

## Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence

Anna Goodman · Einar Heiervang · Bacy Fleitlich-Bilyk ·  
Abdulla Alyahri · Vikram Patel · Mohammad S. I. Mullick ·  
Helena Slobodskaya · Darci Neves dos Santos · Robert Goodman

Received: 3 August 2010 / Accepted: 7 October 2011 / Published online: 28 October 2011  
© Springer-Verlag 2011

### Abstract

**Purpose** To examine whether the widely used Strengths and Difficulties Questionnaire (SDQ) can validly be used to compare the prevalence of child mental health problems cross nationally.

**Methods** We used data on 29,225 5- to 16-year olds in eight population-based studies from seven countries: Bangladesh, Brazil, Britain, India, Norway, Russia and Yemen. Parents completed the SDQ in all eight studies, teachers in seven studies and youth in five studies. We used these SDQ data to calculate three different sorts of “caseness indicators” based on (1) SDQ symptoms, (2) SDQ symptoms plus impact and (3) an overall respondent judgement of ‘definite’ or ‘severe’ difficulties. Respondents also completed structured diagnostic interviews including extensive open-ended questions (the Development and Well-Being

Assessment, DAWBA). Diagnostic ratings were all carried out or supervised by the DAWBA’s creator, working in conjunction with experienced local professionals.

**Results** As judged by the DAWBA, the prevalence of any mental disorder ranged from 2.2% in India to 17.1% in Russia. The nine SDQ caseness indicators (three indicators times three informants) explained 8–56% of the cross-national variation in disorder prevalence. This was insufficient to make meaningful prevalence estimates since populations with a similar measured prevalence of disorder on the DAWBA showed large variations across the various SDQ caseness indicators.

**Conclusions** The relationship between SDQ caseness indicators and disorder rates varies substantially between populations: cross-national differences in SDQ indicators do not necessarily reflect comparable differences in disorder rates. More generally, considerable caution is required when interpreting cross-cultural comparisons of mental health, particularly when these rely on brief questionnaires.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00127-011-0440-2) contains supplementary material, which is available to authorized users.

A. Goodman (✉) · V. Patel  
Faculty of Epidemiology and Population Health,  
London School of Hygiene and Tropical Medicine,  
Keppel Street, London WC1E 7HT, UK  
e-mail: anna.goodman@lshtm.ac.uk

E. Heiervang  
Centre for Child and Adolescent Mental Health, Bergen, Norway

E. Heiervang  
Institute of Clinical Medicine, University of Oslo, Oslo, Norway

B. Fleitlich-Bilyk  
University of São Paulo, São Paulo, Brazil

A. Alyahri  
Hadramaut University, Mukalla, Hadramaut, Yemen

M. S. I. Mullick  
Bangabandhu Sheikh Mujib Medical University, Dhaka,  
Bangladesh

H. Slobodskaya  
Siberian Branch of the Russian Academy of Medical Sciences,  
Novosibirsk State University, Novosibirsk, Russia

D. N. dos Santos  
Federal University of Bahia, Salvador, Brazil

R. Goodman  
King’s College London Institute of Psychiatry, London, UK

**Keywords** Cross cultural · Questionnaire · Prevalence · Reporting bias

## Introduction

A seminal finding of twentieth century epidemiology was that a population's mean predicts the proportion of high-scoring ('deviant') individuals. This was first demonstrated for physical health [1] and has recently been reported for mental health in adult populations across Europe [2] and in child populations within Great Britain [3].

These findings highlight the importance of implementing population-wide interventions alongside interventions which target the highest-risk individuals [1]. They also suggest the possibility of using population mean scores to compare health over time, space or culture. Caution is needed, however, when making such comparisons using subjectively reported outcomes such as mental health. This is because differences in mean scores may not reflect differences in population health but rather systematic bias in how mental health is reported. Such biases may be particularly likely in brief questionnaires which (unlike structured diagnostic interviews), ask only a small number of broad questions and which allow no role for clinical judgement [4, 5].

We have previously shown that, in general, such systematic reporting biases do not seem to apply within Great Britain when using the Strengths and Difficulties Questionnaire (SDQ) [6]. Mean SDQ symptom scores predicted the prevalence of disorder in an accurate and unbiased manner across populations defined by multiple child, family and area characteristics (e.g. ethnicity, family type, and area deprivation) [3]. This was true for the parent, teacher and youth SDQs alike, and allowed us to derive and validate UK 'SDQ prevalence estimators'. For the parent and teacher (but not youth) SDQs, the prevalence of disorder was also closely estimated by (1) the proportion of individuals with high SDQ symptoms plus impact; and (2) the proportion of individuals reporting 'definite' or 'severe' difficulties in a one-item, global rating of child mental health problems.

It would be a great boost to child psychiatric epidemiology if these British findings applied cross nationally, i.e. if the same set of equations could be applied to generate prevalence estimates in and within countries other than Britain. First, it would allow researchers in other settings to treat the SDQ as an accurate and unbiased method for monitoring and comparing child mental health. This could be particularly important in low- and middle-income settings, which frequently lack the money and clinical expertise to conduct prevalence studies using detailed diagnostic interviews and/or to use diagnostic interviews to derive

country-specific prevalence estimating equations. Second, it would greatly facilitate comparisons of child mental health across many different countries, and so aid the identification of population-level determinants of health [7].

Interesting findings regarding cross-cultural similarities and differences in child mental health have already emerged from international comparisons using brief questionnaires [8–10], including the SDQ [11, 12]. Yet, interpreting these findings is substantially complicated by uncertainty about how far these brief questionnaires provide unbiased cross-cultural estimates of disorder prevalence. Several studies indicate that rating norms may differ across cultures [13, 14], providing indirect evidence that brief questionnaires may be problematic. To our knowledge only one study examines this issue directly, demonstrating that differences in mean SDQ scores only sometimes reflected different disorder prevalences in Norway compared to Britain [4]. This paper builds upon this Norway–Britain comparison to examine whether caseness indicators based on the parent, teacher or youth SDQ provide meaningful prevalence estimates cross nationally.

## Methods

### Study samples

Our data come from 29,225 5- to 16-year olds from seven different countries: Britain [15, 16], Norway [17], Brazil [18, 19], Yemen [20], India [21], Bangladesh [22] and Russia [23]. These represent the participants in all published population-based studies which have: administered the parent SDQ; estimated prevalence using a highly comparable form of clinician-rated diagnosis (including shared supervision, as described see below); and based prevalence estimates upon Development and Well-Being Assessment (DAWBA) interviews about at least 100 children.

All these studies have previously been reported in detail individually [15–23] and Table 1 summarises their survey methodology, including sampling procedures and informants used. Four out of eight studies were two-phase, administering the DAWBA to all children who screened positive on the SDQ and to a random subsample of children who screened negative. All studies approached parents for written informed consent to take part and the present analyses include only those children with complete parent SDQ data. With parental permission, 7/8 studies also collected mental health data from teachers (all except India) and 5/8 collected data from youth aged 11–16 (all except Bangladesh, Norway and Yemen). All studies received ethical approval from local and/or UK research ethics committees.

In three studies (from Brazil, Britain and Yemen), we subdivided the study samples a priori into further

**Table 1** Key methodological features and sample characteristics of study populations

Country	Study methodology		Study samples used for present analyses				
	Location, sampling frame and participation rate	One- or two-phase	Informants	Sub-populations	Sample size	Age range (mean)	% male
Britain [15, 16], 1999 and 2004	Stratification and random sampling of postal sectors, followed by random sampling from Child Benefit Register. 69% participation rate.	One-phase	Parents, teachers and youth	Deciles of small-area deprivation, measured using the UK Indices of Multiple Deprivation [47]	18,190; 1,790–1,843 per decile	5–16 ( <i>m</i> = 10.1)	50.7
Norway [17], 2002–2004	Total population sample of all children in three school years in the city of Bergen. 67% participation rate.	Two-phase	Parents and teachers	–	6,658 (1,024 DAWBAs)	8–10 ( <i>m</i> = 9.0)	50.1
Southeast Brazil [18], 2000–2001	Random sampling of schools in Taubaté municipality, stratified into private schools, public rural schools, and public urban schools. Pupils were then randomly sampled from schools. 83% participation rate.	One-phase	Parents, teachers and youth	Middle class in private schools	236	7–14 ( <i>m</i> = 10.1)	47.5
Northeast Brazil [19], 2001	Household census of all children on Ilha de Maré, a small island <i>quilombo</i> (community with particularly strong African–Brazilian heritage). Participating children were then sampled at random. 100% participation rate.	Two-phase	Parents, teachers and youth	Rural	264	7–14 ( <i>m</i> = 10.4)	50.4
Yemen [20], 2002–2003	School-based sampling of children living in the city of Mukalla or the rural area of Tuban. 77% participation rate.	Two-phase urban, one-phase rural	Parents and teachers	Urban slum ( <i>avela</i> ) in urban public schools	745	7–14 ( <i>m</i> = 10.6)	55.6
India [21], 2002–2003	Four rural areas were randomly selected in the Indian state of Goa, along with six out of 16 Wards in the city of Margao. All 12–16 year olds were then approached, being identified by family health registers and door-to-door surveys. 76% participation rate.	One-phase	Parents and youth	–	430 (100 DAWBAs)	7–14 ( <i>m</i> = 10.4)	50.9
Bangladesh [22], 2004	Community samples in or near the city of Dhaka from an urban slum, a prosperous urban area and a nearby rural area. Samples were obtained by random sampling from the electoral registers. 75% participation rate.	One-phase	Parents and teachers	Urban	1,007 (262 DAWBAs)	7–10 ( <i>m</i> = 8.5)	50.1
Russia [23], 2001–2002	Random sampling of 10 schools (stratified by quality) in the city of Novosibirsk, then random sampling of pupils. 83% participation rate.	Two-phase	Parents, teachers and youth	Rural	203	7–10 ( <i>m</i> = 8.2)	51.2
					1,169	12–16 ( <i>m</i> = 13.8)	50.8
					922 (208 DAWBAs)	5–10 ( <i>m</i> = 7.5)	52.6
					447 (172 DAWBAs)	7–14 ( <i>m</i> = 10.5)	48.5

socio-demographic populations. The result was 10 British and 10 non-British populations, the age range and sex composition of which are reported in Table 1.

## Measures

### *Strengths and Difficulties Questionnaire (SDQ)*

The Strengths and Difficulties Questionnaire (SDQ) is a brief questionnaire measure of child mental health problems that can be administered to parents and teachers of children aged 4–16 and to young people aged 11–16 [6, 24]. It contains 20 items covering emotional symptoms, conduct problems, hyperactivity and peer problems, which can be summed to give a ‘total difficulty score’. The total difficulty score is a measure of overall child mental health problems that has been shown to have good psychometric properties in studies from around the world [6, 25–31]. This includes evidence that the total difficulty score is correlated with existing questionnaire and interview measures; differentiates clinic and community samples; and is associated with increasing rates of clinician-rated diagnoses of child mental disorder across its full range.

This paper makes cross-cultural comparisons using three SDQ caseness indicators

- 1) ‘SDQ prevalence estimates’. Within Britain, we have previously derived and validated equations which estimate disorder prevalence based on mean total difficulty scores, adjusting for the population’s age and sex composition [3; prevalence estimator equations in Supplementary material]. We used these prevalence estimates rather than raw mean scores in order to allow for age differences between our study samples.
- 2) SDQ ‘symptoms+impact’. The SDQ impact supplement asks whether reported difficulties cause the child distress (1 item) or impairment in their daily life (4 items for parents and youth, and 2 items for teachers) [29]. We calculated the proportion of children with borderline or high symptoms (total difficulty score cut-points 13/14 for parent SDQ, 11/12 for teacher SDQ, and 15/16 for youth SDQ) plus high impact (impact score cut-point 1/2 for all informants) [32].
- 3) ‘Definite/severe’ difficulties. The SDQ symptom questions are followed by a single item asking whether the child has difficulties with “emotions, concentration, behaviour or being able to get along with other people”. We calculated the proportion of informants reporting ‘definite’ or ‘severe’ difficulties (vs. ‘no’ or ‘minor’ difficulties).

### *Development and Well-being Assessment (DAWBA)*

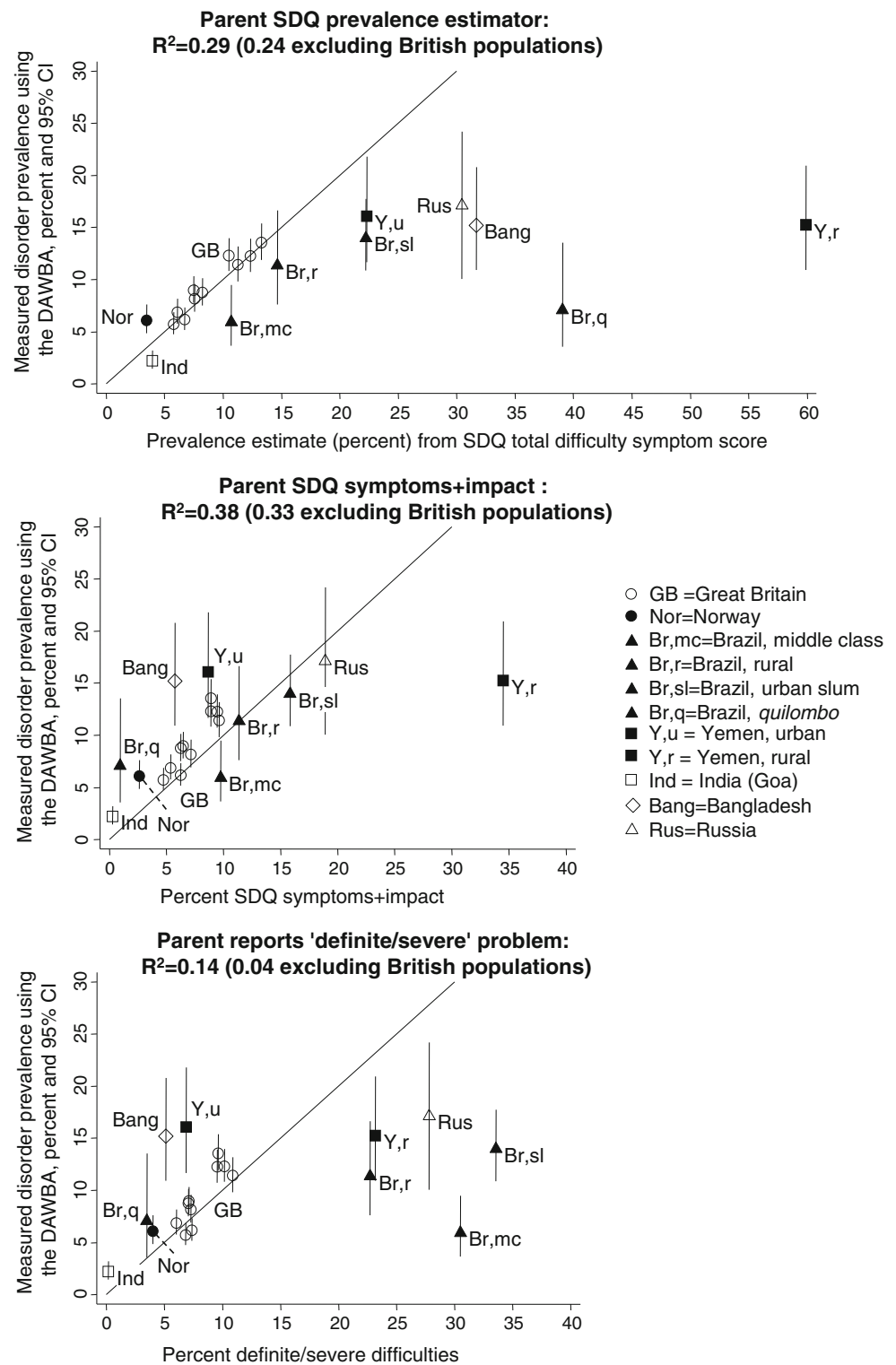
We measured disorder prevalence using the DAWBA. This is a detailed psychiatric interview administered by lay interviewers to parents and youth, and a briefer questionnaire for teachers [33]. The main DAWBA interview is fully structured, closely following the diagnostic criteria operationalised in the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) [34]. Responding parents, teachers and youth are then prompted to describe any reported difficulties in detail, with answers recorded verbatim by the interviewer. Experienced clinicians review the open and closed accounts of all available informants, and rate the presence or absence of individual diagnoses according to DSM-IV [35].

In our eight study populations, the DAWBA diagnoses have been shown to have high inter-rater reliabilities [17, 18, 36, 37], to discriminate clinic and community samples [18, 22, 33, 36], to show plausible patterns of comorbidity and association with risk factors [17, 19, 23, 37, 38], and to be strongly predictive of mental health service contact [17, 39]. All diagnostic ratings were carried out by the DAWBA’s creator (RG) or by experienced local professionals supervised by RG. These experienced local professionals trained initially on the 54 practice cases in the on-line DAWBA manual (<http://www.dawba.info/manual/m0.html>). They were then supervised individually by RG who reviewed a mixture of randomly selected cases and difficult cases that the trainee had provisionally rated.

## Analyses

We calculated all prevalence estimates and confidence intervals using sampling weights to correct for the two-phase design of some studies (see Table 1). We also adjusted for the complex survey design of those studies that used stratification or clustered sampling. We plotted each of our nine SDQ caseness indicators (three measures times three informants) against the measured prevalence of disorder using the DAWBA, deriving the measured prevalence from the same subset of children (e.g. comparing predictors based on teacher SDQs with the prevalence of disorder in children with teacher SDQ data). We fitted nine corresponding linear regression models, with the relevant SDQ caseness indicator as the explanatory variable and giving all study populations equal weight. We present the adjusted  $R^2$  values from these regression models as a measure of how much of the variance in prevalence was explained. All analyses were performed in Stata 10.2

**Fig. 1** Parent SDQ caseness indicators versus prevalence of disorder: data from seven countries

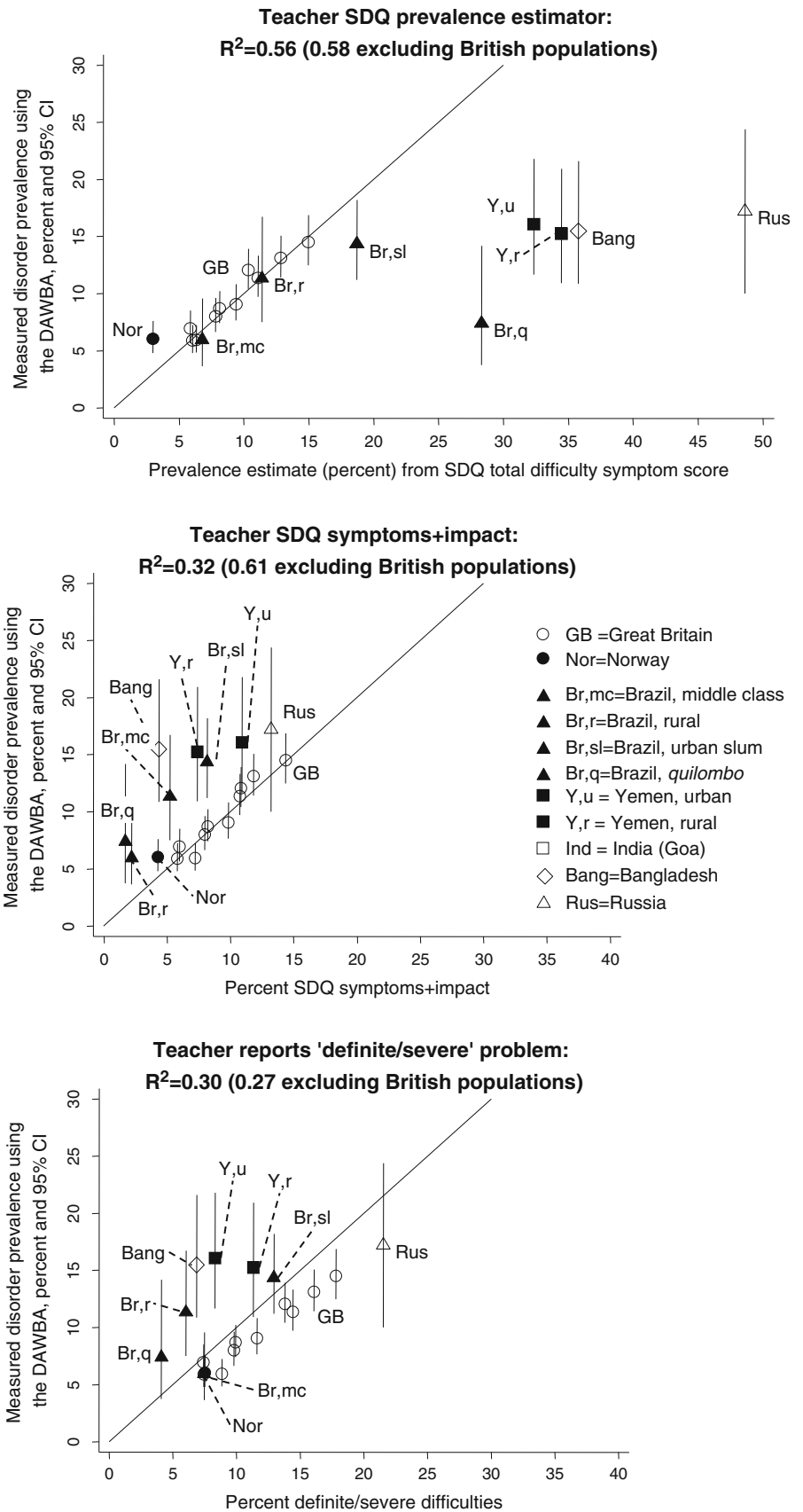


**Results**

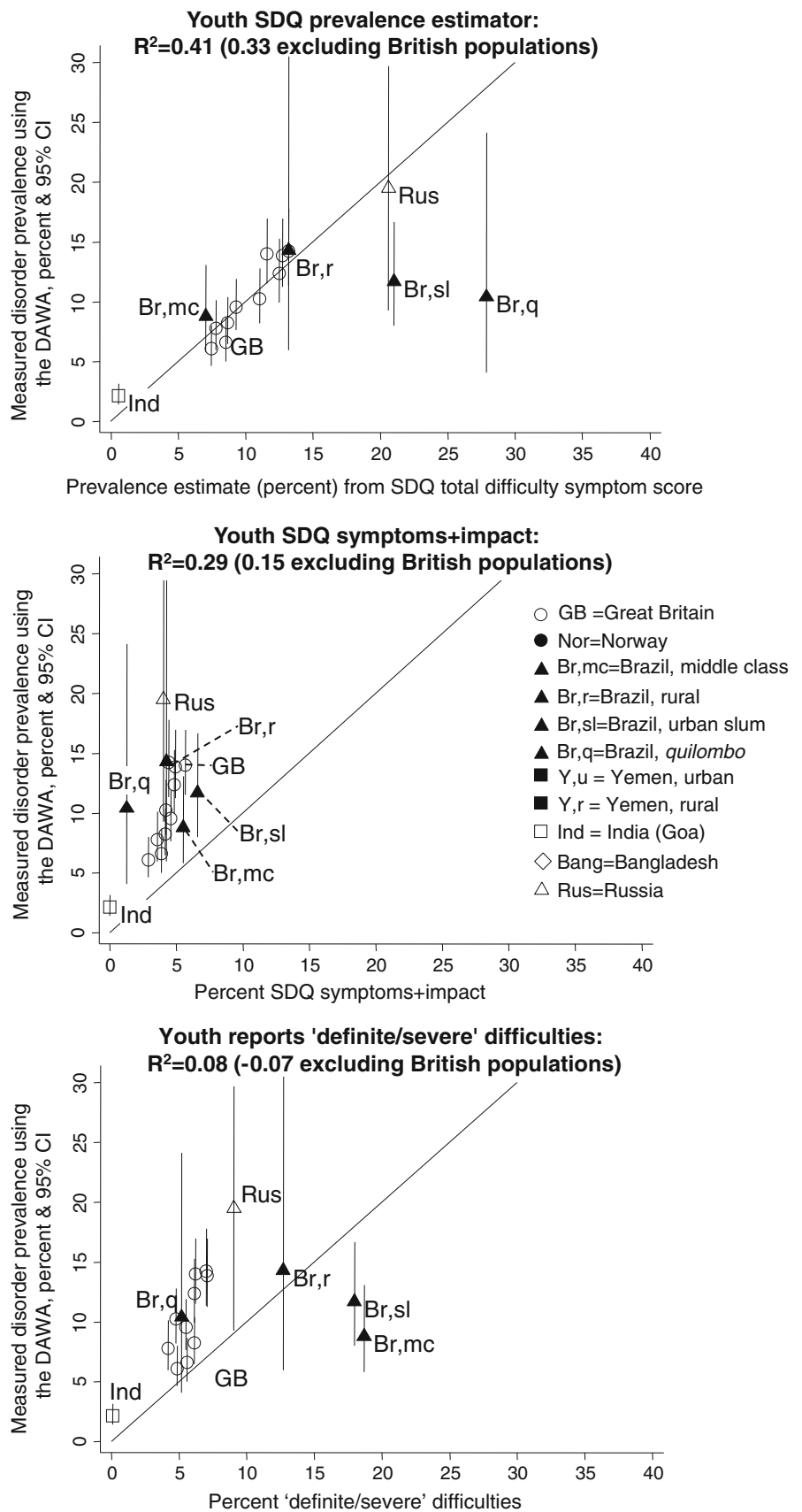
The prevalence of disorder measured using the DAWBA ranged from 2.2% in our Indian sample to 17.1% in our Russian sample. Figure 1 plots these prevalence values

(y-axis), comparing them to the three parent-based SDQ caseness indicators (x-axis) and presenting the  $R^2$  values; Figs. 2, 3 present corresponding graphs for the teacher and youth SDQs. This information is also presented in tables in the Supplementary material, together with the raw mean

**Fig. 2** Teacher SDQ caseness indicators versus prevalence of disorder: data from six countries



**Fig. 3** Youth SDQ caseness indicators versus prevalence of disorder: data from four countries



total difficulty scores upon which the SDQ prevalence estimates are based. The Supplementary material also shows the prevalence rates and relative proportions of emotional, behavioural and hyperactivity disorders; these relative proportions were much less variable than the overall prevalence rates.

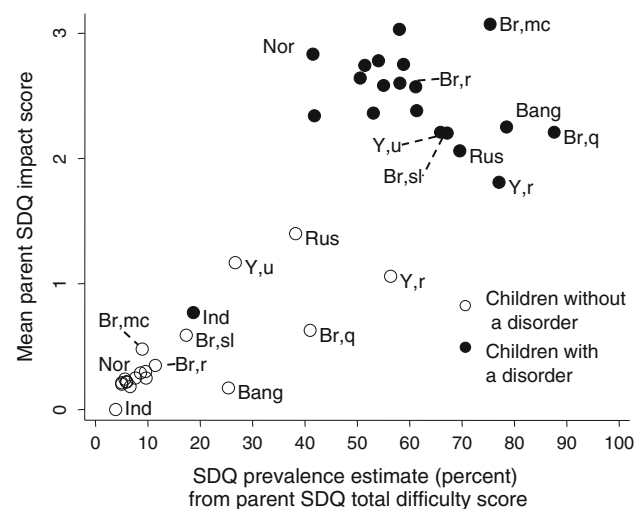
The three parent-based SDQ caseness indicators yielded  $R^2$  values of 0.14–0.38—that is, explaining 14–38% of the observed cross-national variation in the prevalence of disorder ascertained using the DAWBA (see Figures for individual  $R^2$  values). The corresponding  $R^2$  values were 0.30–0.56 for teachers and for 0.08–0.41 for youth. These values were similar when the analyses were repeated separately for study populations aged 5–10 years and for populations aged 11–16 years (see Supplementary material) and generally fell when the British samples were removed. Only within Britain did the SDQ prevalence estimates closely approximate the true prevalence (i.e. lie close to the 45 degree line plotted in the Figures); in most other populations the SDQ prevalence estimator equations overestimated the prevalence, while in Norway they underestimated it.

The result was that none of these SDQ caseness indicators could be used to make meaningful estimates of prevalence across the non-British samples. To illustrate this point, it is useful to consider the performance of the parent SDQ in the 10 populations with the highest measured prevalence of disorder. The *actual* prevalences as measured by the DAWBA ranged from 11 to 17% in these 10 populations (see Fig. 1). By contrast, the *estimated* prevalences from the parent SDQ prevalence estimators were 10–15% in rural Brazil and the four most deprived British population; 22% in urban Yemen and urban slum Brazil; 30–32% in Russia and Bangladesh; and 60% in rural Yemen. The other two parent SDQ caseness indicators did no better, giving values ranging from 5 to 47% for these same populations. An instance of inaccurate prediction affecting a population with a low prevalence of disorder was seen in the Northeastern Brazilian *quilombo* (predominantly African–Brazilian rural area): this had a parent SDQ prevalence estimate of 39%, as compared to a measured prevalence of 7%.

Populations with a similar measured prevalence of disorder therefore showed large variations in the parent SDQ caseness indicators. The same was true for the teacher and youth SDQs, as shown in Figs. 2, 3. Furthermore, the relative ordering of populations was not consistent across these measures. For example, in Bangladesh the parent SDQ prevalence estimate was 32%, reflecting a high level of symptoms reported by parents. Yet only around 5% of the Bangladeshi children had SDQ symptoms+impact or were reported by their parents to have ‘definite/severe’ difficulties, among the lower values in the sample. The

Brazilian *quilombo* likewise had one of the highest parent SDQ prevalence estimates (39%) but only 1–3% had symptoms+impact or ‘definite/severe’ difficulties. The teacher and youth SDQ produced similar findings. These discrepancies suggested cross-cultural variation in the relationship between symptoms and impact within the SDQ. To investigate this, we plotted mean parent SDQ impact scores against the SDQ prevalence estimates—that is, against age-adjusted parent SDQ symptoms. As Fig. 4 shows, Bangladesh and *quilombo* Brazil stand out in having unusually low impact scores at a given level of symptoms. The same was true of rural Yemen, where mean parent impact scores were slightly lower than urban Yemen, but the SDQ prevalence estimates were much higher.

Indeed, parent SDQ symptom scores were so high in rural Yemen that the population mean of non-disordered children was comparable to that of children with a disorder in Britain. The converse was true of the final notably anomalous population in Fig. 4, namely the 26 Indian children with a disorder. These children had mean levels of parent-reported symptoms and impact which were far lower than disordered children in any other population ( $p \leq 0.003$ ), and indeed lower than non-disordered children in Russia and Yemen. This was replicated for the youth SDQ, where again the Indian children with a disorder had mean SDQ symptom and impact scores which were indistinguishable from non-disordered samples in most other populations (see Supplementary material for teacher and youth graphs).



**Fig. 4** Estimated prevalence of disorder from parent SDQ versus mean impact score on parent SDQ, stratified by disorder status. *Bang* Bangladesh; *Br,mc* Brazil, middle class; *Br,r* Brazil, rural; *Br,sl* Brazil, urban slum; *Br,q* Brazil, *quilombo*; *Ind* India; *Rus* Russia; *Nor* Norway; *Y,u* Yemen, urban; *Y,r* Yemen, rural; *Unlabelled points* Great Britain



## Discussion

This study of 29,225 5- to 16-year olds from seven countries has examined whether measures based on the parent, teacher or youth SDQ can be used to estimate the prevalence of child mental disorder cross nationally without the need for population-specific norms. Our findings suggest that this is not possible, and that population-specific norms may be needed when estimating prevalence. Our findings also imply the need for substantial caution when interpreting cross-cultural comparisons of levels of child mental health problems which are based solely upon brief questionnaires.

When interpreting these findings, it is worth bearing in mind the limitations of our study. First, our study populations had different age ranges. However, the low correlations between the SDQ measures and the DAWBA were almost unchanged after stratifying by age, suggesting that this cannot explain the large cross-national discrepancies observed. A second limitation is that although all studies collected mental health data from parents, one study did not include teachers and three did not include youth. This undermines comparability because multi-informant DAWBA information generates slightly higher prevalence estimates (e.g. clinicians in Britain were 6% more likely to diagnose a disorder if teachers completed a DAWBA as well as parents [15]). Again, however, these effects are too small to plausibly affect our substantive conclusions. Finally, the DAWBA-generated prevalence figures are themselves only estimates of the true prevalence. Despite our efforts to standardize ratings through shared training and supervision, the DAWBA diagnoses are themselves subject to measurement error, some of which may be systematic across countries. Nevertheless, we believe that the DAWBA's use of multiple detailed questions, open-ended transcripts and local clinical judgment all render it less prone to cross-cultural bias than the SDQ [4]. Moreover, any bias in the DAWBA cannot plausibly account for the extremely large cross-national differences we observed in the SDQ.

We are therefore confident in our substantive conclusion that the SDQ shows large cross-cultural reporting effects and cannot be assumed a priori to be a valid method for comparing prevalences cross nationally without recourse to population-specific norms. Of course, brief questionnaires may nonetheless be important in monitoring mental health or examining risk factor associations. Moreover, cross-cultural bias *between* countries does not necessarily translate into cross-cultural bias *within* a country. For example, despite the differences between the Indian and the British studies in this paper, the SDQ and DAWBA have very similar psychometric properties between British Indians and British Whites [40]. More broadly, within Britain the parent, teacher and youth SDQs generally provide accurate

and unbiased prevalence estimates for populations defined by multiple child, family and area characteristics [3].

Yet what our findings do indicate is that population-specific SDQ norms may be necessary for valid international comparisons. Moreover, it cannot necessarily be assumed that the same norms will always apply within a single country. For example, parent SDQ symptom scores were far higher in rural Yemen than urban Yemen, despite similar disorder prevalences and SDQ impact scores. Much the same was true comparing the Northeastern Brazilian *quilombo* with the Southeastern Brazilian populations. One possible explanation is that in relatively isolated rural communities, respondents have little experience of completing questionnaires, and may find it hard to know what level of symptoms the investigators are interested in [19]. In Yemen, rural parents may also show lower tolerance for problematic child behaviour than urban parents. This would be consistent with our previous demonstration that harsh physical punishment is particularly common in rural Yemen, perhaps reflecting a higher work burden and reduced childcare support [41]. Thus, SDQ symptom scores may be higher when respondents have little familiarity with questionnaires and perhaps when stressful life circumstances reduce tolerance for troubled children. We believe both factors may partly explaining why, relative to British norms, the SDQ tended to overestimate the prevalence of disorder in all our low and middle-income country samples. Only Norway showed an effect in the opposite direction, possibly reflecting a more 'normalizing' attitude towards some child mental health problems [4].

One final striking cross-national anomaly was the low SDQ symptom *and* impact scores of children with a DAWBA diagnosis in Goa, India. This could reflect a cross-national rating bias, such that the threshold for assigning DAWBA diagnoses was lower in India than elsewhere. This, however, would imply that the true prevalence in our Indian sample was even lower than the (already exceptionally low) 2.2% recorded. Instead the judgement of the experienced local adolescent psychiatrist (VP) is that Indian informants were understating child mental health symptoms and impact. This counterpoint to the overstatement hypothesised in rural Yemen again highlights the importance of using local cultural and linguistic knowledge when reading the DAWBA transcripts and interpreting responses to structured questions.

To summarise, this paper uses a uniquely rich dataset to demonstrate substantial cross-cultural differences in how parents report child mental health problems on the SDQ. Our findings also demonstrate that these cross-cultural differences take many different forms, and do not show any obvious systematic pattern. We conclude that the SDQ cannot be used as a short-cut to comparing prevalence cross nationally. Furthermore we hypothesise that this may

also apply to other widely used questionnaires such as the Rutter [42] and the ASEBA [43–45], which are similar to the SDQ in their format, items and psychometric properties [6, 25, 46]. We therefore recommend that questionnaires are only used in cross-cultural comparisons when their cross-cultural equivalence has been empirically demonstrated. Doing so may require detailed diagnostic measurements that employ local and contextual knowledge in order to provide population-specific reference points for judging the performance of brief questionnaire measures.

Such cross-national comparisons based on detailed culturally sensitive assessments will clearly require substantially more time and resources than questionnaire-based studies. Nonetheless, their potential importance is illustrated by the almost eightfold difference between the 2.2% prevalence of child mental disorder in our Indian sample to the 17.1% prevalence in our Russian sample. This is far greater than the variation typically seen within populations from the same country; for example, prevalence ‘only’ varied from 5.7 to 13.5% between the least deprived and most deprived deciles in Britain. Under such circumstances, multi-population studies may yield powerful new insights into the determinants of population health [7]. Understanding international differences in child mental health therefore remains a key research goal in seeking to improve child mental health worldwide, but achieving this may require more than questionnaire comparisons.

**Acknowledgments** The British studies were funded by the Department of Health and the Scottish Executive. The Norwegian study was funded by the City of Bergen, the Research Council of Norway, the Norwegian Directorate for Health and Social Affairs, Western Norway Regional Health Authority, Haukeland University Hospital, Unifob Health, and the Meltzer legacy. The Brazilian, Bangladeshi and Russian studies were funded by the Wellcome Trust. The Indian study was funded by the WHO’s Special Programme of Research, Development and Research Training in Human Reproduction. The Yemeni study was funded by the World Health Organisation and Save the Children Sweden. VP is supported by a Wellcome Trust Senior Clinical Research Fellowship in Tropical Medicine.

**Conflict of interest** AG and RG are directors and RG is the owner of Youthmind, which provides no-cost and low-cost software and web sites related to the SDQ and the DAWBA. Einar Heiervang is the director and owner of Careahead, which provides teaching and supervision to clinics on the use of the SDQ and the DAWBA. BF-B, AA, VP, MM, HS and DnS have no biomedical financial interests or potential conflicts of interests.

## References

- Rose G, Day S (1990) The population mean predicts the number of deviant individuals. *BMJ* 301(6759):1031–1034
- Veerman JL et al (2009) Population prevalence of depression and mean Beck Depression Inventory score. *Br J Psychiatry* 195(6):516–519
- Goodman A, Goodman R (2011) Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *J Child Psychiatry Psychol* 52(1):100–108
- Heiervang E, Goodman A, Goodman R (2008) The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. *J Child Psychol Psychiatry* 49(6):678–685
- Breslau N (1987) Inquiring about the bizarre: false positives in Diagnostic Interview Schedule for Children (DISC) ascertainment of obsessions, compulsions, and psychotic symptoms. *J Am Acad Child Adolesc Psychiatry* 26(5):639–644
- Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 38(5):581–586
- Rose G (1985) Sick individuals and sick populations. *Int J Epidemiol* 14(1):32–38
- Achenbach TM et al (1990) Epidemiological comparisons of Puerto-Rican and United-States Mainland children—parent, teacher, and self-reports. *J Am Acad Child Adolesc Psychiatry* 29(1):84–93
- Weine AM, Phillips JS, Achenbach TM (1995) Behavioral and emotional-problems among Chinese and American children—parent and teacher reports for ages 6 to 13. *J Abnorm Child Psychol* 23(5):619–639
- Crijnen AAM, Achenbach TM, Verhulst FC (1999) Problems reported by parents of children in multiple cultures: The Child Behavior Checklist Syndrome Constructs. *Am J Psychiatry* 156(4):569–574
- Becker A et al (2006) Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. *Eur Child Adolesc Psychiatry* 15:56–62
- Obel C et al (2004) The Strengths and Difficulties Questionnaire in the Nordic countries. *Eur Child Adolesc Psychiatry* 13(Suppl 2):II32–II39
- Luk S (1996) Cross cultural/ethnic aspects, in hyperactivity disorders of childhood. In: Sandberg S (ed). Cambridge University Press, Cambridge
- Mann EM et al (1992) Cross-cultural differences in rating hyperactive-disruptive behaviors in children. *Am J Psychiatry* 149(11):1539–1542
- Meltzer H et al (2000) Mental health of children and adolescents in Great Britain. The Stationery Office, London
- Green H et al (2005) Mental health of children and young people in Great Britain, 2004. Palgrave MacMillan, Basingstoke
- Heiervang E et al (2007) Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. *J Am Acad Child Adolesc Psychiatry* 46(4):438–447
- Fleitlich-Bilyk B, Goodman R (2004) Prevalence of child and adolescent psychiatric disorders in southeast Brazil. *J Am Acad Child Adolesc Psychiatry* 43(6):727–734
- Goodman R et al (2005) The Ilha de Mare study: a survey of child mental health problems in a predominantly African-Brazilian rural community. *Soc Psychiatry Psychiatr Epidemiol* 40(1):11–17
- Alyahri A, Goodman R (2008) The prevalence of DSM-IV psychiatric disorders among 7–10 year old Yemeni schoolchildren. *Soc Psychiatry Psychiatr Epidemiol* 43(3):224–230
- Pillai A et al (2008) Non-traditional lifestyles and the prevalence of mental disorders in young adolescents in a community in India. *Br J Psychiatry* 192:45–51
- Mullick MS, Goodman R (2005) The prevalence of psychiatric disorders among 5–10 year olds in rural, urban and slum areas in Bangladesh: an exploratory study. *Soc Psychiatry Psychiatr Epidemiol* 40(8):663–671
- Goodman R, Slobodskaya H, Knyazev G (2005) Russian child mental health—a cross-sectional study of prevalence and risk factors. *Eur Child Adolesc Psychiatry* 14(1):28–33

24. Goodman R (2001) Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry* 40(11):1337–1345
25. Goodman R, Scott S (1999) Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *J Abnorm Child Psychol* 27(1):17–24
26. Klasen H et al (2000) Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *Eur Child Adolesc Psychiatry* 9(4):271–276
27. Achenbach TM et al (2008) Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. *J Child Psychol Psychiatry* 49(3):251–275
28. Goodman R, Renfrew D, Mullick M (2000) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *Eur Child Adolesc Psychiatry* 9(2):129–134
29. Goodman R (1999) The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *J Child Psychol Psychiatry* 40(5):791–799
30. Mullick MS, Goodman R (2001) Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. *Soc Psychiatry Psychiatr Epidemiol* 36(2):94–99
31. Goodman A, Goodman R (2009) Strengths and difficulties questionnaire as a dimensional measure of child mental health. *J Am Acad Child Adolesc Psychiatry* 48(4):400–403
32. Goodman R, Ford T, Meltzer H (2002) Mental health problems of children in the community: 18 month follow up. *BMJ* 324(7352):1496–1497
33. Goodman R et al (2000) The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 41(5):645–655
34. American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV). American Psychiatric Association, Washington, DC
35. Goodman R et al (1996) Rating child psychiatric caseness from detailed case histories. *J Child Psychol Psychiatry* 37(4):369–379
36. Alyahri A, Goodman R (2006) Validation of the Arabic Strengths and Difficulties Questionnaire and the Development and Well-Being Assessment. *East Mediterr Health J* 12(Suppl 2):S138–S146
37. Ford T, Goodman R, Meltzer H (2003) The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *J Am Acad Child Adolesc Psychiatry* 42(10):1203–1211
38. Goodman A et al (2007) Child, family, school and community risk factors for poor mental health in Brazilian school children. *J Am Acad Child Adolesc Psychiatry* 46(4):448–456
39. Ford T et al (2008) Predictors of service use for mental health problems among British school children. *Child Adolesc Mental Health* 13(1):32–40
40. Goodman A, Patel V, Leon DA (2010) Why do British Indian children have an apparent mental health advantage? *J Child Psychol Psychiatry* 51(10):1171–1183
41. Alyahri A, Goodman R (2008) Harsh corporal punishment of Yemeni children: occurrence, type and associations. *Child Abuse Negl* 32(8):766–773
42. Rutter M, Tizard J, Whitmore K (1970) *Education, health and behaviour*. Longmans, London
43. Achenbach TM (1991) *Manual for the Teacher's Report Form and 1991 profile*. University of Vermont, Burlington
44. Achenbach TM (1991) *Manual for the Youth Self-Report and 1991 profile*. University of Vermont, Burlington
45. Achenbach TM (1991) *Manual for the Child Behavior Checklist 4-18 and 1991 Profile*. University of Vermont, Burlington
46. Goodman R (1994) A modified version of the Rutter parent questionnaire including extra items on children's strengths: a research note. *J Child Psychol Psychiatry* 35(8):1483–1494
47. Noble M et al (2004) *Indices of deprivation*. Report to the Office of the Deputy Prime Minister. Neighbourhood Renewal Unit, London