, we defined a binary outcome, which is henceforth called binary "or" outcome. The code 1 represents the interviewer adapting the higher score, whereas 0 was coded for cases when the interviewer did not adapt the higher score. According to its definition, the binary outcome was only coded for diverging scores between the parent and the adolescent, which were analyzed over all three assessment time points. Consequently, the binary "or" outcome represents the interviewer's rating behavior for diverging parent's and adolescent's CDRS-R scores, irrespective of whether symptoms are rated high or low. Therefore, potential time effects cannot be explained by the general decrease of symptom severity over time. Descriptive statistics with χ^2 -tests were used to compare frequencies of the binary "or" outcome according to assessment time point, site, and gender.

A generalized linear mixed model (GLMM) for binary outcomes (mixed effects logistic regression) was used to test whether the interviewer's rating and his tendency to endorse the higher rating is affected over time. The binary "or" outcome was defined as the dependent variable, whereas time (baseline, t_1 , and t_2), age, and gender were included as fixed and CDRS-R item and patients as random effects. The GLMM coefficients were estimated using the Laplace approximation (Raudenbush et al. 2000). The global effect of time was estimated by comparing the model with and without time using an analysis of variance (ANOVA test). Odds ratios (ORs) with 95% confidence intervals (CIs) were computed for time, age, and gender, and p -values for the null-hypothesis OR = 1.00.

The GLMM and surface plots of the polynomial regression were calculated with R using the lmer4 and RSA package. Other analyses were carried out using SPSS Statistics version 25. Pairwise comparisons of the multivariate analyses and ICCs were corrected with Bonferroni correction (Bland and Altman 1995).

Results I: Interviewer Agreement with the Parent and the Adolescent

Descriptive results of agreement at baseline

Agreement among the adolescent, the parent, and the interviewer at baseline are shown in Table 1. The interviewer's agreement with the parent and the adolescent was moderate to excellent and the overall agreement was good. The interviewer's overall agreement was not different between the adolescent (ICC = 0.85, 95% CI 0.81–0.88) and the parent (ICC = 0.85, 95% CI 0.77–0.88). In about half of the cases at the item level (52%), the parent, adolescent, and interviewer agreed in their rating. In disagreeing cases, the interviewer used the adolescent's score in 20% of the cases and the parent's score in 21% of the cases, which was not significantly different [$\chi^2(1) = 0.311$, $p = 0.577$]. In 5% of the cases, the interviewer used neither the parent's nor the adolescent's score. In the remaining 5%, one of the scores was missing.

Profile differences among interviewer, parent, and adolescent

A multivariate analysis with the parent's, adolescent's, and interviewer's report was conducted on the 14 CDRS-R symptoms. Gender was defined as a between-subjects factor and the analysis was controlled for age.

The model revealed a significant three-way interaction among CDRS-R symptoms, gender, and report [$F(16.0, 5747.2) = 2.1$, $p = 0.007$, $\eta^2 = 0.006$]. According to our research interests, we were interested in differences between the interviewer on the one hand and the parent and adolescent on the other hand. A detailed result of the pairwise comparisons can be found in Supplementary Table S1. Figure 1 illustrates the mean differences between the interviewer's score and the parent's and adolescent's score separately for boys and girls.

In boys, the interviewer's rating did not differ significantly from the adolescents reports for guilt ($p = 0.424$), morbid ideation ($p = 0.100$), and suicidal ideation ($p = 0.222$), suggesting that the interviewer tended to assign similar values than the adolescent himself. For all other symptoms, the interviewer assigned significantly higher scores to the adolescent compared with the adolescent himself. Furthermore, the interviewer scored the adolescent significantly higher on all items compared to the parent ($p = 0.024$ for

TABLE 1. INTERVIEWER'S AGREEMENT WITH PARENT AND ADOLESCENT AT BASELINE

	<i>M</i> (A)	<i>M</i> (P)	<i>M</i> (I)	<i>ICC</i> (A, I)	<i>ICC</i> (P, I)	<i>I=A=P</i> (%)	<i>I=A</i> (%)	<i>I=P</i> (%)	<i>I else</i> (%)
Schoolwork	4.4 (1.8)	4.5 (1.9)	4.8 (1.7)	0.76	0.86	48	17	26	7
Anhedonia	4.6 (1.6)	4.7 (1.6)	5.1 (1.3)	0.68	0.73	42	22	26	9
Social withdrawal	3.5 (1.6)	3.6 (1.7)	3.9 (1.5)	0.76	0.85	45	19	26	9
Sleep	3.4 (1.7)	3.2 (1.7)	3.7 (1.5)	0.84	0.71	53	21	17	3
Appetite	2.7 (1.5)	2.6 (1.5)	3.0 (1.5)	0.76	0.72	53	17	21	8
Fatigue	4.7 (1.8)	4.6 (1.9)	5.2 (1.5)	0.73	0.67	46	23	26	5
Physical Complaints	2.5 (1.7)	2.5 (1.7)	2.9 (1.7)	0.77	0.81	49	21	23	6
Irritability	4.2 (1.7)	4.4 (1.7)	4.8 (1.5)	0.70	0.78	44	19	32	6
Guilt	2.0 (1.4)	1.7 (1.2)	2.1 (1.4)	0.90	0.71	64	18	11	4
Self-esteem	4.3 (1.8)	4.5 (1.7)	4.9 (1.5)	0.71	0.73	40	22	28	7
Depressed feelings	4.5 (1.7)	4.6 (1.5)	5.0 (1.3)	0.65	0.64	41	25	29	5
Morbid ideation	1.9 (1.5)	1.5 (1.2)	2.0 (1.5)	0.94	0.60	73	17	4	2
Suicidal ideation	1.7 (1.3)	1.3 (0.9)	1.7 (1.3)	0.92	0.62	77	14	5	2
Weeping	2.5 (1.8)	2.1 (1.6)	2.8 (1.8)	0.85	0.70	47	28	18	5
Overall	3.3 (2.0)	3.3 (2.0)	3.7 (1.9)	0.85	0.84	52	20	21	5

ICCs are Bonferroni corrected for 30 correlations and are all significant at level $p < .001$ respectively $p < 3.3e-5$. A, adolescent; I, interviewer; ICC, intraclass coefficient; P, parent.

schoolwork, $p = 0.005$ for depressed feeling; $p = 0.011$ for morbid ideation; $p = 0.006$ for suicidal ideation, $p = 0.009$ for weeping; else $p < 0.001$).

In girls, the interviewer rated the items morbid ($p = 0.999$) and suicidal ideation ($p = 0.999$) similar to the adolescents, whereas the interviewer assigned significantly higher scores compared with the adolescent's reports for all other items ($p = 0.001$ for appetite; $p = 0.006$ for physical complaints; $p = 0.002$ for guilt; $p = 0.004$ for weeping; else < 0.001). The parent's report was lower than the interviewer's rating for all items ($p < 0.001$ for all items).

In conclusion, the interviewer's rating corresponded with the adolescent's rating for morbid and suicidal ideation and in boys additionally for guilt. Otherwise, the interviewer assigned significantly higher scores compared with both the adolescent's and the parent's report.

Influence of the agreement between parent and adolescent on the interviewer's rating

To test whether the agreement between parent's and adolescent's rating has an influence on the interviewer's rating, a polynomial regression was conducted. The adolescent's rating, the adolescent's rating squared, the parent's rating, the parent's rating squared, and the interaction between parent's and adolescent's rating were regressed on the interviewer's rating. The polynomial regression was significant [$F(4, 357) = 551.6, p < 0.001, R^2 = 0.86$]. Adolescent score ($\beta = 0.605, t = 26.4, p < 0.001$), adolescent score squared ($\beta = 0.168, t = 7.9, p < 0.001$), and parent score ($\beta = 0.425, t = 19.0, p < 0.001$) significantly predicted the interviewer's rating. Parent score squared was not included in the final model due to collinearity indicating high intercorrelations between predictors.

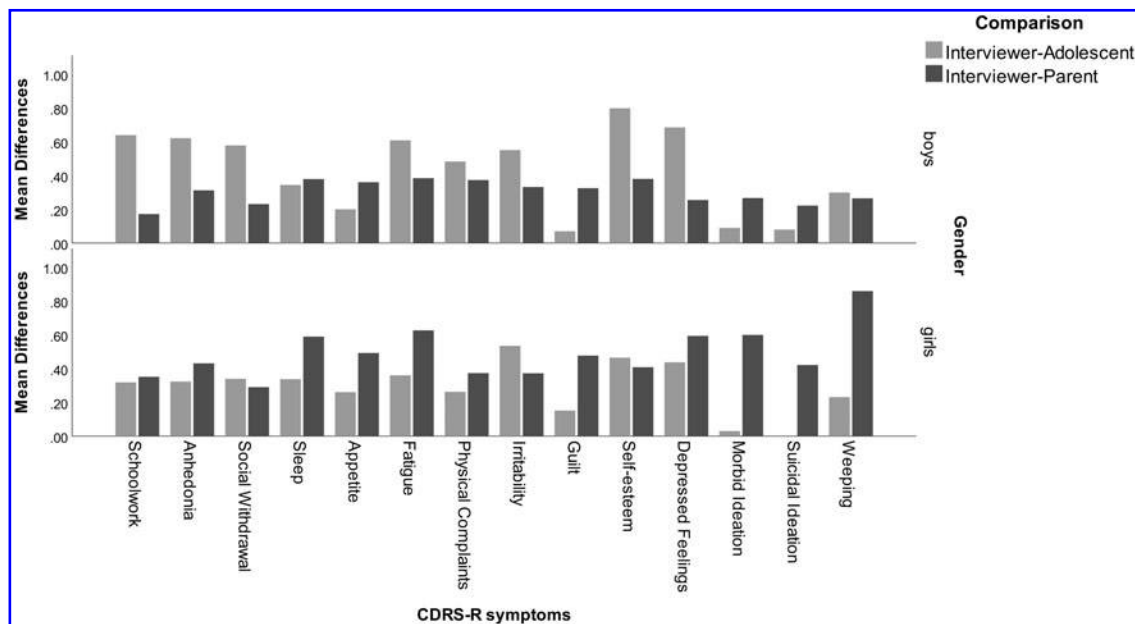


FIG. 1. CDRS-R profile mean differences of interviewer's, parent's, and adolescent's reports controlling for age. CDRS-R, Children's Depression Rating Scale-Revised.

The interaction between adolescent's and parent's report was significant ($\beta=0.077$, $t=3.6$, $p<0.001$) indicating that the agreement between these reports influence the interviewer's rating. Surface plotting suggest that the parent's report had a higher influence on the interviewer's rating when the adolescent's CDRS-R score was high. The surface plot in Figure 2 illustrates the nonlinear relationship between the interviewer's score and the adolescent's score. When parent and adolescent both reported a high CDRS-R score, the interviewer's score was disproportionately higher compared with cases where only one informant reported a high score.

Results II: The Use of the "Or" Algorithm

Profile differences between the interviewer's rating and the higher score

In the second part of our results, we specifically aim to investigate the integration of discrepant reports. Specifically, we were interested to assess whether the interviewer is simply using the higher scores, representing a consistent use of the "or" algorithm.

In a first step, we have computed a new variable representing the maximum score of the parent or adolescent, adapting either the one of the parent or the one of the adolescent. We then compared in a multivariate analysis whether this higher score differed from the interviewer's rating by entering the highest score and the interviewer rating into a repeated measure ANOVA with gender and assessment time point as factors and age as a covariate.

There was a three-way interaction among rating, gender, and assessment time point [$F(2, 16,492)=3.78$, $p=0.023$ $\eta^2=0.001$]. The higher score was significantly higher than the interviewer's score for all assessment time points and for both genders ($p<0.001$, see Supplementary Table S2).

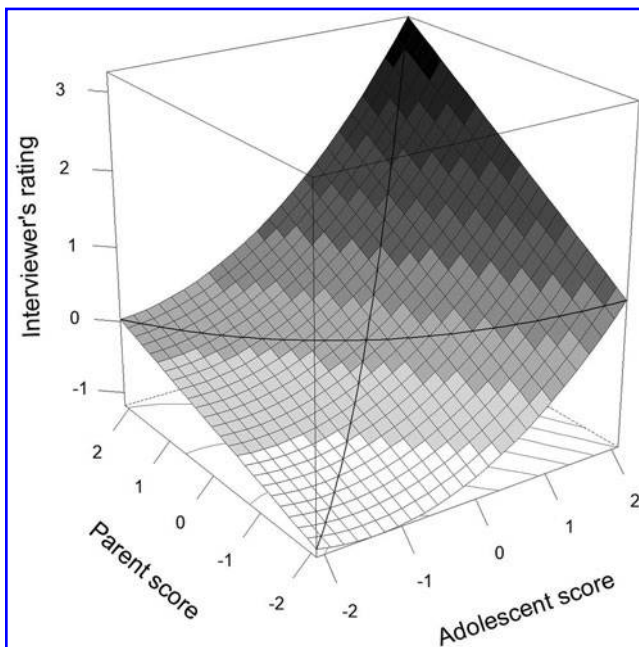


FIG. 2. Surface plot of the polynomial regression showing the interaction between the parent's and adolescent's score on the interviewer's rating. The interviewer's rating increases nonlinearly given a high parent and adolescent score. If the parent-child agreement would have had no effect on the interviewer's score, the relationship would have been linear, resulting in a less concave surface plot.

At baseline, girls and boys significantly differed according to the interviewer's ($p\leq 0.001$) and the higher score ($p\leq 0.001$). After 6 weeks though, there was no gender difference according to the interviewer's score ($p=0.171$), while the gender difference remained according to the higher score ($p=0.020$). At the 12 weeks' assessment, gender differences were significant for the interviewer's ($p=0.0401$) and the higher score ($p=0.009$). Gender interactions over time are shown in Figure 3.

Descriptive results of the "or" rule (higher rating)

We then created a new dichotomized variable representing the use of the "or" algorithm by scoring 0 when the interviewer did not adapt the higher score, while 1 was coded in cases where the higher score was adapted. Over all the assessment times, the interviewer used the "or" rule in 5459 CDRS-R item ratings (77.4%). At baseline, 2850 CRDS item ratings were divergent between parents and adolescents (Table 2). In 80.4% of these cases, the interviewer adapted the "or" rule, while in 19.6% of the cases he/she did not. The frequency of using the higher CDRS score of the two ratings was significantly less [$\chi^2(2)=25.147$, $p<0.001$] after 6 weeks (75.1% of the cases) and 12 weeks (75.7% of the cases). In girls, the interviewer used the "or" rule significantly less often over time [$\chi^2(2)=26.2$, $p<0.001$], whereas in boys the change in use did not reach level of significance [$\chi^2(2)=2.7$, $p=0.262$]. The frequency of following the "or" rule significantly differed between sites [$\chi^2(12)=215.5$, $p<0.001$], with adherence rates ranging between 65.1% and 92.1% of cases.

Longitudinal changes in the interviewer's rating behavior

A GLMM was calculated to further assess factors associated with the strategy used to combine discrepant reports. We observed a significant main effect of time [$\chi^2(2)=17.3$, $p<0.001$]. The model resulted in a significant effect for t_1 ($b=-0.30$, standard error [SE]=0.1, $z=-4.0$, $p<0.001$, OR=0.74, 95% CI=0.64–0.86) and t_2 ($b=-0.25$, SE=0.1, $z=-3.1$, $p=0.002$, OR=0.78, 95% CI=0.66–0.91). The negative coefficients indicate that the interviewer was less frequently endorsing the higher rating of the parent or adolescent after 6 and 12 weeks. Thus, the interviewer was less likely to use the "or" rule over time. Furthermore, gender emerged as a significant predictor ($b=0.027$, SE=0.1, $z=2.1$, $p=0.040$, OR=1.30, 95% CI=1.01–1.69). This suggests that the interviewer was more frequently endorsing the "or" rule in boys than girls. Age did not emerge as a predictor for the interviewer's rating behavior ($b=0.06$, SE=0.1, $z=0.9$, $p=0.358$, OR=1.06, 95% CI=0.93–1.20).

Discussion

The present study systematically investigated how the interviewer incorporates information from parents and adolescents and factors influencing the final interviewer rating. The adolescent, parent, and interviewer CDRS-R ratings used in this analysis stem from the TADS study sample, consisting of 439 adolescents diagnosed with an MDD. The TADS sample was a large clinical trial investigating treatment effectiveness of four different treatments in depressed adolescents (The Treatment for Adolescents with Depression Study Team 2003). The dataset was obtained through a special license to a controlled dataset access.

In a first part, we compared whether the interviewer's ratings corresponded more closely to the parent's or the adolescent's score.

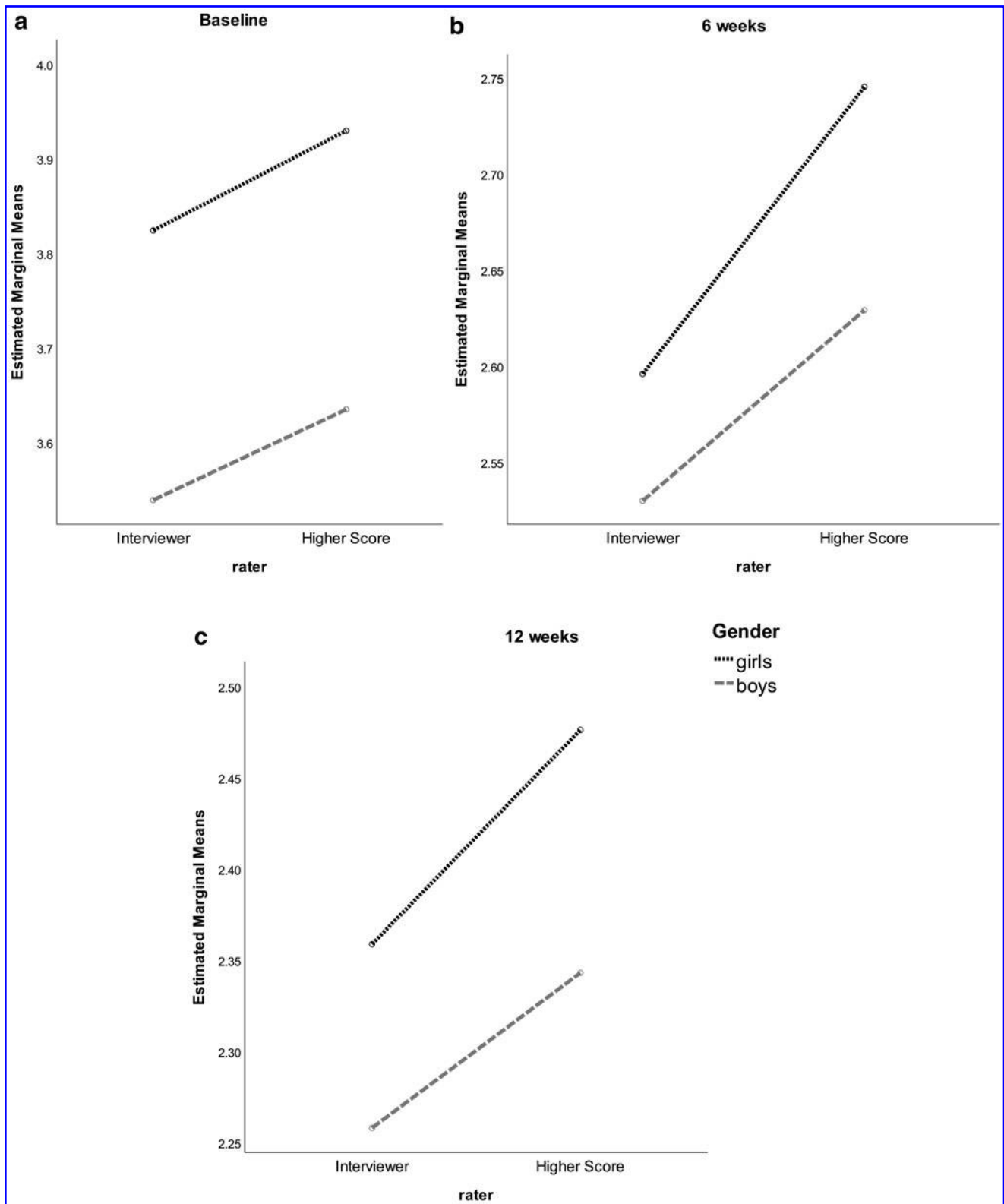


FIG. 3. Results of multivariate analyses showing a three-way interaction among rating and gender at baseline (a), after 6 weeks (b), and after 12 weeks (c) [$F(2, 16,492) = 3.78, p = 0.023, \eta^2 = 0.001$]. Age is held constant at 14.7 years. In Figure 3, the scores given by the interviewer and the ones that would have been reached by following the “or” algorithm (higher score) are shown separately for boys and girls, and for each time point. Following the “or” algorithm results in higher scores than given by the interviewer over all time points ($p < 0.001$). Furthermore, while according to the interviewer’s score girls score higher than boys at the baseline ($p < 0.001$) and the 12 weeks assessment ($p = 0.04$), this effect was not found for the 6 weeks assessment ($p = 0.171$). In contrast, when the “or” rule is applied consistently, girls would score higher than boys at each time point ($p < 0.001, p = 0.020, p = 0.009$).

TABLE 2. FREQUENCY OF THE INTERVIEWER USING THE “OR” RULE OVER TIME

	<i>Baseline</i>	<i>6 Weeks</i>	<i>12 Weeks</i>
Girls	1306 (80.3%)	997 (73.2%)	801 (73.6%)
Boys	986 (80.6%)	740 (77.8%)	629 (78.5%)
Overall	2292 (80.4%)	1737 (75.1%)	1430 (75.7%)

Overall, the interviewer’s agreement with the parent and the adolescent was good contrary to studies reporting a low agreement (Grills and Ollendick 2003; Storch et al. 2012; Hamblin et al. 2016). In over half of the CDRS-R item ratings, there was a full agreement among interviewer, parent, and adolescent severity ratings. The CDRS-R is conceptualized as a semistructured interview, which allows the interviewer to ask detailed questions that can be adapted to each individual patient. Through the possibility to seek clarifications in cases of ambiguous answers, the interviewer might be able to minimize some of the parent–child disagreement usually observed within diagnostic interviews (Grills and Ollendick 2003; Storch et al. 2012; Hamblin et al. 2016). Furthermore, while previous studies consistently report a better agreement between interviewer and parent (Grills and Ollendick 2003; Hawley and Weisz 2003; Kramer et al. 2004; De Los Reyes et al. 2011; Storch et al. 2012; Hamblin et al. 2016), this pattern was not observed in this dataset, possibly reflecting the “a priori” rule to apply an “or” algorithm in case of disagreement between adolescent’s and parent’s ratings.

The “or” algorithm resulted in the interviewer’s score being higher than the parent’s and the adolescent’s scores for all items, except for morbid and suicidal ideations for both genders, and for guilt in boys only, suggesting that for these items, the adolescents consistently scored higher values compared with their parents. Over all symptoms though, reflected in a significant gender interaction, the interviewer’s score was more similar to the parent’s scores in boys while being more similar to the adolescent’s score in girls. While girls might simply rate their symptoms more severely than their parents, the use of the “or” rule probably exacerbates gender differences. The “or” algorithm has been criticized due to the risk of overestimating symptom severity and the fact that random errors are not cancelled out symmetrically (Martel et al. 2017). However, the choice of the best algorithm might also depend on the data structure and on the level of agreement between parents and children.

Our polynomial regression suggests that the agreement between parent’s and adolescent’s report has a predictive ability on the rating of the interviewer above the two reports themselves. Namely, the rating of the interviewer is more likely to be high when both reports show a severe CDRS-R score. Contrary, when only one informant reports a high CDRS-R score, while the other does not, the interviewer is more likely to assign a lower score. Consequently, the agreement of the two informants influences the interviewer’s adaptation beyond the judged reliability or validity of the informants.

While the first part of our study investigated the effect of the “or” algorithm on interviewer–adolescent and interviewer–parent agreement, the second part analyzed more in detail the use of the “or” rule in cases of discrepant reports. The CDRS-R (Poznanski and Mokros 1996) and TADS manual advise the interviewer to use the “or” rule (meaning using the higher rating) in case of divergent adolescent–parent reports when both informants are judged to be valid and reliable. Consequently, we would expect the interviewer

to mostly adapt the higher rating whenever parents and adolescents disagree. However, in the TADS dataset in about 20% of discrepant reports, the interviewer does not follow this rule, resulting in the interviewer’s scores to be significantly lower than the higher rating for all CDRS-R symptoms.

Even more so, gender and assessment time points influence the deviation from the higher score. While at baseline, the difference between the interviewer’s final score and the highest score is somewhat similar for boys and girls, at the two follow-up time points the relationship between the interviewer’s score and the highest score changes nonlinearly between the genders. This pattern is corroborated by the results of the mixed model analysis. Interviewers were less likely to assign the higher score to girls compared with boys and they were less likely to use the “or” algorithm during the follow-up assessments compared with the baseline visit.

According to the manual, the “or” algorithm should not be applied in cases an informant is judged to be invalid or nonreliable. This recommendation is certainly useful in settings in which the interviewer is familiar with the adolescent’s and parent’s histories. However, if the interviewer is an IE blind to treatment condition and clinical information, the interviewer’s ability to judge the reliability and validity of an informant’s report is questionable. The observed influence of gender and assessment time point on the frequency of invalidity judgments suggests a certain bias when integrating information. For example, the interviewer might tend to assign the higher CDRS-R score in the beginning of the trials for participants to meet the inclusion criteria. At later assessments, the interviewer less frequently adapts the higher rating, especially in girls, potentially due to the assumption that the participant must feel better. Furthermore, implicit gender biases might also play a role, for example girls might be seen as more prone to exaggerate their symptoms compared with boys. The adherence to the “or” rule varied across recruitment sites, suggesting that different interviewers adapted different integration strategies, despite a centralized training. More frequent trainings or a regular centralized supervision of the interviews might be warranted.

Limitations

Our analysis of the CDRS agreement between adolescent and parents and interviewers has some limitations. First of all, the agreement among adolescents, parents, and interviewers was generally very good, with an identical score being rated by all parties in over half of the cases. Also, in the vast majority of the cases, the interviewer followed the “or” algorithm, with the difference between the interviewer’s score and the highest score probably not amounting to any clinically relevant differences. While therefore the conclusions of the TADS study are not disputed in any way, the dataset is used as an example to highlight the importance of defining “a priori” how to deal with adolescent–parent disagreement, in particular for primary outcome measures. Furthermore, we did not have access to data explaining why an informant might have been judged to be unreliable. Future research should investigate different influencing factors, such as the gender of the interviewer or the effect on specific CDRS-R symptoms. Analyses of these data might give a more in-depth and accurate picture of the decision making of the interviewer. Furthermore, it would be interesting to compare rule-based algorithms to the best-estimate method, identifying factors favoring one approach over the other, such as the expertise of the interviewer or his familiarity with the patients and their circumstances.

Conclusion

Composite scores are the main outcome in many studies in child research and most importantly the evaluation of treatment methods is based on such scores. The integration of discrepant reports affects treatment evaluation (Weisz et al. 2006). When only one informant evaluates a treatment as effective, the integration could lead to a loss of important information, namely that the second informant did not observe a positive treatment effect.

However, if discrepant reports are not handled in a consequent way, they pose the danger of introducing potentially systematic errors and subjective biases. Therefore, we suggest the following recommendation for future clinical trials: First of all, a manual with clear rules on how to handle discrepant reports should be established, aiming to minimize subjective evaluations of the interviewer, and these rules should be clearly stated in the method section. While a rigid algorithm might be the most transparent way of integrating information, these algorithms have their own fallacies. Rigid methods might ultimately result in a loss of information and might also introduce biases depending on the underlying data structure, while less rigid integration methods inevitably lead to more subjective outcome measures. Possibly, an integration algorithm such as the “or” rule might be compared with the results obtained by the interviewer’s integration to determine whether essential differences exist and whether the final outcome might be affected by the way divergent reports are reconciled. We suggest that additionally the level of agreement between parent’s and adolescent’s report should be mentioned in a study. In cases of good agreement, the outcome will not vary substantially depending on how the information was integrated. In cases of high discrepancy however, it might be preferable to calculate effect sizes and primary outcomes separately for the parent and child reports, allowing a differentiated evaluation of the results without introducing a bias based on the interviewer’s integration.

Clinical Significance

The integration of discrepant parent’s and child’s reports in clinical interviews affects diagnostic decisions, prevalence estimates, and treatment outcomes, highlighting the need to further investigate how such reports are integrated. The results of this study indicate that despite clear guidance, the interviewer’s integration is affected by factors such as gender or assessment time, suggesting that implicit biases might hamper the validity of the outcome measure. Therefore, the level of agreement between parents’ and children’s reports should always be considered when reporting results of clinical interviews. Future studies investigating the influences of the interviewer’s decision making on research outcomes are warranted.

Disclosures

S.W. has received royalties from Thieme Hogrefe, Kohlhammer, Springer, and Beltz in the last 5 years. S.W. has received lecture honoraria from Opopharma in the last 5 years. Her work was supported in the last 5 years by the Swiss National Science Foundation (SNF), diff. EU FP7s, HSM Hochspezialisierte Medizin of the Kanton Zurich, Switzerland, Bfarm Germany, ZInEP, Hartmann Müller Stiftung, Olga Mayenfisch, and Gertrud Thalman Fonds. G.B. was supported by the Swiss National Science Foundation, the Stanley Foundation, the Gertrud Thalman Fonds, and the Ebnet Foundation and has received lecture honoraria from Lundbeck, Opopharma, Antistress AG (Burgerstein) in the last 5 years. N.B., S.E., S.F., and I.H. declare no conflict of interest.

Supplementary Material

Supplementary Table S1
Supplementary Table S2

References

- American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Washington, DC: American Psychiatric Association; 1994.
- Bird HR, Gould MS, Staghezza B: Aggregating data from multiple informants in child psychiatry epidemiological research. *J Am Acad Child Adolesc Psychiatry* 31:78–85, 1992.
- Bland JM, Altman DG: Multiple significance tests: The Bonferroni method. *BMJ* 310:170, 1995.
- Compton SN, Walkup JT, Albano AM, Piacentini JC, Birmaher B, Sherrill JT, Ginsburg GS, Rynn MA, McCracken JT, Waslick BD, Iyengar S, Kendall PC, March JS: Child/Adolescent Anxiety Multimodal Study (CAMS): Rationale, design, and methods. *Child Adolesc Psychiatry Ment Health* 4:1, 2010.
- De Los Reyes A: Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. *Dev Psychopathol* 25:669–682, 2013.
- De Los Reyes A, Alfano CA, Beidel DC: Are clinicians’ assessments of improvements in children’s functioning “Global”? *J Clin Child Adolesc Psychol* 40:281–294, 2011.
- De Los Reyes A, Kazdin AE: Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychol Bull* 131:483–509, 2005.
- De Los Reyes A, Thomas S, Goodman KL, Kundey SMA: Principles underlying the use of multiple informants’ reports. *Annu Rev Clin Psychol* 9:123–149, 2013.
- 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, 1997.
- Grills AE, Ollendick TH: Multiple informant agreement and the anxiety disorders interview schedule for parents and children. *J Am Acad Child Adolesc Psychiatry* 42:30–40, 2003.
- Hamblin RJ, Salloum A, Nadeau JM, Ansel R, McBride NM, Lewin AB, Storch EA, Hamblin RJ: Predictors of parent-child agreement on child anxiety diagnoses on the ADIS-IV-C/P. *Psychiatry Res* 245:303–310, 2016.
- Hawley KM, Weisz JR: Child, parent, and therapist (dis)agreement on target problems in outpatient therapy: The therapist’s dilemma and its implications. *J Consult Clin Psychol* 71:62–70, 2003.
- Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N: Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* 36:980–988, 1997.
- Klein DN, Dougherty LR, Olino TM: Toward guidelines for evidence-based assessment of depression in children and adolescents. *J Clin Child Adolesc Psychol* 34:412–432, 2005.
- Koo TK, Li MY: A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15: 155–163, 2016.
- Kramer T, Phillips SD, Hargis MB, Miller TL, Burns BJ, Robbins JM: Disagreement between parent and adolescent reports of functional impairment. *J Child Psychol Psychiatry Allied Discip* 45:248–259, 2004.
- Laird RD, De Los Reyes A: Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores

- cannot tell you what you want to know and how polynomial regression may. *J Abnorm Child Psychol* 41:1–14, 2013.
- Laird RD, Weems CF: The equivalence of regression models using difference scores and models using separate scores for each informant: Implications for the study of informant discrepancies. *Psychol Assess* 23:388–397, 2011.
- March J, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, Burns B, Domino M, McNulty S, Vitiello B, Severe J, Casat C, Kolker J, Riedal K, Goldman M, Feeny N, Findling R, Stull S, McNamara N, Weller E, Robins M, Weller R, Jessani N, Waslick B, Sweeney M, Kandel R, Schoenholz D, Walkup J, Ginsburg G, Kastelic E, Koo H, Kratochvil C, May D, LaGrone R, Harrington M, Albano AM, Hirsch G, Knibbs T, Capili E, Reinecke M, Leventhal B, Nageotte C, Rogers G, Pathak S, Wells J, Arszman S, Danielyan A, Simons A, Rohde P, Grimm J, Nguyen L, Emslie G, Kennard B, Hughes C, Ruberu M, Rosenberg D, Benazon N, Butkus M, Bartoi M, Clarke G, Brent D, Koch G: The Treatment for Adolescents with Depression Study (TADS): Demographic and clinical characteristics. *J Am Acad Child Adolesc Psychiatry* 44:28–40, 2005.
- March J, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, Burns B, Domino M, McNulty S, Vitiello B, Severe J, Treatment for Adolescents With Depression Study (TADS) Team: Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA* 292:807–820, 2004.
- Martel MM, Markon K, Smith GT: Research review: Multi-informant integration in child and adolescent psychopathology diagnosis. *J Child Psychol Psychiatry Allied Discip* 58:116–128, 2017.
- 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.
- Piacentini JC, Cohen P, Cohen J: Combining discrepant diagnostic information from multiple sources: Are complex algorithms better than simple ones? *J Abnorm Child Psychol* 20:51–63, 1992.
- Polanczyk GV, Willcutt EG, Salum GA, Kieling C, Rohde LA: ADHD prevalence estimates across three decades: An updated systematic review and meta-regression analysis. *Int J Epidemiol* 43:434–442, 2014.
- Poznanski E, Mokros H: Children's Depression Rating Scale, Revised (CDRS-R). Manual. Los Angeles (California), Western Psychological Services, 1996.
- Raudenbush SW, Yang M-L, Yosef M: Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. *J Comput Graph Stat* 9:141–157, 2000.
- Smith SR: Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. *J Psychoeduc Assess* 25:139–149, 2007.
- Steinhausen H-C, Winkler Metzke C: Prevalence of affective disorders in children and adolescents: Findings from the Zurich epidemiological studies. *Acta Psychiatr Scand* 108:20–23, 2003.
- Storch EA, Ehrenreich May J, Wood JJ, Jones AM, De Nadai AS, Lewin AB, Arnold EB, Murphy TK: Multiple informant agreement on the anxiety disorders interview schedule in youth with autism spectrum disorders. *J Child Adolesc Psychopharmacol* 22:292–299, 2012.
- The Treatment for Adolescents with Depression Study Team: Treatment for Adolescents With Depression Study (TADS): Rationale, Design, and Methods. *J Am Acad Child Adolesc Psychiatry* 42:531–542, 2003.
- Weisz JR, McCarty CA, Valeri SM: Effects of psychotherapy for depression in children and adolescents: A meta-analysis. *Psychol Bull* 132:132–149, 2006.

Address correspondence to:

Noemi Baumgartner, PhD
Department of Child and Adolescent Psychiatry
and Psychotherapy
Psychiatric University Hospital Zurich
University of Zurich
Neumünsterallee 9
Zurich, CH-8032
Switzerland

E-mail: noemi.baumgartner@uzh.ch