European Child & Adolescent Psychiatry
9:129–134 (2000) © Steinkopff Verlag 2000

ORIGINAL CONTRIBUTION

R. Goodman
D. Renfrew
M. Mullick

Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka

Accepted: 25 October 1999

R. Goodman (✉)
Department of Child
and Adolescent Psychiatry
Institute of Psychiatry
Kings College London
London SE5 8AF, UK

D. Renfrew
Child and Family Service
Ayrshire Central Hospital
Kilwinning Road
Irvine, Ayrshire KA12855, UK

M. Mullick
Department of Psychiatry
Bangabandhu Sheikh Mujib
Medical University
Dhaka-1000, Bangladesh

Abstract A computerised algorithm was developed to predict child psychiatric diagnoses on the basis of the symptom and impact scores derived from Strengths and Difficulties Questionnaires (SDQs) completed by parents, teachers and young people. The predictive algorithm generates ''unlikely'', ''possible'' or ''probable'' ratings for four broad categories of disorder, namely conduct disorders, emotional disorders, hyperactivity disorders, and any psychiatric disorder. The algorithm was applied to patients attending child mental health clinics in Britain (N = 101) and Bangladesh (N = 89). The level of chance-corrected agreement between SDQ prediction and an independent clinical diagnosis was substantial and highly significant (Kendall's tau b between 0.49 and 0.73; $p < 0.001$). A ''probable'' SDQ prediction for any given disorder correctly identified 81–91% of the children who definitely had that clinical diagnosis. There were more false positives than false negatives, i.e. the SDQ categories were over-inclusive. The algorithm appears to be sufficiently accurate and robust to be of practical value in planning the assessment of new referrals to a child mental health service.

Key words Children – mental health – psychiatric diagnosis – questionnaire-prediction

Introduction

The Strengths and Difficulties Questionnaire (SDQ) is a brief behavioural screening questionnaire that asks about 25 attributes, some positive and others negative. Despite being brief enough to fit on one piece of paper, the SDQ functions as well as longer-established questionnaires (3, 7). The 25 SDQ items are divided between five scales of five items each, generating scores for conduct problems, hyperactivity-inattention, emotional symptoms, peer problems and prosocial behaviour; all but the last one are summed to generate a total difficulties score. An informant-rated version of the SDQ can be completed by either the parents or teachers of 4–16 year olds (3), while a self-report version of the SDQ can be completed by 11–16 year olds themselves (6). The extended version of the SDQ includes not just the 25 items on symptoms and positive attributes but also an impact supplement that asks whether the respondent thinks that the child or teenager has a problem, and if so, enquires further about overall distress, social impairment, burden and chronicity. For clinicians and researchers with an interest in psychiatric caseness and the determinants of service use, the impact supplement appears to provide useful additional information without taking up much more of respondents' time (4).

Diagnostic predictions from questionnaire data are likely to be more accurate if they draw both on impact and symptom scores, and if they use information from multiple informants. Both DSM-IV (1) and the research version of ICD-10 (9) define most of the common child psychiatric disorders in terms of impact as well as symptoms. The operational criteria for these disorders

stipulate that symptoms must result either in substantial distress for the child or in significant impairment in the child's ability to fulfil normal role expectations in everyday life. Defining disorders solely in terms of symptoms results in implausibly high caseness rates, with most of the supposed cases not being significantly socially impaired by their symptoms, not seeming in need of treatment, and not corresponding to what clinicians would normally recognise as cases (2). A previous study showed that the SDQ impact score was a better guide to psychiatric caseness than SDQ symptom scores, and suggested the use of predictive algorithms based on a combination of SDQ impact and symptom scores (4).

There are several reasons why it is valuable to draw on information from multiple informants when making a child psychiatric disorder. Firstly, hyperkinesis and ADHD can only be diagnosed when there is evidence that the symptoms are present in two or more settings, usually home and school. Secondly, other behavioural problems may be highly situational, e.g. severe conduct problems may be present at school but not at home, or vice versa. Although it is possible to question parents about whether teachers have complained of problems at school, it is clearly preferable not to rely on such "hearsay" evidence but to collect information directly from teachers as well as parents. Young people's self-reports can provide a valuable third source of information. For example, teenagers may describe worries or antisocial activities that they have successfully hidden from the adults around them. For all these reasons, diagnostic predictions generated by the SDQ are more likely to be accurate when based on multiple informants.

The present investigation examines whether an algorithm that predicts diagnosis on the basis on multi-informant SDQ symptom and impact scores can be sufficiently accurate in a clinic sample to be of practical use. The algorithm was designed to identify three broad-band diagnoses, namely conduct disorders, emotional disorders and hyperactivity disorders. If it could do so accurately enough in advance of the first clinical assessment at a child mental health clinic, the clinic could use the SDQ diagnostic predictions to influence their decisions on how and when to see the child. For example, hyperkinesis (9) or attention deficit hyperactivity disorder (ADHD) (1) affect a substantial minority of children referred with disruptive behaviour, and this minority is particularly likely to benefit from the involvement of a child psychiatrist or paediatrician trained in the recognition and pharmacotherapy of hyperactivity disorders. Consequently, if it were possible to use the SDQ to predict in advance which disruptive children were particularly likely to have hyperactivity disorders, medically qualified professionals could be involved in their assessment from the outset. Children with "pure" conduct disorders are less likely to require the involvement of medically qualified professionals from the start.

The present study started as an evaluation of the accuracy of the predictive algorithms in a child psychiatric clinic in London. However, the algorithm was slightly modified as a result of the experience gained during the study, and this made it possible that the SDQ's good predictive power in London was capitalising on chance and would not be replicable. (Algorithms are always likely to work best in the sample on which they were developed.) It was for this reason that we subsequently extended the study to Dhaka with no further changes in the algorithm. If the algorithm continued to work well when the questionnaire was administered in a very different language and cultural context, this replication would suggest that the predictive properties of the diagnostic algorithm were robust.

Method

London sample

A multidisciplinary child mental health clinic in outer London administered the SDQ routinely to parents, teachers and 11–16 year olds prior to the first clinic assessment. The present sample consists of a consecutive sample of 101 new referrals aged 4–16 years who had both parent and teacher SDQs completed, and who were subsequently assessed at the clinic and found not to have either an autistic or a psychotic disorder. Children with an autistic or psychotic disorder were excluded for three reasons. Firstly, the SDQ is clearly focused on common forms of psychopathology and does not include the sorts of questions that would allow the recognition of autistic or psychotic disorders with confidence. Secondly, it is generally easy to recognise children at risk of psychosis or autism from the referral letter, so there would be little additional merit in predicting these disorders from prior SDQs even if this were possible. Thirdly, new referrals with these disorders are relatively rare in district clinics, so numbers would have been insufficient for statistical analyses. While parent and teacher SDQs were completed on all subjects, the self-report SDQ was only available on 23 (55%) of the 42 subjects aged between 11 and 16. The mean age of the sample was 10.3 years (SD 3.2 years) and 80 (79%) were male. On the basis of the clinical diagnoses described below, 48 (48%) had a conduct disorder, 21 (21%) had an emotional disorder, 36 (36%) had hyperkinesis, and 26 (26%) had none of these; the percentages add up to considerably more than 100% because of comorbidity, particularly between conduct disorder and hyperkinesis.

Dhaka sample

A multidisciplinary child mental health clinic in Dhaka, the largest city in Bangladesh, administered the SDQ to parents, teachers and 11–16 year olds at the time of the first assessment. The questionnaires were administered as a structured interviews when the respondents' literacy skills were insufficient for them to complete the questionnaire directly. The present sample consists of a consecutive sample of 89 new referrals aged 4–16 years who had both parent and teacher SDQs completed, and who did not have either an autistic or a psychotic disorder. While parent and teacher SDQs were completed on all subjects, the self-report SDQ was only available on 36 (62%) of the 58 subjects aged between 11 and 16. The mean age of the sample was 12.4 years (SD 3.5 years) and 48 (54%) were male. On the basis of the clinical diagnoses described below, 21 (24%) had a conduct disorder, 46 (52%) had an emotional disorder, 9 (10%) had hyperkinesis, and 18 (20%) had none of these; the percentages add up to slightly more than 100% because of comorbidity, particularly between conduct disorder and hyperkinesis.

Questionnaires

The London clinic used the original English SDQs designed for adult informants and self-report; questionnaires included the impact supplement and were scored in the standard manner (3, 4, 6). Samples of the questionnaires and descriptions of their psychometric properties are available from published papers (3, 4, 6, 7) and from the SDQ web site (http://www.iop.kcl.ac.uk/Iop/Departments/ChildPsy/sdq/a1.stm). The Dhaka clinic used translations of the same questionnaires into Bangla. A panel of expert translators carried out the translation, whose accuracy was subsequently checked by independent back-translation. Great care was taken to ensure that the translation was culturally sensitive, using only those words and idioms that would readily be understood by all Bangla-speakers irrespective of their social or educational backgrounds.

Diagnostic algorithm

The predictive algorithm generates "unlikely", "possible" or "probable" ratings for four categories of disorder, namely conduct disorders, emotional disorders, hyperactivity disorders, and any psychiatric disorder. The original a priori algorithm predicted that a disorder was probably present on the basis that the relevant symptom score was above the 95th centile and the impact score was two or more (representing a minimum of "quite a lot" of impact in two domains or "a great deal" of impact in one domain). The criteria could be met according to just one rater for the prediction of conduct or emotional disorders but had to be met according to parent and teacher for the prediction of a hyperactivity disorder. These a priori criteria were adjusted slightly in the light of the initial findings on the London sample. The most marked change was made when it became apparent that many children with conduct or hyperactivity disorders who did not have a comorbid emotional disorder according to the clinical assessment were nevertheless predicted to have probable emotional disorders according to the SDQ. This over-diagnosis of probable emotional disorders was reduced by including an additional diagnostic criterion, namely that if the algorithm predicted that the child had a conduct or hyperactivity disorder, it would only predict a probable emotional disorder if the criteria were met according to at least two informants. Whereas the previously mentioned criteria form the basis for a prediction of "probable" disorder, more inclusive criteria form the basis for a prediction of "possible" disorder, e.g. lower symptom scores, lesser impact, lack of pervasiveness. The algorithm, as programmed in SAS version 6.12 (SAS Institute Inc.), is available from the SDQ web site (http://www.iop.kcl.ac.uk/Iop/Departments/ChildPsy/sdq/a1.stm). The algorithm is also incorporated into the scoring and report-writing computer program available from the web site.

Clinical diagnosis

In both London and Dhaka, children were assigned ICD-10 clinical diagnoses (9) on the basis of detailed information on symptoms and resultant impairments gathered from multiple informants. These diagnoses were collapsed into three broad categories to provide cell sizes that would be sufficient for meaningful analysis. These categories were hyperkinesis, conduct disorder (including oppositional disorder) and emotional disorder (including anxiety, depressive and obsessive compulsive disorders). For all subjects, each disorder was rated as absent, borderline or present by an experienced child psychiatrist who was blind to the SDQ findings. In the London clinic, these diagnostic ratings were made by one of the authors (DR) after reviewing the extensive clinical case notes that include a diagnostic formulation as well as detailed accounts of the semi-structured enquiries made of multiple informants. Another of the authors (RG) independently rated the first 20 subjects; the level of agreement between the two raters was such that the kappa coefficients were 0.92 for hyperkinesis, 1.0 for conduct disorder, and 0.67 for emotional disorder. In Dhaka, the clinical diagnosis was rated at the time of the initial assessment by the senior clinician involved in the assessment (MM), using the pheno-

menologically based approach acquired during specialist training in London with another of the authors (RG).

Results

Table 1 presents the cross-tabulation of clinical diagnosis and SDQ predictions for the London sample, considering conduct, emotional and hyperactivity diagnoses separately. Table 2 presents equivalent cross-tabulations for the Dhaka sample. For each disorder in each clinic, the level of chance-corrected agreement between clinical diagnosis and SDQ prediction was substantial and highly significant (Kendall's tau b between 0.49 and 0.73; $p < 0.001$).

The same pattern was evident for both clinics and for all three types of disorder. When a child had a definite clinical diagnosis of some particular disorder, the SDQ prediction for that disorder was "probable" for the great majority of these children and "possible" for practically all the rest. Serious false negatives – children who had a definite disorder but who were rated "unlikely" by the SDQ algorithm – were rare (1 child in London, none in Dhaka). There were more false positives than false negatives. Nevertheless, for all three disorders in the London sample, and for two of the three disorders in the Dhaka sample, over half of the children rated as "probable" by the SDQ algorithm did indeed have the relevant disorder according to clinicians, i.e. the prediction of "probable" seemed justified. However, in the case of children predicted to have probable hyperkinesis in Dhaka, only just over a third did.

It is only possible to describe screening efficiency in terms of specificity, sensitivity, positive predictive value and negative predictive value when both the screen and the 'gold standard' are dichotomous variables (i.e. positive or negative). In order to present the screening efficiency of the SDQ in these terms, the algorithm's

Table 1 London clinic (N = 101): cross-tabulation of clinical diagnosis and SDQ prediction

a

Conduct disorder	Clinical diagnosis		
	Not conduct disorder	Borderline conduct disorder	Definite conduct disorder
SDQ predicts that conduct disorder is:			
Unlikely	11	5	1
Possible	3	6	4
Probable	3	25	43

Kendall's tau b = 0.50; $p < 0.001$

b

Emotional disorder	Clinical diagnosis		
	Not emotional disorder	Borderline emotional disorder	Definite emotional disorder
SDQ predicts that emotional disorder is:			
Unlikely	20	4	0
Possible	23	17	4
Probable	4	12	17

Kendall's tau b = 0.55; $p < 0.001$

c

Hyperkinesis	Clinical diagnosis		
	Not hyperkinesis	Borderline hyperkinesis	Definite hyperkinesis
SDQ predicts that hyperkinesis is:			
Unlikely	17	5	0
Possible	13	16	4
Probable	1	13	32

Kendall's tau b = 0.67; $p < 0.001$

Table 2 Dhaka clinic (N = 89): cross-tabulation of clinical diagnosis and SDQ prediction

a

Conduct disorder	Clinical diagnosis		
	Not conduct disorder	Borderline conduct disorder	Definite conduct disorder
SDQ predicts that conduct disorder is:			
Unlikely	45	0	0
Possible	10	1	3
Probable	7	5	18

Kendall's tau b = 0.68; $p < 0.001$

b

Emotional disorder	Clinical diagnosis		
	Not emotional disorder	Borderline emotional disorder	Definite emotional disorder
SDQ predicts that emotional disorder is:			
Unlikely	20	6	0
Possible	7	3	4
Probable	2	5	42

Kendall's tau b = 0.73; $p < 0.001$

c

Hyperkinesis	Clinical diagnosis		
	Not hyperkinesis	Borderline hyperkinesis	Definite hyperkinesis
SDQ predicts that hyperkinesis is:			
Unlikely	56	0	0
Possible	9	0	1
Probable	14	1	8

Kendall's tau b = 0.49; $p < 0.001$

Table 3 Screening efficiency (predicting definite clinical disorders from algorithm ratings of 'probable')

	Conduct		Emotional		Hyperactivity	
	London	Dhaka	London	Dhaka	London	Dhaka
Sensitivity	90%	86%	81%	86%	89%	89%
Specificity	47%	82%	80%	84%	78%	81%
PPV	68%	60%	52%	86%	70%	35%
NPV	83%	95%	94%	90%	93%	98%

PPV = positive predictive value; NPV = negative predictive value

predictions and the clinical ratings were both reduced from trichotomies to dichotomies. 'Probable' predictions of the algorithm were counted as positive, whereas 'unlikely' and 'possible' predictions were both counted as negative. Likewise, clinical diagnoses of definite disorders were counted as positive, whereas clinical diagnoses of no disorder or borderline disorder were counted as negative. It is worth noting that many 'false negatives' were children with definite disorders who were rated as 'possible' by the algorithm. In other words, many false negatives were partial rather than complete. The indices of screening efficiency are presented separately for the two clinic samples in Table 3. These indices confirm that the algorithm is good at detecting disorders (sensitivities from 81 to 90%) but at the expense of being over-inclusive (positive predictive values from 35 to 86%).

Discussion

An algorithm drawing on multi-informant SDQ symptom and impact scores was able to predict broad-band psychiatric diagnoses in child mental health clinics with a fair degree of accuracy. Parent and teacher SDQs were available in all cases, but self-report SDQs were only completed on just over half of all eligible subjects; this may have reduced predictive accuracy. Although the independent clinical diagnoses that served as the 'gold standard' were made by experienced clinicians on the basis on detailed information from multiple informants, the process did not involve standardised and validated psychiatric interviews – a limitation that will need to be addressed in future studies. Nevertheless, the initial results are encouraging.

When the SDQ predicted that a specific type of disorder was probable, over half did indeed have that disorder (except for predictions of hyperkinesis in Dhaka, where only just over a third did turn out to have the disorder). Perhaps more importantly, relatively few children with definite disorders did not receive a "probable" rating. So although the SDQ predictions were over-inclusive, they functioned well as a message saying "Think carefully about this particular type of disorder since it is present in over half of all clinic attenders with a similar SDQ profile." This is comparable to a message on a haematology result reminding you, if there is a macrocytic anaemia, to think about B12 and folate deficiency.

The level of accuracy of the SDQ predictions could potentially be clinically useful. Judging from the present study, SDQs administered before the first clinic appointment could potentially have predicted about 90% of the children who would subsequently be diagnosed as hyperkinetic. Consequently, if a clinic allocated children with SDQ-predicted hyperkinesis to members of the team particularly specialised in the assessment and treatment of hyperactivity disorders, then relatively few children with hyperkinesis would be allocated to less appropriate team members in the first instance. Knowing in advance that a particular child is at high risk of hyperkinesis could also help relatively inexperienced clinicians avoid the trap of focusing excessively on a child's conduct problems while failing to notice the comorbid hyperkinesis. Predicting comorbid emotional disorders in advance of the first clinic assessment may be similarly useful. Future studies could examine whether using the algorithm does increase a clinic's efficiency and diagnostic sensitivity.

The finding that a predictive algorithm developed for use in England worked well in Bangladesh despite the differences in language, culture and socio-economic circumstances provides strong evidence for the cross-cultural robustness of both the SDQ and the predictive algorithm. It also provides the first evidence for the validity of the Bangla translation of the SDQ. Since the economic circumstances of Bangladesh preclude the routine use of computers in clinics at present, we have gone on to develop a simpler paper-and-pencil version of the algorithm that is more suitable for local circumstances (8).

Although the present study demonstrates that the algorithm based on multi-informant SDQs can predict diagnoses in clinic samples with a useful degree of accuracy, this does not demonstrate that the same algorithm is equally useful when applied to community samples. In a community sample, psychiatric disorders are rarer (which will tend to increase the rate of false positives) and milder (which will tend to increase the rate of false negatives). A subsequent study comparing SDQ predictions with independent psychiatric diagnoses in a large community study has indeed shown that the algorithm is not as successful at predicting specific groups of disorders in the community (5).

Acknowledgements We are very grateful to all participating children, parents and teachers, and also to the entire staff of the two clinics involved.

References

1. American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders (4th edn). American Psychiatric Association, Washington, DC
2. Bird HB, Yager TJ, Staghezza B, Gould MS, Canino G, Rubio-Stipec M (1990) Impairment in the epidemiological measurement of childhood psychopathology in the community. Journal of the American Academy of Child and Adolescent Psychiatry 29:796–803
3. Goodman R (1997) The strengths and difficulties questionnaire: a research note. Journal of Child Psychology and Psychiatry 38:581–586
4. Goodman R (1999) The extended version of the strengths and difficulties questionnaire as a guide to child psychiatric caseness and consequent burden. Journal of Child Psychology and Psychiatry 40:791–799
5. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H (2000) Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample, submitted for publication.
6. Goodman R, Meltzer H, Bailey V (1998) The strengths and difficulties questionnaire: a pilot study of the validity of the self-report version. European Child and Adolescent Psychiatry 7:125–130
7. Goodman R, Scott S (1999) Comparing the strengths and difficulties questionnaire and the child behavior checklist: is small beautiful? Journal of Abnormal Child Psychology 27:17–24
8. Mullick MSI, Goodman R (2000) Is questionnaire screening for child mental health problems feasible in the developing world? A preliminary study applying the Strengths and Difficulties Questionnaire (SDQ) to community and clinic samples of children in Bangladesh. submitted for publication.
9. World Health Organization (1994) The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research. World Health Organization, Geneva