

Associations between family and clinician ratings of child mental health: A study of UK CAMHS assessments and outcomes

Dion N Terrelonge^{1,2} and Andrew JB Fugard^{1,3,4}

¹Child Outcomes Research Consortium, London, UK

²Educational Psychology Service, Social Care and Wellbeing Directorate, Tower Hamlets, London, UK

³Research Department of Clinical, Educational and Health Psychology, University College London, UK

⁴Department of Psychosocial Studies, Birkbeck, University of London, UK

Abstract

Background: The rated severity of child mental health problems depends on who is doing the rating, whether child, carer or clinician. It is important to know how these ratings relate to each other.

Aims: To investigate to what extent clinicians' views are associated with carers' and young people's views in routine care in the United Kingdom.

Method: Ratings of clinician and parent/child viewpoints from a large Child and Adolescent Mental Health Services (CAMHS) sample (*n*s 1773–47,299), as measured by the Children's Global Assessment Scale (CGAS) and Strengths and Difficulties Questionnaire (SDQ) respectively, were analysed. The parent SDQ added value score (AVS), which adjusts for regression to the mean and other non-treatment change, was also included in the analyses.

Results: Small-to-medium correlations were found between family and clinician ratings; however, ratings diverged for the lowest-function CGAS bands. Regression analyses showed that pro-social ratings from both child and parent contributed to clinician ratings. Knowing child-reported emotional problem severity made parent ratings of emotions irrelevant to clinician judgements. There was a positive association between SDQ AVS and CGAS; as hypothesised, CGAS showed more change than the SDQ AVS, suggesting that clinicians over-estimate change.

Conclusion: This study shows the importance of multi-informant data gathering and the integration of multiple views by clinicians when monitoring outcomes.

Keywords

Clinician, parent, child, CGAS, SDQ, added value score, outcomes, CAMHS, clinician-rated presenting problems

Introduction

Assessment of the severity of a child's mental health difficulties depends on who is doing the rating, for example, the child themselves, carers, or clinicians. As a consequence of this, it has been recommended to take multiple perspectives into consideration, and in practice, mental health services often use multiple measures (Essau & Petermann, 1999; Uher et al., 2012; Wolpert et al., 2012). For example, teachers report more frequently on children's behavioural difficulties than do parents and less frequently on children's emotional difficulties (Brown et al., 2006; Jensen et al., 1999; Verhulst & Van der Ende, 2008). This difference may be due to how salient such problems are within the classroom, compared to home, and differing expectations of emotional and behavioural norms (Atzaba-Poria, Pike, & Barrett, 2004; Stanger & Lewis, 1993). When compared to children's self-report, parents were less sensitive to emotional problems and more sensitive to behavioural problems (Sourander, Helstelä, & Helenius, 1999).

Studies of associations between parent and clinician perspectives have found low-to-moderate correlations (Shaffer et al., 1983; Stavrakaki, Vargo, Roberts, & Boodoosingh, 1987) and decidedly different ratings on measures of autism spectrum condition (ASC; Lemler, 2012). Lemler (2012), like Stavrakaki et al. (1987), hypothesised that differences in ratings were due to differences in relationships; that is, the parent-child relationship is likely to be closer compared to the less familiar child-clinician relationship. Bastiaansen, Koot, Ferdinand, and Verhulst (2004) found moderate correlations between parent and clinician report on quality-of-life ratings, and between parent and child, but low correlations between child and clinician. They also found that clinicians rated children with pervasive developmental disorders as having a lower quality of life than did the children themselves and their parents.

Achenbach, McConaughy and Howell (1987) performed a meta-analysis of correlations between reports of the emotional and behavioural difficulties of children and young people, gathered from 119 studies, and found significant correlations between all informant types. However, lower levels of correlation were found between mental health workers and families (parents, $r = .24$; self, $r = .27$) than mental health workers and other professionals or non-family members (teachers, $r = .34$; observers $r = .42$; peers, $r = .44$). A later study by DeFife, Drill, Nakash and Westen (2010) reported greater correlation between self and therapist ($r_s = .40$ to $.66$).

This study explores to what extent clinicians' views, as measured by the Children's Global Assessment Scale (CGAS), are associated with parents' and young people's views as measured by the Strengths and Difficulties Questionnaire (SDQ) at both outset and follow-up using a large national database. There is a scarcity of research on correlations in symptom reports at the end of treatment. Understanding when raters agree and differ in their judgments is important for evaluating the effectiveness of treatments and making clinical decisions. For example, it will be helpful for clinicians to understand for what symptoms they can rely on their judgements about when someone has improved enough to end treatment or whether they should always seek formal feedback from children and young people using outcomes measures. We expected the various perspectives to correlate with each other; however, we had no prior expectation about whether correlations would be of the same magnitude at the end of treatment compared to at outset. We hypothesised that the CGAS would show more change than the SDQ added value score (AVS) since the latter adjusts for regression to the mean and other non-treatment factors.

Method

Sample

The data used in this article come from a data set of 161,979 episodes of care collected by the Child Outcomes Research Consortium (CORC), a learning collaboration of service managers, researchers, clinicians, commissioners and other professionals working together to use routine outcome monitoring for service evaluation and improvement. Of the 161,979 episodes of care,

Table 1. Descriptive statistics of each measure at Time 1 and Time 2: size (*n*), mean (*M*), standard deviation (*SD*), skewness and kurtosis.

Measure	Time 1					Time 2				
	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	<i>n</i>	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
CGAS	71,626	56.5	12.8	-0.12	0.87	29,851	64.5	14.3	-0.26	0.16
Parent SDQ										
Hyperactivity	81,858	6.2	2.9	-0.36	-0.92	17,294	5.4	3.0	-0.04	-1.09
Emotion	81,899	5.3	2.8	-0.11	-0.94	17,284	4.2	2.8	0.30	-0.86
Conduct	82,169	4.2	2.7	0.29	-0.76	17,297	3.3	2.6	0.66	-0.32
Peer problems	81,896	3.7	2.5	0.36	-0.64	17,290	3.2	2.4	0.57	-0.40
Pro-social	82,110	6.6	2.5	-0.48	-0.46	17,284	7.1	2.4	-0.72	-0.17
Total difficulties	81,750	19.4	7.3	-0.15	-0.47	17,269	16.0	8.0	0.18	-0.72
Child SDQ										
Hyperactivity	47,685	5.7	2.6	-0.19	-0.71	10,759	4.7	2.6	0.06	-0.75
Emotion	47,852	5.4	2.7	-0.16	-0.87	10,761	4.3	2.7	0.20	-0.84
Conduct	47,880	3.8	2.3	0.38	-0.46	10,757	2.9	2.2	0.68	-0.02
Peer problems	47,680	3.1	2.3	0.61	-0.19	10,755	2.7	2.2	0.73	0.00
Pro-social	47,699	7.1	2.1	-0.61	-0.04	10,751	7.5	2.3	-1.19	1.37
Total difficulties	47,613	18.0	6.5	-0.04	-0.28	10,804	14.5	7.2	0.12	-0.40

CGAS: Children's Global Assessment Scale; SDQ: Strengths and Difficulties Questionnaire.

sample subsets of 1733–47,299 were analysed (see individual analysis *ns*) with the age range of the largest available sample ($N = 47,299$) being 4–18 years, $M = 12$ years (standard deviation (SD) = 3.56 years), where both CGAS and parent SDQ data were present. Cases were submitted by 45 NHS Child and Adolescent Mental Health Services (CAMHS) from across the United Kingdom.

The ages of the CORC data set sample primarily lay between 6 and 11 years (48%) and 12 and 18 years (44%). In total, 55.9% were males and 44.1% females; over half (62.8%) were White British, 26.8% from other ethnic backgrounds and 10.4% unrecorded. The most common presenting problems, as recorded by clinicians using a checklist developed by CORC (Wolpert et al., 2012), were emotional disorder (47.5%), conduct disorder (17.9%), other (21.6%), conduct (15.3%), hyperkinetic (12.1%) and autism spectrum disorder (10.2%). Out of the measures included in the study, parent SDQ was the most prevalent in the CORC data set. There was also more data submitted at Time 1 (onset) than at Time 2 (follow-up; see Table 1).

Measures

CGAS. The CGAS is a clinician-rated scale of general functioning rated as an integer from 1 to 100, with 100 signifying superior functioning and 1 indicating extreme impairment (Shaffer, 1983). The scale is divided into 10 deciles with detailed descriptions to help raters navigate the range; they then choose a number within the decile. The CGAS is a widely used measure, shown

to have high inter-rater reliability, with intra-class correlations (ICCs) of .84 between clinicians' initial ratings, and excellent test–retest stability with ICCs of .85 between raters 6 months after initial ratings (Shaffer et al., 1983). More modest ICCs of .73 and .61 were found with larger samples (Hanssen-Bauer, Aalen, Ruud, & Heyerdahl, 2007; Lundh, Kowalski, Sundberg, Gumpert, & Landén, 2010). CGAS is sensitive to ‘differences in level of impairment’, with the tool yielding statistically significant differences between outpatient and inpatient scores (Shaffer et al., 1983).

SDQ. The SDQ is a standardised emotional and behavioural screening questionnaire comprising 25 items on five scales: emotional symptoms, hyperactivity, conduct, pro-social behaviour and peer relationships (Goodman, 1997). Each scale results in a score in the range 0–10. Summing the total of all scales except the pro-social scale generates a total difficulties score. Parent and teacher versions are available for 4- to 16-year-olds and a self-report version for 11- to 16-year-olds. The SDQ is one of the most widely used measures in CAMHS, and its validity, reliability and ability to accurately distinguish between clinical and non-clinical samples have been extensively investigated (e.g. Goodman, 2001; Goodman, Meltzer, & Bailey, 1998; Goodman & Scott, 1999; Johnson, Hollis, Marlow, Simms, & Wolke, 2014).

SDQ AVS. The SDQ AVS estimates the impact of treatment over and above that which would be expected without intervention. It was developed using a community sample of children and young people with clinical levels of difficulty, most of whom were yet to receive intervention by the second time point around 6 months later (Ford, Hutchings, Bywater, Goodman, & Goodman, 2009). It is computed using initial and follow-up parent SDQ ratings. Scores above zero suggest more improvement than would be expected by non-treatment factors alone. SDQ AVS has been validated by comparing its estimates with those from randomised controlled trials, where the treatment effect was known (Rotheray et al., 2014). Although still experimental, it does appear to help estimate change in routine care where control groups are unavailable.

Presenting problems checklist. The presenting problems checklist was developed by CORC and is rated by the clinician. It includes hyperkinetic disorder, emotional disorder, conduct disorder, eating disorder, psychotic disorder, self-harm, autistic spectrum disorder, learning disability, developmental disability, habit disorder, substance disorder and other presenting problems (Wolpert et al., 2012).

Statistical analysis

Statistical analyses were conducted using SPSS version 19. Pearson's correlations and regression analyses were used to estimate agreement between respondents at Time 1 and Time 2. The level of statistical significance was set at $\alpha = .05$. Cohen's (1992) criteria were used for effect size interpretations.

Results

To help contextualise results, Table 2 shows the distribution of CGAS scores by the presenting problems at Time 1 ($N = 25,162$). Very few people had CGAS scores below 30 (severe problems to extremely impaired). People presenting with psychosis or learning disability were most likely to have a score in this range (11.1% and 11.6%, respectively), with percentages in other groups ranging from 0.6% to 6.2% (for autism). These groups were correspondently less likely to be in the higher CGAS bands.

All SDQ subscores correlated with clinician CGAS. At Time 1, there were small correlations between parent and child SDQ subscales and clinician CGAS (parent $r = -.25$ to $-.14$, child $r = -.23$ to $-.13$; see Table 3). At Time 2, medium correlations were observed between parent SDQ subscales and CGAS ($r = -.45$ to $-.29$) and small-to-medium correlations between child SDQ subscales and CGAS ($r = -.35$ to $-.25$). Correlations between CGAS and the pro-social behaviour subscale, which is a positively rated item, were small at Time 1 for parent ($r = .20$) and child ($r = .14$), and small to medium at Time 2 for parent ($r = .34$) and child ($r = .25$). Using Fisher's r -to- z transformation, all correlations between SDQ subscales and clinician CGAS were statistically significantly stronger at Time 2 than at Time 1.

Table 2. Distribution of Time 1 CGAS scores, by presenting problem.

Presenting problem	CGAS band									
	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
Hyperkinetic disorder (12.1%)	0.3	0.3	0.9	8.4	32.9	32.7	18.6	4.6	1.1	0.1
Emotional disorder (47.5%)	0.3	0.4	0.9	6.6	28.5	30.6	22.9	7.9	1.9	0.1
Conduct disorder (15.3%)	0.3	0.4	0.5	8.9	31.2	33.0	18.9	5.4	1.4	0.1
Eating disorder (3.9%)	0.2	0.4	0.8	11.7	28.5	27.2	20.2	7.5	3.4	0.1
Psychotic disorder (1.9%)	2.3	2.3	6.5	18.4	24.3	19.0	16.2	8.6	2.3	0.0
Deliberate self-harm (6.2%)	0.3	0.9	0.7	6.9	29.8	28.4	21.6	8.8	2.6	0.1
Autistic spectrum disorder (10.2%)	1.4	2.3	2.5	15.2	35.0	24.9	14.6	3.4	0.7	0.0
Learning disability (4.8%)	2.9	4.3	4.4	17.9	30.0	21.2	13.7	4.2	1.2	0.2
Developmental disability (4.6%)	0.2	1.2	1.0	9.7	31.7	29.4	19.3	6.1	1.3	0.0
Habit disorder (3.6%)	0.0	0.4	0.3	4.7	23.8	29.1	28.4	9.9	3.1	0.2
Substance disorder (1.8%)	0.2	0.2	0.2	9.9	28.7	27.8	21.1	9.1	2.6	0.2
Other (21.6%)	0.2	0.2	0.9	5.7	21.9	27.9	25.8	12.9	3.8	0.6

CGAS: Children's Global Assessment Scale.

Table 3. Pearson r correlations between clinician CGAS and parent and child SDQ subscale scores at Time 1 ($N = 47,299$) and Time 2 ($N = 14,668$).

	Time 1		Time 2	
	Parent	Child	Parent	Child
Hyperactivity	-.15	-.13	-.35	-.25
Emotional problems	-.14	-.16	-.29	-.28
Conduct problems	-.17	-.14	-.33	-.26
Peer problems	-.22	-.19	-.36	-.25
Pro-social behaviour	.20	.14	.34	.25
Total difficulties	-.25	-.23	-.45	-.35

CGAS: Children's Global Assessment Scale; SDQ: Strengths and Difficulties Questionnaire.

Note all correlations are highly significant ($p < .01$, two tailed).

Tables 4 and 5 show that for CGAS bands of obvious problems (41-50) to doing very well (91-100), there is a linear relationship between clinicians' ratings and both parents and children's ratings. This linearity breaks down for the lowest bands (1-30). The lack of concordance is most marked in parent mean scores; for instance, the mean parent SDQ total difficulties scores for the two lowest functioning CGAS bands (1-20) are similar to those for the higher bands (51-70; see Table 4). A similar pattern is shown for the child-rated SDQ against the CGAS (see Table 5).

Table 6 shows how transitions between CGAS bands from Time 1 to Time 2 correspond with AVSs. Generally, greater change on the CGAS is associated with a larger AVS. However, staying in CGAS band 41-50 or in 51-60 between Time 1 and Time 2 is associated with a

negative AVS, which indicates deterioration. To have a positive AVS starting in band 41–50 requires jumping two bands to 61–70. All transitions beginning 41–70 and ending 71 or above (interpreted as at worst a minor impairment) were associated with a statistically significantly positive AVS.

AVS was regressed on Time 1 and Time 2 CGAS scores. No interaction was found between Time 1 and Time 2 CGAS ($t = 0.25, p = .25$). Nor was there a main effect of Time 1 at the 5% level of statistical significance; however, the coefficient was negative as one would expect with regression to the mean, that is, the further the Time 1 CGAS is away from the mean, the more the AVS adjusts (Est = $-.01$, standardised Est = $-.12, t = -1.88, p = .06$). A statistically significant main effect of CGAS score at Time 2 on AVS was found (Est = $.03$, standardised Est = $.36, t = 6.25, p < .001$). The intercept was -1.26 (standard error (SE) = $.43$). Overall, the model explained 11% of the variance in AVS, $F(2, 4243) = 261.5, p < .01$.

Table 4. Mean (SD) parent SDQ subscale scores by CGAS band at Time 1 ($N = 44,073$).

CGAS band	Hyperactivity	Emotional Conduct	Peer problems	Pro-social behaviour	Total difficulties	
1–10	5.8 (3.10)	4.8 (2.71)	4.0 (2.56)	3.4 (2.46)	5.8 (3.03)	17.7 (7.21)
11–20	6.4 (3.03)	5.5 (2.80)	4.4 (2.89)	4.2 (2.67)	5.3 (3.08)	20.4 (7.26)
21–30	6.0 (2.89)	6.0 (2.81)	3.9 (2.76)	5.0 (2.58)	5.5 (2.95)	20.8 (7.14)
31–40	6.9 (2.79)	5.8 (2.83)	4.9 (2.80)	4.7 (2.45)	5.6 (2.67)	22.3 (6.88)
41–50	6.5 (2.85)	5.9 (2.78)	4.5 (2.74)	4.2 (2.55)	6.2 (2.52)	21.1 (7.00)
51–60	6.4 (2.87)	5.6 (2.73)	4.3 (2.64)	3.7 (2.42)	6.5 (2.42)	20.0 (6.93)
61–70	5.7 (2.92)	5.2 (2.74)	3.7 (2.56)	3.2 (2.40)	7.0 (2.33)	17.8 (7.14)
71–80	5.3 (2.93)	4.8 (2.77)	3.3 (2.41)	2.9 (2.31)	7.3 (2.21)	16.3 (7.17)
81–90	4.7 (2.87)	4.1 (2.69)	2.7 (2.33)	2.5 (2.17)	7.7 (2.14)	14.1 (7.02)
91–100	3.4 (2.65)	3.5 (3.04)	2.0 (2.07)	2.0 (1.89)	8.4 (1.90)	11.0 (7.58)

CGAS: Children’s Global Assessment Scale; SDQ: Strengths and Difficulties Questionnaire; SD: standard deviation.

Table 5. Mean (SD) child SDQ subscale scores by CGAS band at Time 1 ($N = 27,171$).

CGAS band	Hyperactivity	Emotional Conduct	Peer problems	Pro-social behaviour	Total difficulties	
1–10	5.5 (2.53)	4.5 (2.39)	3.7 (2.41)	2.1(1.87)	6.8 (2.48)	16.2 (5.84)
11–20	5.9 (2.53)	5.5 (2.81)	4.3 (2.53)	3.6 (2.50)	6.2 (2.56)	19.8 (6.65)
21–30	5.5 (2.56)	6.1 (2.84)	3.6 (2.55)	4.3 (2.61)	6.4 (2.50)	20.6 (7.45)
31–40	6.2 (2.64)	5.9 (2.81)	4.2 (2.58)	3.8 (2.44)	6.5 (2.31)	20.2 (6.76)
41–50	6.0 (2.55)	6.0 (2.66)	4.0 (2.37)	3.5 (2.37)	6.8 (2.14)	19.5 (6.34)
51–60	5.8 (2.54)	5.5 (2.62)	3.9 (2.23)	3.0(2.19)	7.0 (2.08)	18.3 (6.17)
61–70	5.4 (2.59)	5.1 (2.62)	3.5 (2.21)	2.6 (2.14)	7.3 (1.98)	16.6 (6.43)
71–80	5.1 (2.56)	4.7 (2.60)	3.1 (2.11)	2.4 (2.06)	7.5 (1.97)	15.4 (6.49)
81–90	4.6 (2.58)	4.1 (2.59)	2.8 (2.00)	2.0 (1.96)	7.7 (1.91)	13.5 (6.28)
91–100	3.5 (2.31)	3.9 (2.77)	1.6 (1.91)	1.6 (1.84)	8.8 (1.12)	10.7 (7.19)

CGAS: Children’s Global Assessment Scale; SDQ: Strengths and Difficulties Questionnaire; SD: standard deviation.

Table 6. 95% Confidence intervals for mean added value score effect size within each CGAS band transition (where each cell $n \geq 20$; total $n = 4246$).

CGAS T1	CGAS T2					
	31–40	41–50	51–60	61–70	71–80	81–90
31–40	[-0.34, 0.15]	[-0.77, -0.29]	[-0.47, 0.19]	[-0.07, 0.51]	[-0.31, 0.49]	
41–50	[-0.69, 0.01]	[-0.48, -0.28]	[-0.34, -0.10]	[0.08, 0.39]	[0.50, 0.92]	[0.53, 1.08]
51–60		[-0.77, -0.41]	[-0.37, -0.17]	[0.02, 0.22]	[0.47, 0.76]	[0.79, 1.16]
61–70		[-0.45, 0.21]	[-0.50, 0.07]	[-0.13, 0.09]	[0.25, 0.48]	[0.59, 0.92]
71–80				[-0.48, 0.24]	[-0.07, 0.31]	[0.29, 0.61]
81–90						[0.29, 0.85]

CGAS: Children’s Global Assessment Scale.

The interval is in bold if it is statistically significantly ($p < .05$) above 0 and in bold italic if below zero.

Regression models were used to predict CGAS score based on parent and child SDQ subscales at Time 1 and Time 2 (see Table 7; one model was used for each time point). At Time 1, 11% of the variance in clinicians CGAS ratings was accounted for by parent and child SDQ scores, $F(12, 21,175) = 210.9, p < .01$. At Time 2, 27% of the variance in CGAS scores could be accounted for by SDQ, $F(12, 1760) = 56.8, p < .01$.

Parent-rated pro-social behaviour was the strongest predictor of Time 1 CGAS, and second strongest when taken from children’s report. The strongest predictor of CGAS at Time 1 from child’s report was emotional problems, but this was not statistically significant for parent report. At Time 2, child-rated emotional problems and pro-social behaviour were the strongest predictors of Time 2 CGAS. From the parent’s perspective, pro-social behaviour was the strongest predictor, alongside peer problems.

Table 7. Coefficients for a regression of T1 parent and child SDQ subscales on T1 CGAS, and T2 parent and child SDQ on T2 CGAS, $N = 21,340$.

	Time 1					Time 2				
	<i>B</i>	<i>SE</i>	Standardised Est	<i>t</i>	<i>p</i>	Est	<i>SE</i>	Standardised Est	<i>t</i>	<i>p</i>
Parent										
Hyperactivity	.04	.04	.01	0.93	.35	-.13	.15	-.03	-0.86	.39
Emotional problems	-.06	.04	-.02	-1.71	.09	-.30	.15	-.06	-2.07	<.05
Conduct problems	-.08	.05	-.02	-1.83	.07	-.22	.19	-.04	-1.11	.27
Peer problems	-.12	.04	-.03	-2.72	<.05	-.45	.17	-.08	-2.67	<.05
Pro-social behaviour	.24	.04	.05	5.67	<.001	.48	.17	.08	2.91	<.05
Child										
Hyperactivity	.03	.04	.01	0.71	.48	.10	.16	.02	0.62	.54
Emotional problems	-.26	.04	-.06	-6.43	<.001	-.51	.16	-.10	-3.30	<.05
Conduct problems	-.01	.05	-.00	-0.25	.80	.08	.20	.01	0.42	.68
Peer problems	-.24	.05	-.05	-5.22	<.001	.17	.18	.03	0.96	.34
Pro-social behaviour	.33	.05	.06	7.28	<.001	.57	.17	.09	3.25	<.05

CGAS: Children’s Global Assessment Scale; SDQ: Strengths and Difficulties Questionnaire; SE: standard error.

Discussion

In this study, we investigated associations between family reports of child mental (using the self and parent SDQ) and clinician-reported levels of functioning (CGAS). We also used the SDQ AVS, which attempts to adjust the estimated change for regression to the mean and other non-treatment-related factors. The main research questions were (a) how does change on the CGAS relate to the SDQ AVS and (b) what strengths and difficulties, from the perspective of parent and child, are most predictive of the clinician's CGAS rating, at Time 1 and Time 2.

We found small-to-moderate correlations between SDQ and CGAS, with correlations being significantly stronger at Time 2 than at Time 1. The relationship between the different informant views was non-linear: when clinicians rated children in the lowest three functioning bands (1–30: 'extremely impaired', 'very severely impaired' or with 'severe problems'), the relationship between family and clinician views begun to diverge. SDQ rating was higher than would be expected with a linear relationship.

There are several possible explanations for the observed non-linearity. Problems with data quality cannot be ruled out, for example, children and young people rated as having low CGAS scores should have been rated higher. Lundh et al. (2010) point out that although training is required to ensure accurate CGAS ratings, it is common in practice for clinicians to use the measure without training. They found that untrained professionals' rating significantly disagreed with expert ratings. An alternative explanation is that the SDQ does not accurately capture the difficulties of children falling within the more severe bands of the CGAS such as psychosis, learning disabilities and ASC.

Another possibility is that since these lower bands were more likely to contain children with pervasive conditions, parents may have different norms and see characteristics as part of the young person rather than impairment. This is supported by Brown et al. (2006), who found that despite teachers and parents identifying a similar proportion of children out of a sample of 227 as having high levels of difficulty, they did not agree on which children. Parents did not identify 53% of those identified by teachers. This emulates the non-linear pattern observed in our data among the lower functioning bands (greater impairment) between parents and clinicians.

It should also be considered that although the CGAS and SDQ are measures of mental health, they do not aim to measure exactly the same thing. While the CGAS measures general functioning on one scale, the SDQ measures strengths and difficulties across five key areas. Also, the SDQ asks informants to rate how things have been over the last 6 months (Goodman, 1997), whereas CGAS asks clinicians to rate based on the worst level of functioning over the last 3 months (Shaffer et al., 1983). These differences could also contribute to the lack of observed linearity.

Our study also found that correlations between clinician-rated CGAS score and parent and child SDQ scores were stronger at Time 2 than at Time 1. This suggests greater agreement in the level of functioning and severity of problems after approximately 6 months of treatment. This finding is consistent with previous findings in adult literature (e.g. Dorz, Borgherini Conforti, Scarso, & Magni, 2004; Mattila-Evenden, Svanborg, Gustavsson, & Åsberg, 1996) and may suggest that family members and clinicians increase in their level of agreement the longer they work together because they are better able to communicate with each other. The increase in correlation at Time 2 may also simply be due to clinicians having more evidence on which to base their ratings since they have had more contact with the family, that is, it is not that they had trouble communicating but just that it takes time to communicate and observe symptoms, signs and strengths.

The results of regression analyses showed the impact of each SDQ subscale on CGAS to differ not only by time but by informant. The subscale pro-social behaviour was one of the strongest predictors of CGAS at Time 1 and Time 2 when taken from parents and self-report. At Time 1, emotional problems were not a significant predictor of CGAS when taken from parent report but were the strongest predictor when taken from self-report. The weaker correlations

between CGAS and parent-reported emotional problems on the SDQ compared to child report are in line with findings from Stavrakaki et al. (1987), who reported poorer agreement between parents and children when rating internalised problems such as depression. Clinicians heavily rely on informant information; therefore, if parents report little about or place little emphasis on emotional problems, then this may make it more difficult for clinicians to pick up on these and, subsequently, have less impact on their judgement ratings.

The SDQ is one of two child measures to provide an estimate of value added (AVS) by an intervention (the other is the YP-CORE for young people; Cooper, Fugard, Pybis, McArthur, & Pearce, 2015). There is evidence of regression to the mean in CGAS scores (Table 6) since the CGAS suggests improvement yet the AVS shows no change. Table 6 is a potentially useful clinical tool as it demonstrates what a change in CGAS banding equates to in terms of AVS.

A limitation of this study was the relatively low number of cases with SDQ and CGAS data, both at intake and follow-up 6 months later. This is a common problem in routine outcomes monitoring and has prompted the introduction of session-by-session measures to UK CAMHS (Wolpert, Fugard, Deighton, & Görzig, 2012). However, increasing the frequency of questionnaire use comes with the cost of practice effects and the requirement to ask fewer questions, hence providing a less comprehensive evaluation of progress. Additionally, the use of problem-specific measures, as proposed in the United Kingdom, presupposes that new problems will not develop over the course of care, which is not necessarily the case. Ultimately, poor return rates may only be improved by increasing funding for outcomes evaluation.

Results reported in this article demonstrate agreement in reports of child mental health, but in many instances, correlations were low. This finding does not necessarily reflect inaccuracy in reporting – it is difficult to argue who is best positioned to be objective, if that is possible at all. This article emphasises the importance of multiple perspective data collection in child mental health care, taking account of a range of viewpoints when making appraisals of functioning and symptom severity. It is hoped that the reported findings will help clinicians integrate information from different sources and use this to improve the care that children and their families receive.

Acknowledgements

The authors would like to thank Miranda Wolpert, Director of the Evidence Based Practice Unit and all members of the Child Outcomes Research Consortium (CORC) committee at the time of writing: Ashley Wyatt, Alison Towndrow, Duncan Law, Evette Girgis, Julie Elliott, Ann York, Mick Atkinson and Alan Ovenden. The CORC central team at the time of writing: Jenna Jacob, Emily Stapley, Charlotte Wray, Harriet Hockaday, Slavi Savic and Thomas Booker. Thanks also to Roger Davies, Robert Goodman and Anna Lundh for their helpful comments.

References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioural and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232.
- Atzaba-Poria, N., Pike, A., & Barrett, M. (2004). Internalising and externalising problems in middle childhood: A study of Indian (ethnic minority) and English (ethnic majority) children living in Britain. *International Journal of Behavioural Development*, *28*, 449–460.
- Bastiaansen, D., Koot, H. M., Ferdinand, R. F., & Verhulst, F. C. (2004). Quality of life in children with psychiatric disorders: self-, parent, and clinician report. *Journal of the American Academy of Child and Adolescent Psychiatry*, *43*, 221–230.
- Brown, J. D., Wissow, L. S., Gadowski, A., Zachary, C., Bartlett, E., & Horn, I. (2006). Parent and teacher mental health ratings of children using primary care services: Inter-rater agreement and implications for mental health screening. *Ambulatory Pediatrics*, *6*, 347–351.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

- Cooper, M., Fugard, A. J. B., Pybis, J., McArthur, K., & Pearce, P. (2015). Estimating effectiveness of school-based counselling: Using data from controlled trials to predict improvement over non-intervention change. *Counselling and Psychotherapy Research, 15*, 262–273.
- DeFife, J. A., Drill, R., Nakash, O., & Westen, D. (2010). Agreement between clinician and patient ratings of adaptive functioning and developmental history. *The American Journal of Psychiatry, 167*, 1472–1478.
- Dorz, S., Borgherini, G., Conforti, D., Scarso, C., & Magni, G. (2004). Comparison of self-rated measures of depressive symptoms: A naturalistic study. *Psychology and Psychotherapy: Theory, Research and Practice, 77*, 353–361.
- Essau, C., & Petermann, F. (1999). *Depressive disorders in children and adolescents: Epidemiology, risk factors, and treatment*. Northvale, NJ: Jason Aronson.
- Ford, T. J., Hutchings, J., Bywater, T., Goodman, A., & Goodman, R. (2009). Evaluation of the Strengths and Difficulties Questionnaire added value score as a method for estimating effectiveness in child mental health interventions. *The British Journal of Psychiatry, 194*, 552–557.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1337–1345.
- Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry, 7*, 125–130.
- Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the child behaviour checklist: Is small beautiful? *Journal of Abnormal Child Psychology, 27*, 17–24.
- Hanssen-Bauer, K., Aalen, O. O., Ruud, T., & Heyerdahl, S. (2007). Inter-rater reliability of clinician rated outcome measures in child and adolescent mental health services. *Administration and Policy in Mental Health, 34*, 504–512.
- Jensen, P., Rubio-Stipec, M., Canino, G., Bird, R., Dulcan, M. K., Scwab-Stone, M. E., . . . Lahey, B. B. (1999). Parent and Child Contributions to Diagnosis of Mental Disorder. Are both informants always necessary? *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 1569–1579.
- Johnson, S., Hollis, C., Marlow, N., Simms, V., & Wolke, D. (2014). Screaming for childhood mental health disorders using the Strengths and Difficulties Questionnaire: The validity of multi-informant reports. *Developmental Medicine & Child Neurology, 56*, 453–459.
- Lemler, M. (2012). Discrepancy between parent report and clinician observation of symptoms in children with autism spectrum disorders. *Discussions, 8*(2). Retrieved from <http://www.studentpulse.com/a?id=803>
- Lundh, A., Kowalski, J., Sundberg, C. J., Gumpert, C., & Landén, M. (2010). Children's Global Assessment Scale (CGAS) in a naturalistic clinical setting: Inter-rater reliability and comparison with expert ratings. *Psychiatry Research, 177*, 206–210.
- Mattila-Evenden, M., Svanborg, P., Gustavsson, P., & Åsberg, M. (1996). Determinants of self-rating and expert rating concordance in psychiatric out-patients, using the affective subscales of the CPRS. *Acta Psychiatrica Scandinavica, 94*, 386–396.
- Rotheray, S., Racey, D., Rodgers, L., McGilloway, S., Berry, V., & Ford, T. (2014). Innovations in Practice: Further evidence on the effectiveness of the strengths and difficulties added value score as an outcome measure for child and adolescent services. *Child and Adolescent Mental Health, 19*, 270–273.
- Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., . . . Aluwahlia, S. (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry, 40*, 1228–1231.
- Sourander, A., Helstelä, L., & Helenius, H. (1999). Parent-adolescent agreement on emotional and behavioral problems. *Social Psychiatry and Psychiatric Epidemiology, 34*, 657–663.
- Stanger, C., & Lewis, M. (1993). Agreement among parents, teachers and children on internalizing and externalizing behaviour problems. *Journal of Clinical Child Psychology, 22*, 107–115.
- Stavrakaki, C., Vargo, B., Roberts, N., & Boodoosingh, L. (1987). Concordance among sources of information for ratings of anxiety and depression in children. *Journal of the American Academy of Child & Adolescent Psychiatry, 26*, 733–737.
- Uher, R., Perlis, R. H., Placentino, A., Dernovšek, M. Z., Henigsberg, N., Mors, O., . . . Farmer, A. (2012). Self-report and clinician-rated measures of depression severity: Can one replace the other? *Journal of Depression and Anxiety, 29*, 1043–1049.
- Verhulst, F. C., & Van der Ende, J. (2008) Using rating scales in a clinical context. In: Rutter, M., Bishop, D., Pine, D., Scott, S., Stevenson, J., Taylor, E., & Thapar, A. (Eds.), *Rutter's child and adolescent psychiatry* (5th ed., pp. 289–298). Oxford, UK: Blackwell.

- Wolpert, M., Ford, T., Trustam, E., Law, D., Deighton, J., Flannery, H., . . . Fugard, A. J. (2012). Patient-reported outcomes in child and adolescent mental health services (CAMHS): Use of idiographic and standardized measures. *Journal of Mental Health, 21*, 165–173.
- Wolpert, M., Fugard, A. J. B., Deighton, J., & Görzig, A. (2012). Editorial commentary: Routine outcomes monitoring as part of Children and Young People's Improving Access to Psychological Therapies (CYP IAPT) – improving care or unhelpful burden? *Child and Adolescent Mental Health, 17*, 129–130.