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

Rotheray, Sebastian, Racey, Daniel, Rodgers, Lauren, McGilloway, Sinead, Berry, Vashti Louise and Ford, Tamsin

Available at http://clok.uclan.ac.uk/23192/


It is advisable to refer to the publisher’s version if you intend to cite from the work. http://dx.doi.org/10.1111/camh.12059

For more information about UCLan’s research in this area go to http://www.uclan.ac.uk/researchgroups/ and search for <name of research Group>.

For information about Research generally at UCLan please go to http://www.uclan.ac.uk/research/

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the http://clok.uclan.ac.uk/policies/
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

Abstract

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.

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)

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
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 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 lead to apparent improvement at 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.
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.

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:

1. The SDQ score measured at baseline and follow-up with the Impact subscale.
A statistically significant difference in outcome between the intervention and control groups on the SDQ.

Children between the age of 2 and 16

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, the first two of the trials listed above 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).

The results from the single remaining study (McGilloway et al, 2012) involved 149 children from Ireland, aged between 2 and 8 years who scored above the cut-off for conduct disorder on the Eyberg 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 intervention had a significant effect size using ANCOVA calculated using Cohen’s guidelines, of 0.53 (95% CI 0.2 to 0.9), according to the parental SDQ.
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 baseline total difficulties score + 0.2 x baseline impact score - 0.3 x baseline emotional difficulties subscale score - follow-up total difficulties score

Raw change score (in SDQ points)
= baseline total difficulties score - 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
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 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.

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.

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.

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. 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.

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

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

* 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

<table>
<thead>
<tr>
<th></th>
<th>Predicted value</th>
<th>AVS</th>
<th>Change score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean &amp; 95% CI</td>
<td>p-value</td>
<td>Mean &amp; 95% CI</td>
</tr>
<tr>
<td>Control</td>
<td>0</td>
<td>0.15 (-0.01 to 0.30)</td>
<td>0.08</td>
</tr>
<tr>
<td>Intervention</td>
<td>0.53 (0.2 to 0.9)</td>
<td>0.62 (0.42 to 0.83)</td>
<td>0.37</td>
</tr>
</tbody>
</table>

References


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