

Further Evidence on the Effectiveness of the Strengths and Difficulties Added Value Score as an Outcome Measure for Child and Adolescent Services

Dr Seb Rotheray¹, Dr Daniel Racey¹, Dr Lauren R Rodgers¹, Dr Sinead McGilloway², Dr Vashti Berry³ Professor Tamsin Ford¹

¹ Exeter Medical School, Exeter, Devon

² Department of Psychology, National University of Ireland, [Maynooth](#)

³ People and Families Research Group, University of Central Lancashire

Abbreviated title (running head – up to 80 characters):

Effectiveness of the SDQ Added Value Score

Author for correspondence:

Email address corresponding author: t.j.ford@exeter.ac.uk

Summary / Abstract

Background: The Strengths and Difficulties Added Value Score (SDQ AVS) uses a large epidemiological study to predict follow-up parental SDQ scores for the evaluation of routine outcomes.

Method: We tested the prediction of the SDQ AVS derived from a national population survey separately on scores for the waiting list control and intervention groups in a randomised controlled trial. If the SDQ AVS is to be clinically useful, it needs to function as expected across different populations.

Formatted: Not Highlight

Results:

In the control arm, the SDQ AVS predicted an effect size of 0.15 (95% CI -0.01 to 0.30) compared to an expected effect size of 0, as the children in this arm received no treatment. In the experimental arm, the SDQ AVS predicted an effect size of 0.62 (95% CI 0.42 to 0.83) compared to the study effect size of 0.53. Change scores overestimated the effect size in both arms (control 0.50 95% CI 0.34 to 0.66, intervention 0.85 95% CI 0.66 to 1.04)

Formatted: Not Highlight

Conclusion: Our findings suggest that the SDQ AVS adjusts for spontaneous improvement, regression to the mean and attenuation.

Keywords (3-6)

SDQ AVS, Routine Outcome Measurement, Service evaluation

Formatted: Not Highlight

Introduction

Child and Adolescent Mental Health Services (CAMHS) face increasing demands to routinely monitor outcomes (Department of Health, 2013). The interpretation of change in outcome measures pre-and-post-intervention is difficult (Wolpert, Ford, & Trustam, 2012) as several factors may influence change scores. Regression to the mean is a random measurement error whereby particularly high or low baseline estimates scores tend to be followed by closer to scores nearer the mean when measured at follow-up (Last, 2001). Attenuation is a type of respondent fatigue that leads to fewer problems being reported by the respondent on successive tests (Jensen et al, 1995). Finally, the inherently fluctuating nature of childhood psychopathology tends to will may lead to apparent improvement at scores on follow-up as children are often referred at the peak of their difficulties (Ford et al, 2007). The use of an experimental design with a comparison group and random allocation, which should account for both known and unknown confounders, ameliorates these difficulties in randomised controlled trials (RCTs), but is not practical in the measurement of outcome in routine practice.

The Strengths and Difficulties (SDQ) score is a patient reported outcome measure (PROM) that is commonly used in CAMHS (Wolpert, Ford, & Trustam, 2012). It is a widely used, reliable and valid general scale of psychopathology (Goodman, 2001; see www.sdqinfo.org). The SDQ Added Value Score (SDQ-AVS) compares outcome scores predicted from a high risk epidemiological sample at 4-8 months after baseline with those actually obtained in groups of young people in receipt of targeted or indicated interventions (Ford et al, 2009). The aim is to estimate change attributable to clinical intervention in a manner analogous to growth

Formatted: Not Highlight

Commented [S1]: Seb just wondering if you actually managed to locate Last's dictionary because the reviewer didn't seem too sure about this and I am not terribly keen on his suggested wording – which you have replicated here. There would appear to be quite a few definitions of regression to the mean!

Commented [FT2]: Taken from Last!- amend wording in table

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

Field Code Changed

charts commonly used in children's services to monitor height and weight. It is calculated from the parental SDQ measured at assessment and four to eight months later, and was generated by applying linear regression to the baseline and follow-up SDQ scores of 609 children from the British Child and Adolescent Mental Health Survey 2004 (Ford et al, 2009; www.sdqinfo.org). Children were included if they either had a psychiatric disorder and / or their parents had sought advice from teachers and primary health care about their child's mental health. The aim was to produce a "control group" for clinical services. The SDQ AVS is the predicted score minus the actual follow-up SDQ total difficulties score; so if it is a score greater than zero, it suggests that the child is doing better than predicted. Similarly a negative additive value score suggests that the young person is doing worse than predicted.

Formatted: Not Highlight

Preliminary support for the validity of the SDQ AVS was demonstrated by testing it against results from single trial of a parenting intervention for behaviour problems in three- and four-year-olds (Ford et al, 2009). We aimed to further evaluate the reliability of the SDQ AVS by seeking other trials against which to test it.

Method

This study was a secondary analysis of data already obtained; the original trial received ethical approval from >>>>>, while the secondary analysis related to the SDQ AVS was covered by approval from the Peninsula School of Medicine and Dentistry Research Ethics Committee. We searched for eligible trials in trial databases, literature and contacted colleagues conducting RCTs of interventions designed to influence child mental health using the following inclusion criteria:

Commented [FT3]: Siobhan?

- (1) The SDQ score measured at baseline and follow-up with the Impact subscale.

Formatted: Not Highlight

- (2) A statistically significant difference in outcome between the intervention and control groups on the SDQ.
- (3) Children between the age of 2 and 16
- (4) The follow-up SDQ was administered 4-8 months after baseline.

Four potential trials were identified. One was telephone CBT for OCD (Robinson et al, 2012), one was for CBT and fluoxetine for depression (Goodyer et al, 2007) and two were group-based parent training programmes for emotional and behavioural problems (Little et al, 2012 and McGilloway et al, 2012). Unfortunately? Sadly, the first two of the trials listed above were the interventions for emotional disorders were equivalence studies, whilst and one parent training trial did not show sufficient difference between control and intervention arm once imputation for missing data was removed (Little et al, 2012).

Commented [S4]: Sorry Seb – the wording here seems odd to me so I have suggested an amendment
Formatted: Font: Bold

The R results from the a-single remaining study (McGilloway et al, 2012) involved 149 children from Ireland Eire, aged between 2 and 8 years who scored above the cut-off for conduct disorder on the Eyeberg Child Behaviour Inventory (ECBI). Their parents were randomly allocated to an Incredible Years parenting programme course (n=103) or waiting list control group (n=46). Twelve participants were lost to follow-up and one parent had a missing value for SDQ impact at baseline. As we wanted to avoid assumptions about missing data in order to test how the SDQ AVS predicts actual data, our analysis only included parents with complete data; that is, 94 parents in the intervention arm and 42 in the control arm. In the Trial report, The trial reported the intervention had a significant effect size, s using ANCOVA calculated using Cohen’s guidelines, of 0.53 (95% CI 0.2 to 0.9), according to the parental SDQ.

Formatted: Not Highlight

Commented [S5]: Seb the word Eire is a little contentious in some quarters! Best to refer to ‘Ireland’ or the ‘Republic of Ireland’ just to be on the safe side.

Commented [S6]: Not quite accurate – our age range was 32-88 months so approx. 3-7 years

Formatted: Not Highlight

Formatted: Not Highlight

Commented [S7]: Seb can I suggest that you re-word this sentence because it does not seem to flow terribly well?

Formatted: Not Highlight

Statistical analysis

The analysis was conducted using Stata Version 12.1. The sample from the included trial was compared with the sample from which the SDQ AVS was derived using t -tests to compare the age and SDQ scores and a Chi-squared test to compare gender. [The assumptions of all tests were checked using standard diagnostics.](#)

We calculated the AVSs and simple change scores for each child using the equations below.

Raw SDQ AVS (in SDQ points)

= $2.3 + 0.86 \times \text{baseline total difficulties score} + 0.2 \times \text{baseline impact score} - 0.3 \times \text{baseline emotional difficulties subscale score} - \text{follow-up total difficulties score}$

Raw change score (in SDQ points)

= $\text{baseline total difficulties score} - \text{follow-up total difficulties score}$

We calculated effect sizes for both the added value and change scores by dividing the raw scores by their respective standard deviations in normative samples (5.8 for the change score, 5 for the AVS; see www.sdqinfo.org). We predicted that the AVS for the control group would be zero (i.e. no change as no intervention while on the waiting list), and that the AVS for the intervention group should approximate to the per-protocol effect size reported in the original trial (0.53). A one-sample t -test compared the SDQ Added Value Scores and the change scores from the experimental sample with the expected value for each group (i.e. 0.53 for the intervention group and 0 for the control group).

[How was the per protocol effect size calculated? Was it using ANCOVA?](#)

Results

Table 1 Here

Commented [S8]: Seb I wonder would it be worth adding something brief like this just to show that the comment of the second reviewer has been addressed?

Commented [FT9]: Could do – have changed table to indicate that could take out if the word count was now too long

Formatted: Not Highlight

Formatted: Not Highlight

There were statistically significant differences in age and parent-reported SDQ at baseline between the [Irish](#) children ~~from Eire~~ and the derivation sample; children from [Ireland Eire](#) tended to be younger and were reported to have more difficulties (see Table 1). There was no significant difference found in the gender distribution.

As Table 2 shows, the SDQ AVS effect for the control arm was 0.15 (95% CI -0.01 to 0.30) compared to an expected effect size of 0 and the effect found using the SDQ AVS for the intervention group ~~was~~^{is} higher, 0.63 (95% CI 0.42 to 0.83), compared to the expected effect 0.53 (95% CI 0.2 to 0.9). However, these differences were not significant (p=0.08 in the control arm and p=0.37 in the intervention). The change score effect sizes were significantly different from expected values (p<0.001 in both the control and intervention arms). The change scores suggest that being in the control group has an effect size of 0.50 (95% CI 0.34 to 0.66), while being in the intervention group appears to have a large impact with an effect size of 0.85 (95% CI 0.66 to 1.04), higher than found in the study. The change scores would appear to be overestimating the effects of both waiting list and intervention as predicted.

Formatted: Not Highlight

Insert Table 2 here

Discussion

These results offer the second validation of the SDQ AVS tool, providing further evidence that it may be a clinically useful tool to measure the impact of clinical interventions while adjusting for the tendency for change scores to overestimate change produced by interventions (Ford et al, 2009). The SDQ AVS modulated results from the original trial,

producing an effect size that was close to zero for the control group and an effect size for the intervention group that was closer to that calculated using SDQ total difficulties scores in the original trial than simple change scores. ~~The SDQ scores in the control arm of 0.15 (-0.01 to 0.30) include confidence intervals that are close to excluding zero and, although not statistically significant, they may suggest improvement beyond what was predicted; this which may be explained by the relatively higher SDQ scores at baseline in the trial group, when compared to the derivation sample.~~

Formatted: Not Highlight

Commented [S10]: This is rather a long sentence- hence the semi-colon!

Commented [S11]: Yes almost two-thirds of our sample were socially disadvantaged which would explain the higher scores perhaps – not sure if you would like to add something on this?

Commented [FT12]: No space

The SDQ AVS was derived from 5-16 year old British children in 2004 (Green et al, 2004), who are a geographically distinct population from the study described in this paper (McGilloway S et al, 2012), with a different age-profile (age 3-8 versus 5-16) and were targeted for behaviour problems toward conduct disorder. However, ethnicity, culture and degree of psychological morbidity were broadly similar. Similar approximations from epidemiological samples to measures of change, such as growth charts, have been extremely valuable in other health settings. Alternative approaches to these norm-based trajectories come with their own difficulties. We could attempt to characterise dose-response relationships to see if more treatment sessions leads to greater improvement (Bickman, Andrade, & Lambert, 2002). However, this kind of observational study is prone to bias as duration of attendance may be positively or negatively correlated to surrogate factors affecting outcome. For example those who drop out may be either too well to need continued intervention or too distressed to engage. Another approach has been to compare treated children with other children who were offered treatment but did not attend their allocated sessions (Weisz & Jensen 2001). Again, this is unlikely to be a random sample with potential biases either favouring improvement or continued deterioration.

Limitations

While providing further evidence of the ability to adjust for change not derived from clinic attendance, the SDQ AVS has still only been tested in two RCTs, both of which are both involved of the same intervention in similar age groups. We need more trial data of different interventions among children of different ages and with a range of different difficulties in order to ascertain see how the SDQ AVS functions. Currently, we cannot be sure that our findings would generalise beyond parenting interventions among children with challenging behaviour. We need a variety of RCTs with appropriate outcome measures to calibrate the SDQ AVS.

Formatted: Not Highlight

Formatted: Not Highlight

Formatted: Not Highlight

The use of Using normative data with trajectories over time is only helpful?useful when those norms adjust for important background factors or the algorithm is robust to differences in background characteristics. The confidence intervals around SDQ scores in the control arm (-0.01 to 0.30) come close to excluding zero and, although not statistically significant, may suggest improvement beyond what was predicted. This might relate to relatively higher SDQ scores at baseline, which might in turn related to high levels of socioeconomic deprivation in the Irish sample, when compared to the derivation sample.

Formatted: Not Highlight

Commented [S13]: This is rather a long sentence- hence the semi-colon!

Commented [S14]: Yes almost two-thirds of our sample were socially disadvantaged which would explain the higher scores perhaps – not sure if you would like to add something on this?

Commented [FT15]: No space

This issue is why height charts are created for different genders. The SDQ AVS was remarkably robust to measures of case complexity in the sample from which it was derived; only 0.6% of variation in the SDQ AVS is accounted for by a wide range of case complexity variables (Ford et al, 2009). It is possible that the calculation acts as a good surrogate

variable, accounting for a wide range of case complexity. Alternatively, case complexity may not be an important predictor of the trajectory [of](#) childhood psychopathology in clinical samples, which seems unlikely, although measures of case complexity have not reliably predicted routine outcomes (Garralda, Yates & Higginson, 2000). Regrettably, there were too few background variables that we could test in the current sample.

Formatted: Not Highlight

Table 1: Comparison of the samples from which the strengths and difficulties AVS (SDQ AVS) was derived and evaluated

	SDQ AVS Derivation Sample n=609	Irish EIRE Sure Start Trial n=136
Age, years		
Range	5-16	Approx. 3-7
Mean (s.d)	11.0 (3.3)	4.9 (1.3)*
Male Gender (%)	61.1	62.5
SDQ parental total difficulties score at baseline, mean (s.d.)	15.5 (7.2)	18.6 (6.1)*

Commented [S16]: No - Sure Start was only in the UK

* P<0.001.

Table 2: Comparison of the added value Strengths and Difficulties Questionnaire (SDQ) scores and change scores with the expected effect sizes for control and intervention groups separately

	Predicted value	AVS		Change score	
		Mean & 95% CI	p-value	Mean & 95% CI	p-value
Control	0	0.15 (-0.01 to 0.30)	0.08	0.50 (0.34 to 0.66)	<0.001
Intervention	0.53 (0.2 to 0.9)	0.62 (0.42 to 0.83)	0.37	0.85 (0.66 to 1.04)	0.001

References

1. Bickman, L., Andrade, A. R., & Lambert, E. W. (2002). Dose response in child and adolescent mental health services. *Mental Health Services Research, 4*(2), 57-70.
2. Department of Health, (2013, February 19) Improving Children and Young People's Health Outcomes: A System Wide Response. Retrieved From https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/141430/9328-TSO-2900598-DH-SystemWideResponse.pdf.pdf
3. Ford, T., Collishaw, S., Meltzer, H. & Goodman, R. (2007). A prospective study of childhood psychopathology; predictors of change over three years. *Social Psychiatry and Psychiatric Epidemiology, 42*: 953–61.
4. Ford, T., Hutchings, J., Bywater, T., Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire AVSs: evaluating effectiveness in child mental health interventions. *The British Journal of Psychiatry, 194*(6), 552-558.
5. Garralda, M. E., Yates, P., & Higginson, I. (2000). Child and adolescent mental health service use HoNOSCA as an outcome measure. *The British Journal of Psychiatry, 177*(1), 52-58.
6. Goodyer, I., Dubicka, B., Wilkinson, P., Kelvin, R., Roberts, C., Byford, S, & Harrington, R. (2007). Selective serotonin reuptake inhibitors (SSRIs) and routine specialist care with and without cognitive behaviour therapy in adolescents with major depression: randomised controlled trial. *BMJ, 335*(7611), 142.

7. Green, H., McGinnity, A., Meltzer, H., Ford, T. & Goodman, R. (2004). *Mental Health of Children and Young People in Great Britain*, The Stationary Office. 2005.
8. Jensen, P.S., Roper, M., Fisher, P., Piacentini, J., Canino, G., Richters, J., et al. (1995) Test–retest reliability of the Diagnostic Interview Schedule for Children (DISC 2.1). *Archives Of General Psychiatry* 52: 61–71.
9. Last, J.M. (2001). *A Dictionary of Epidemiology*, 3rd edn: 144. Oxford University Press
10. Little, M., Berry, V., Morpeth, L., Blower S., Axford, N., Taylor, R., Bywater, T., Lehtonen, M. & Tobin, K. (2012). The impact of three evidenced-based programmes delivered in public systems in Birmingham, UK. *International Journal of Conflict and Violence*, 6(2), 260-272
11. McGilloway, S., Mhaille, G. N., Bywater, T., Furlong, M., Leckey, Y., Kelly, P. & Donnelly, M. (2012). A parenting intervention for childhood behavioral problems: a randomized controlled trial in disadvantaged community-based settings. *Journal of Consulting and Clinical Psychology; Journal of Consulting and Clinical Psychology*, 80(1), 116.
12. Robinson, S., Turner, C., Heyman, I. & Farquharson, L. (2012). The feasibility and acceptability of a cognitive-behavioural self-help intervention for adolescents with obsessive-compulsive disorder. *Behavioural and Cognitive Psychotherapy*, 41(1), 117-122 doi: 10.1017/S1352465812000562. Epub 2012 Aug 16.
13. Weisz, J.R. & Jensen, A.L. (2001). Child and adolescent psychotherapy in research and practice contexts: review of the evidence and suggestions for improving the field. *European Child and Adolescent Psychiatry*, 10, Supplement 12-18.
14. Wolpert, M., Ford, T., Trustring, E., Law, D., Deighton, J., Flannery, H. & Fugard, A.R. (2012). Patient-reported outcomes in Child and Adolescent Mental Health Services: Use of Idiographic and standardised measures. *Journal of Mental Health* 21, 165-173.

Field Code Changed

Field Code Changed

Field Code Changed

Field Code Changed

NB: As this study did not involve any clinical participants, informed consent was not required.