

The Value of Measuring Impact Alongside Symptoms in Children and Adolescents: A Longitudinal Assessment in a Community Sample

Argyris Stringaris · Robert Goodman

Published online: 16 May 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract The impact that psychiatric symptoms have on the lives of young people is central to clinical practice and classification. However, there is relatively little research on impact and its association with symptoms. This paper examines how well impact can be measured and how it relates to psychiatric outcomes. On four separate occasions over 3 years, symptoms and impact were assessed in a UK epidemiological sample ($n=4,479$; 51.5 % boys) using the Strengths and Difficulties Questionnaire (SDQ) as reported by parents, youths and teachers. Disorders were ascertained using the Development and Well-Being Assessment. An impact scale made of items about distress and impairment demonstrated considerable internal consistency, cross-informant correlations, and longitudinal stability by all reporting sources. Impact at baseline was a strong predictor of psychiatric disorder 3 years later after accounting for psychiatric disorders and symptoms measured at baseline: odds ratio $OR=2.10$, 95 % Confidence Interval (CI) [1.50, 2.94] according to parent-rated impact and $OR=1.71$, CI [1.08, 2.72] according to teacher-rated impact. Changes in impact over time were predicted, but not fully accounted for, by symptoms measured at baseline. Impact can be reliably and easily measured across time, and it may be clinically useful as an independent predictor of future symptoms and psychiatric disorders. More studies are needed to understand inter-individual variation in the impact caused by equivalent symptoms.

Keywords Impact · Impairment · Symptoms · Distress · Psychopathology

A. Stringaris (✉) · R. Goodman
King's College London, Institute of Psychiatry, PO 85, Denmark Hill, London SE5 8AF, UK
e-mail: argyris.stringaris@kcl.ac.uk

R. Goodman
e-mail: robert.goodman@kcl.ac.uk

Introduction

Psychiatric symptoms result in distress and impairment. This *impact* due to psychiatric symptoms is central to clinical decision-making in both the DSM-IV and the upcoming DSM-5. However, there is relatively little research on how well impact can be measured and on how much information it adds (Rapee et al. 2012). In this paper we examine this issue in a large, multi-informant, longitudinal study of a general population sample.

Clinicians inquire about their patients' symptoms and also ask about the impact caused by these symptoms. DSM-IV defines the clinical significance of symptoms according to whether they lead to distress or functional impairment (American Psychiatric Association 2000; Ustun and Kennedy 2009). Distress is defined in terms of the worry and upset caused by the symptoms. Functional impairment is not strictly defined in DSM-IV (Ustun and Kennedy 2009), but is meant to capture a reduced level of adaptive functioning in social and educational (in adults, occupational) domains of life.

When distress or impairment criteria are required to make a diagnosis, the prevalence of psychiatric disorders is often reduced (Bird et al. 1992). For example, separation anxiety and simple phobias are about five times less common if criteria for impairment—and not only symptoms—are required for diagnosis (Simonoff et al. 1997). However, this may not be the case for other disorders, such as depression (Simonoff et al. 1997), where symptoms (Pickles et al. 2001) are the best predictors of later diagnosis.

The decision to require impact as a criterion for psychiatric disorders has attracted criticisms (Rutter 2011). By analogy to other branches of medicine, symptoms of a psychiatric disorder may merit medical attention even in the absence of current impact. Thus, symptoms of diabetes, hypertension or a transient ischemic attack require diagnosis

and intervention even if they have not yet caused impact (Rutter 2011). This and related considerations support the ICD-10 approach that does not generally require impairment before a diagnosis can be made, though impairment can optionally be coded separately.

However, there are important counter-arguments in favor of the DSM's focus on impact. For example, information about impact may help identify children who are suffering (and may be in need of intervention) even if they do not meet the full symptom criteria for an operationalized psychiatric disorder (Angold et al. 1999).

More fundamentally, both DSM-IV's major focus on impact, and ICD-10's optional coding of impact implicitly assume that impact is sufficiently separable from symptoms to be worth considering separately. Thus distress and impairment resulting from symptoms ought to be adequately measurable in their own right and add predictive value to that of symptom assessment alone. This assumption, which has received little empirical scrutiny, is tested in this paper. We examine two sets of questions within a large 3-year longitudinal epidemiologic study using multiple informants (parents, youth and teachers reporting on both symptoms and impact).

The first set of questions relates to whether a short scale can measure impact with satisfactory psychometric properties. Do questions about impact form a scale with acceptable factorial structure and internal consistency? What is the stability of such questions across time? How well do informants agree with each other when they rate impact? Also, what is the relationship between measures of impact and relevant outcomes of psychosocial adjustment, such as contact with psychiatric services, self harm, truancy and contact with police (a measure of concurrent validity)?

The second set of questions concerns the relationship between impact and symptoms across time. Does impact add to the prediction of future psychiatric symptoms even after accounting for psychiatric symptoms at baseline? Also, does knowledge about symptoms at baseline contribute to the prediction of impact 3 years later (after adjusting for baseline impact)? Is it possible to distinguish impact from symptoms across time or can the changes in impact be fully accounted for by changes in symptom levels?

Methods

Sample

The 2004 British Child and Adolescent Mental Health Survey (B-CAMHS04) involved a sample of 5–16 year olds ($n=7,977$) representative of the general British population; it has previously been described in detail (Green et al. 2005).

The study used “child benefit” (a universal state benefit payable in Great Britain for each child in a family) to develop a sampling frame of 5–16 year olds in different postal sectors in England, Wales, and Scotland. After excluding families with no recorded postal code, it was estimated that this represented 90 % of all British children. Out of the 12,294 contacted, there were $n=1,085$ who opted out and $n=713$ who were non eligible or had moved without trace, leaving 10,496 who were approached in person. Of those, $n=7,977$ participated (65 % of those selected; 76 % of those approached). After 12 and again after 24 months (i.e., in 2005 and 2006), parents who had agreed to be followed up again were mailed a Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997). 36 months after the baseline survey i.e., in 2007 (Parry-Langdon 2008), families were approached once more unless they had previously opted out or the child was known to have died. Of the original $n=7,977$ participants, $n=5,326$ (67 %) participated in the detailed follow-up (Parry-Langdon 2008). In this paper we present the data on those children whose families participated at all 4 study time points. This yielded a sample of $n=4,479$ (56 % of original participants; 52 % boys). Attrition analyses show that participants with data at all 4 time points did not differ by gender (52 % vs. 52 % were boys; $\chi^2(1, n=7977)=0.00$; $p=0.95$), were more likely to be younger ($M=10.38$, $SD=3.31$ vs. $M=10.75$, $SD=3.50$ years; $t(7975)=4.84$, $p<0.001$), less likely to suffer from a psychiatric disorder (7 % vs. 13 %; $\chi^2(1, n=7977)=66.03$, $p<0.001$), and less likely to come from a family that owned rather than rented its home; (19 % vs. 42 %; $\chi^2(1, n=7972)=492.57$; $p<0.001$).

Assessment

Strengths and Difficulties Questionnaire (SDQ) The SDQ has robust psychometric properties (Bourdon et al. 2005; Goodman 1997, 2001) and separately inquires about symptoms and impact.

Symptoms The SDQ asks 5 questions each about the following domains: hyperactivity/inattention; behavior problems; emotional symptoms; and peer problems. Summing the items generates scores ranging from 0 to 10 for each scale. A total difficulties score created by the addition of these scales ranges from 0 to 40.

Impact The SDQ impact score is generated for the parent and self report by the sum of 5 items: one item about distress; plus 4 items on social impairment in a) family life, b) friendships, c) learning and d) leisure activities. Teachers are only asked about distress and impairment in learning and friendships (Goodman and Scott 1999).

Parents (95 % were mothers, 4 % fathers, 1 % other sources) were asked to fill in the SDQ at baseline and at each of the study's follow-up points, namely 12, 24, and 36 months. The completion rates of the SDQ were very high in those parents who agreed to participate at each time point: $n=4,474$ of 4,479 (99.9 %) at baseline, 100 % at 12 and 24 months, and $n=4,449$ of 4,479 (99 %) at 36 months.

Children's teachers were asked to fill in the SDQ at baseline and at 36-month follow-up. In Great Britain, it would be unusual for the rating to be completed by the same teacher on both occasions. The completion rates of the teacher SDQ were moderate in those families who had participated at baseline and all follow-ups: $n=3,507$ of 4,479 (78 %) at baseline and $n=2,644$ of 4,479 (59 %) at 36 months. Children aged 11 and above were asked to fill in the SDQ at baseline and at 36 months. At baseline there were $n=2,207$ children who were eligible to provide SDQ data and whose families had participated at baseline and all follow-ups, of whom 1995 (90 %) completed the SDQ. At 36 months there were $n=3,419$ children who were eligible to provide SDQ data and whose families had participated at baseline and follow-up, of whom $n=2,926$ (86 %) completed the SDQ.

Development and Well-Being Assessment (DAWBA) The DAWBA (Ford et al. 2003; Goodman et al. 2000; Messer et al. 2006) is a structured interview administered by lay interviewers who also record verbatim accounts of problems. The questions are closely related to *DSM-IV* and *ICD-10* (APA 2000; World Health, O 1994) diagnostic criteria and focus on current problems. The κ statistic for chance-corrected agreement between two raters was 0.86 for any disorder (SE 0.04), 0.57 for internalizing disorders (SE 0.11), and 0.98 for externalizing disorders (SE 0.02) (Ford et al. 2003). Values of $\kappa < 0$ indicate no agreement, 0–0.20 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1 almost perfect agreement (Landis and Koch 1977). Children were assigned a diagnosis only if their symptoms were causing significant distress or social impairment. The DAWBA interview was administered to all parents and to all children aged 11 or over; a shortened version of the DAWBA was mailed to the child's teacher. Further information on the DAWBA is available from <http://www.dawba.info>. The DAWBA was completed at baseline and at 36 months.

This paper focuses on the overall presence of disorder (i.e., any DSM-IV disorder), externalizing disorders (the combination of conduct, oppositional defiant and attention deficit/hyperactivity disorders), and internalizing disorders (the combination of depressive and anxiety disorders).

Other Outcomes To assess the concurrent validity of the impact ratings, the following information available to the baseline (2004) part of the survey were used: contact with psychiatric

services ($n=113$, 3 %), self harm ($n=151$, 3 %), truancy ($n=98$, 2 %) and contact with police ($n=122$, 3 %). A participant was coded as having experienced one of these outcomes if so rated by youth-, teacher- or parent report. For self harm, information was only available from the youth and parents. To assess predictive validity, we also examined whether impact ratings in 2004 were associated with new onset (i.e. reported for the first time) of the following in the 36-month follow up of the 2007 survey: contact with psychiatric services ($n=127$, 3 %), self harm ($n=199$, 4 %), truancy ($n=47$, 1 %) and contact with police ($n=201$, 5 %).

Analysis

Internal Consistency This was estimated using Cronbach's alpha across each informant across time.

Factorial Structure The single-factor structure of the impact score was tested in a confirmatory factor analysis using the 5 SDQ impact items for the parent and self report (but not the 3-item teacher report). Fit was assessed by the following indices: Comparative Fit index (CFI; 0.95 and above indicates good fit) the Tucker Lewis Index (TLI; values close to 1 indicate good fit) the Root Mean Square Error of Approximation (RMSEA; values smaller than 0.06 indicate good fit) and the Standardized Root Mean Square Residual (SRMR; values smaller than 0.07 indicate good fit) (Yu 2002).

Concurrent Validity This is presented in a figure as the standardized mean of impact by one of the following binary categories from the baseline 2004 survey: contact with psychiatric services, self harm, truancy and contact with police. A table presents the results from logistic regression models in which each of these categories (from the baseline 2004 survey) was the outcome and baseline impact the independent variable. The models are presented unadjusted or adjusted for baseline total SDQ score.

Predictive Validity A table presents the results from logistic regression models in which the independent variable is impact at baseline, and the dependent variables were new onset of each of the following binary categories in the 36-month follow up of the 2007 survey: contact with psychiatric services, self harm, truancy and contact with police. The models are presented unadjusted or adjusted for baseline total SDQ score.

Longitudinal Stability This was tested in regression models within informants (e.g., parent report at baseline predicting parent report at 36 months) and across informants (e.g., parent report at baseline predicting teacher report at 36 months).

Impact Ratings Across Informants Concurrent association of ratings between informants was assessed using correlation models, whereas the longitudinal associations were assessed using regression models.

Association Between Symptoms and Other Outcomes

Association Between Symptoms and Impact The prediction of symptoms by impact was estimated in regression models with the standardized total SDQ symptom score as the outcome (e.g., at 36 months) and standardized baseline impact score as the predictor. Adjusted models covaried for standardized baseline total SDQ symptom score. This was reversed for the prediction of impact score from the total SDQ symptom score.

Association Between Psychiatric Disorders and Impact The prediction of psychiatric diagnoses at 36 month follow-up (dependent variable) by baseline impact was estimated in logistic regression models where the standardized impact score was used as an independent variable. In adjusted models, diagnoses at baseline were used as covariates.

Trajectories of Symptoms and Impact Across Time A path analysis model was estimated to test whether total SDQ symptom score and impact followed distinguishable trajectories across time. The model is schematically represented in Fig. 2.

Analyses were conducted in Mplus, Version 5 (Muthen and Muthen 2007). The study option was used to standardize scores and all analyses were estimated with robust maximum likelihood that uses sandwich estimators for the standard errors according to Huber and White (Huber 1967; White 1980).

Ethical Approval

All study procedures received multicentre research ethics committee approval and informed consent was obtained from parents and assent from children participants.

Results

Internal Consistency The internal consistencies for the impact scales were high: by parent report, 0.88 at baseline, 0.89, 0.82, and 0.88 at 12, 24, and 36 months respectively. For youth-report, they were 0.82 at baseline and 0.88 at outcome; for teacher report 0.86 at baseline and 0.87 at outcome.

Factorial Structure As shown in Table 1, a single-factor structure was acceptable at most time points for the parent- and youth-reported impact scales. In particular, the CFI and the SRMR showed that the one-factor solution was a good fit to the data. The RMSEA indicated acceptable fit at all time points, with the exception of 12 months, while the TLI was below the 0.95 cut off at 12 and 36 months by parent report. All factor loadings were excellent (Table 1).

Inter-Rater Correlation and Longitudinal Stability There was a moderately strong correlation at baseline between impact reported by all three informant sources (Table 2). The 12-, 24-, and 36-month stability of the impact scores within informants was considerable (Table 2). Parent- and teacher-reported impact at baseline were highly significant in predicting impact 3 years later, within as well as across informants.

Concurrent Validity As shown in Fig. 1, impact, as rated by any informant, was higher in those who had experienced contact with psychiatric services, self harm, truancy and contact with police compared with those who had not experienced such outcomes. Table 3 presents the odds ratios of the association between each of these outcomes and impact. As can be seen, impact rated by any of the three informants was a significant predictor of psychiatric service use, even after adjustment for baseline total SDQ score. However, neither teacher- nor self-report were significant predictors of self harm after adjustment for baseline symptoms. Similarly, neither parent- nor teacher-report was significantly associated with police contact once adjusted for baseline symptoms.

Predictive Validity Table 4 presents the odds ratios of the association between each of the validating outcomes in year 2007 and baseline impact. As can be seen, parent- and teacher- rated impact was a significant predictor of future psychiatric first-time service use and new-onset self harm in 2007, even after adjustment for baseline total SDQ score. However, neither parent-rated nor teacher-rated impact were predictive of new-onset truancy or first-time police contact in 2007 after adjustment for baseline SDQ scores. Self-reported impact was not a significant predictor of any of these 2007 outcomes once baseline SDQ score had been adjusted for.

Impact Predicting Future Symptoms Impact measured at baseline was significantly predictive of the SDQ total symptom score at 36-month follow-up in unadjusted models (Table 5). This was true both within as well as across informants. In models adjusted for SDQ total symptom score at baseline, parent- or teacher-rated impact measured at baseline remained a significant predictor of SDQ total symptom

Table 1 Confirmatory factor analysis of impact items by informant source at various time points

Reporting source	Parent				Youth	
	Baseline	12 months	24 months	36 months	Baseline	36 months
	Distress	0.80	0.85	0.84	0.81	0.82
Impairment at home	0.82	0.82	0.82	0.84	0.79	0.84
Impairment with friends	0.83	0.83	0.82	0.78	0.67	0.64
Impairment in class	0.71	0.75	0.76	0.72	0.66	0.67
Impairment in leisure activities	0.75	0.74	0.76	0.79	0.58	0.56
CFI	0.98	0.97	0.98	0.97	0.99	0.98
TLI	0.96	0.93	0.95	0.94	0.98	0.97
RMSEA	0.05	0.07	0.06	0.06	0.03	0.04
SRMR	0.02	0.03	0.02	0.03	0.02	0.02

Factor loadings and fit indices from the confirmatory factor analysis (CFA) of impact items by parent and self report. Each column represents a separate CFA (each CFA had five degrees of freedom)

score measured by parent, teacher, or youth report at 36-month follow-up. However, youth-reported impact at baseline was not predictive of youth-reported symptoms at 36-month follow-up. The effect sizes of the associations between impact at baseline and symptoms at outcome varied from moderate in unadjusted models to small in adjusted models (Table 5).

Impact Predicting Future Psychiatric Disorders We assessed how impact predicted psychiatric disorders at 36-month outcome. Parent-rated impact was a significant predictor of psychiatric disorders, although the strength of its association diminished progressively with adjustment for baseline disorder (Table 6). The results were similar for teacher-reported impact (Tables 6), while youth-reported impact was a significant predictor of each disorder domain except for the prediction to disruptive behavior disorders which was non-significant after adjustment.

Symptoms Predicting Future Impact SDQ total symptom score at baseline was highly predictive of impact score at 36-month follow-up in unadjusted models (Table 7). This was true both within and across informants. In models adjusted for impact measured at baseline, total SDQ symptom score measured by any informant at baseline remained a significant predictor of impact score measured at 36-month follow-up. However, the effect sizes of the associations were small, as indicated by the standardized coefficients (Table 7).

Predictions Between Impact and Symptoms Across Time Focusing just on data from parent reports—since these were available at baseline, 12, 24, and 36 months—we studied the association between impact and symptoms across time in a path analytic model (Fig. 2). The within-domain

stability is stronger than the across-domain stability: impact is always a better predictor of impact (the coefficients are significant even for the 36 month predictions), whereas the SDQ total symptom score is always a better predictor of SDQ symptom score. This makes it unlikely that one is merely secondary to the other. Figure 2 also demonstrates that impact is directly predictive of future symptoms only in the short term: for example, impact at 12 months is predictive of SDQ symptom score at 24 months (adjusting for the stability of total SDQ symptom score and the prediction from 24-month impact), but it is not predictive of SDQ symptom score at 36 months with comparable adjustments. Similarly, the SDQ symptom score predicts impact in the short term (12 months). However, Fig. 2 also shows that impact is predictive of SDQ total symptom score in the longer run but indirectly. For example, 12-month impact score predicts 36 month SDQ symptom score indirectly via the 12- and 24-month SDQ total symptom scores (which strongly predict the 36-months SDQ symptom score), as well as via its significant associations with impact scores at 12 and 24 months (which predict SDQ symptom score at 36 months). The same pattern of relationships applies in the prediction from SDQ total symptom to impact.

Discussion

This paper examined the value of measuring impact alongside psychiatric symptoms. Impact was defined as the presence of distress or impairment in different settings and analyses were conducted in a large epidemiologic sample using a longitudinal and multi-informant (parent, youth, and teacher) design. Overall, we found that even a brief measure

Table 2 Association of impact within and across informants at baseline and follow-up

		Baseline impact score			12 month impact score	24 month impact score	36 month impact score	
		Parent	Teacher	Youth	Parent	Parent	Parent	Teacher
		<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>
		95 % CI	95 % CI	95 % CI	95 % CI	95 % CI	95 % CI	95 % CI
		<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>	<i>n</i>
Baseline impact score	Parent							
	Teacher	0.47						
		0.41, 0.53						
		3492						
	Youth	0.38	0.34					
		0.30, 0.47	0.23, 0.44					
		1986	1502					
12 month impact score	Parent	0.50	0.38	0.39				
		0.45, 0.55	0.32, 0.45	0.30, 0.49				
		4478	3493	1986				
24 month impact score	Parent	0.31	0.28	0.26	0.52			
		0.28, 0.35	0.22, 0.35	0.17, 0.36	0.47, 0.58			
		4478	3493	1986	4479			
36 month impact score	Parent	0.38	0.39	0.30	0.48	0.47		
		0.34, 0.41	0.32, 0.45	0.20, 0.41	0.44, 0.52	0.44, 0.51		
		4450	3469	1502	4451	4451		
	Teacher	0.19	0.43	0.23	0.20	0.22	0.36	
		0.14, 0.24	0.35, 0.51	0.06, 0.39	0.15, 0.25	0.17, 0.27	0.29, 0.42	
		2670	2130	936	2671	2671	2666	
	Youth	0.14	0.14	0.22	0.18	0.18	0.38	0.35
		0.09, 0.19	0.07, 0.21	0.13, 0.31	0.12, 0.24	0.12, 0.23	0.30, 0.45	0.33, 0.37
		2915	2245	1712	2915	2915	2887	1673

p < 0.001 in all cells; *r* robust regression coefficients are presented, *CI* confidence interval, *n* number of observations
 All findings in bold are significant (*p* < 0.05)

of impact had adequate psychometric properties and that it added predictive value.

The first set of questions related to whether impact could be measured appropriately using a short scale. The impact scale had excellent internal consistency by parent-, youth-, and teacher report. In addition, there was support for a single-factor structure for the parent- and youth-reported scales, although one of the indicators (TLI) was just below the generally acceptable threshold on two measurement occasions by parent report. We also found that the impact scales showed substantial stability across time both within as well as across informants. The stability is particularly notable in teacher report, where children were rated by different teachers at baseline and follow up. Also, the agreement (ranging from 0.34 to 0.47) between informants compares favorably with cross-informant correlations for

symptoms in previous meta-analyses (Achenbach et al. 1987). Finally, impact showed a strong concurrent association with relevant outcomes of psychosocial adjustment, namely contact with psychiatric services, self harm, truancy and contact with police, although some of these associations were diminished to non-significance after adjustment for psychiatric symptoms. In addition, parent- and teacher-rated impact showed good predictive validity for new-onset of contact with psychiatric services and new-onset self harm; however, self-rated impact was not predictive of any of these outcomes and parent- and teacher-rated impact was not predictive of either truancy or police contact after adjustment for baseline symptoms. It is notable that impact predicted close to an 80 % increase in the probability of contact with psychiatric services concurrently and 20 % for new onset of service contact in 2007, even after psychiatric

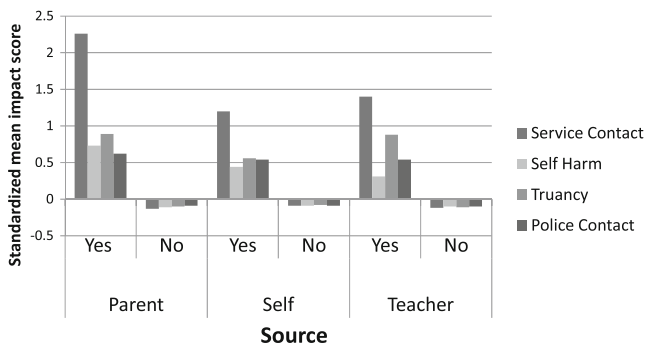


Fig. 1 The association between mean impact and adverse outcomes by reporting source. Standardized (Cohen’s d) mean scores for impact are presented on the y axis

symptoms were taken into account. This highlights the importance of impact (particularly as reported by the parents of individuals) in decisions about use of psychiatric services. This raises the possibility that changes in impact may be a good indicator of the perceived effectiveness of interventions offered by psychiatric services. If so, this would have the major advantage of universal applicability since the same measure of impact is potentially relevant to all disorders (or combination of disorders), whereas a separate set of symptoms is potentially needed for every single disorder or combination of disorders. On the other hand, our

findings also suggest that some negative outcomes are better predicted by psychiatric symptoms than by impact, at least by some reporting sources. Given this uncertainty about the relative benefits of monitoring impact, symptoms or both, future studies could profitably compare the utility of impact- and symptom-based measures when used in session-by-session monitoring or longer-term follow-up.

The second set of questions concerned the relationship between symptoms and impact across time. The number of significant findings (at the pre-set level of $p < 0.05$) was far greater than would be expected by chance, as shown in the paper’s tables. In dimensional analyses, impact at baseline was a significant predictor of future symptoms both within and across informants. The predictions were still significant even after accounting for baseline symptoms, with the exception of youth-reported impact adjusted for youth-reported symptoms. The unadjusted effect sizes of these associations were moderate (e.g., 0.34 for parent reported impact to symptoms 3 years later), but were in many of the models substantially attenuated after adjusting for baseline symptoms (e.g., 0.04 for parent reported impact to symptoms 3 years later). We also show that standardized impact scores at baseline added considerably to the prediction of psychiatric disorders, even years later. We show that every increase in parent-rated impact score increased the probability of future psychiatric disorder over 3 years

Table 3 Concurrent validity: association between impact and psychosocial adjustment

Outcome	Predictors	Reporting source		
		Parent OR 95 % CI	Teacher OR 95 % CI	Youth OR 95 % CI
Outcome: service contact	Impact only	2.63 2.35, 2.93	2.14 1.89, 2.42	1.73 1.50, 1.99
	Impact adjusted for baseline total SDQ score	1.76 1.12, 1.52	1.74 1.39, 2.18	1.37 1.25, 1.60
Outcome: self harm	Impact only	1.65 1.49, 1.82	1.39 1.21, 1.60	1.41 1.25, 1.60
	Impact adjusted for baseline total SDQ score	1.31 1.12, 1.52	1.06 ^{ns} 0.85, 1.32	1.14 ^{ns} 0.99, 1.31
Outcome: truancy	Impact only	1.70 1.52, 1.90	1.76 1.54, 2.02	1.45 1.27, 1.66
	Impact adjusted for baseline total SDQ score	1.35 1.13, 1.61	1.20 ^{ns} 0.95, 1.53	1.16 ^{ns} 0.99, 1.36
Outcome: police contact	Impact only	1.56 1.40, 1.74	1.55 1.33, 1.79	1.46 1.28, 1.65
	Impact adjusted for baseline total SDQ score	1.10 ^{ns} 0.93, 1.31	0.97 ^{ns} 0.76, 1.24	1.23 1.07, 1.43

OR odds ratio; CI 95 % confidence interval. Odds ratios with confidence intervals are presented from logistic regression models with each of the outcomes as dependent variables and impact as an independent variable either unadjusted or adjusted for baseline total SDQ score

All findings in bold are significant ($p < 0.05$)

Table 4 Predictive validity: association between impact and psychosocial adjustment

	Predictors	Reporting source		
		Parent OR 95 % CI	Teacher OR 95 % CI	Youth OR 95 % CI
Outcome: service contact	Impact only	1.56 1.40, 1.74	1.69 1.49, 1.91	1.10 ^{ns} 0.87, 1.41
	Impact adjusted for baseline total SDQ score	1.19 1.01, 1.41	1.36 1.09, 1.70	0.92 ^{ns} 0.70, 1.23
Outcome: self harm	Impact only	1.39 1.26, 1.55	1.46 1.30, 1.64	1.24 1.08, 1.43
	Impact adjusted for baseline total SDQ score	1.24 1.07, 1.44	1.37 1.12, 1.67	1.08 ^{ns} 0.92, 1.28
Outcome: truancy	Impact only	1.42 1.19, 1.71	1.50 1.20, 1.87	1.20 ^{ns} 0.94, 1.53
	Impact adjusted for baseline total SDQ score	0.93 ^{ns} 0.70, 1.24	0.88 ^{ns} 0.61, 1.27	0.95 ^{ns} 0.70, 1.29
Outcome: police contact	Impact only	1.28 1.15, 1.44	1.44 1.27, 1.63	1.46 1.28, 1.65
	Impact adjusted for baseline total SDQ score	1.01 ^{ns} 0.86, 1.19	1.16 ^{ns} 0.94, 1.42	1.12 ^{ns} 0.96, 1.31

OR odds ratio; CI 95 % confidence interval. Odds ratios with confidence intervals are presented from logistic regression models with each of the outcomes as dependent variables and impact as an independent variable either unadjusted or adjusted for baseline total SDQ score

All findings in bold are significant ($p < 0.05$)

between 67 % (parent report) and 22 % (self report). As with the previous analyses, youth-reported impact was the weakest predictor of future outcomes, similar to other findings about the relative weakness of youth-reported psychopathology (Goodman and Goodman 2009). Our results agree in part with those of Pickles et al. (2001). Their results suggest that baseline impact adds to the prediction of future CD/ODD diagnosis and future impact, independently of baseline symptoms. However, these authors found that when predicting future depression symptoms and impairment, baseline symptoms are a better predictor than baseline impact. We found that impact independently predicts emotional disorders (which includes depressive disorder) to about the same degree as it predicts externalizing disorders. Pickles et al. (2001) measured domain-specific impairment, rather than global impact as we did—a difference that conceivably accounts for the discrepant findings. The results of this paper also suggest that impact was sensitive to changes in symptoms across time. Thus, an increase in psychiatric symptoms led to more subsequent impact (adjusted for impact levels present at baseline), even 3 years later. Using a path analytic model, we showed that impact was best predicted by impact and symptoms were best predicted by symptoms.

The Strengths and Difficulties Questionnaire (SDQ) has been shown to be a dimensional measure of psychopathology (Goodman and Goodman 2009), useful for measuring

mental health problems in young people around the world (Achenbach 2012). However, little previous research has been done to validate its impact supplement. A previous study has shown that including the score of the SDQ impact supplement as part of an added value score is helpful in assessing the effectiveness of clinical interventions (Ford et al. 2009). To our knowledge this is the first study that uses longitudinal epidemiological data to examine whether the impact supplement adds value to the prediction of psychiatric disorder.

Our findings suggest two reasons why impact should be measured in addition to symptoms. The first reason relates to measurement utility: impact added to the prediction of new-onset future symptoms, a finding consistent with previous results (Costello et al. 1999). It could be argued that this prediction is merely because symptoms at baseline have been measured imperfectly. However, any measurement contains noise. This may be especially true of clinical assessment. Finding that the addition of a concise measure of impact improved the prediction of future symptoms is therefore important.

The second reason for measuring impact in addition to symptoms concerns decisions about service provision. Impact may guide whether a child warrants referral to specialist services and treatment. Particularly when it comes to

Table 5 Association between impact score at baseline and SDQ total score at follow-up for each informant source (only baseline and 36 month follow-up shown)

		Impact score at baseline			
		Reporting source at baseline			
		Predictor variables	Parent	Teacher	Youth
			β	β	β
			95 % CI	95 % CI	95 % CI
			R^2	R^2	R^2
Outcome: SDQ score at 36 months follow-up	Parent	Impact score only	0.34	0.41	0.32
			0.31, 0.36	0.37, 0.46	0.25, 0.40
			$R^2=0.21$	$R^2=0.13$	$R^2=0.08$
		Impact score adjusted for baseline SDQ score	0.04	0.11	0.13
			0.01, 0.07	0.07, 0.15	0.07, 0.19
			$R^2=0.48$	$R^2=0.48$	$R^2=0.22$
	Teacher	Impact score only	0.22	0.45	0.31
			0.19, 0.26	0.40, 0.51	0.19, 0.44
			$R^2=0.09$	$R^2=0.16$	$R^2=0.07$
		Impact score adjusted for baseline SDQ score	0.09	0.14	0.17
			0.06, 0.13	0.06, 0.23	0.05, 0.29
			$R^2=0.25$	$R^2=0.23$	$R^2=0.18$
Youth	Impact score only	0.19	0.24	0.29	
		0.16, 0.21	0.18, 0.29	0.22, 0.35	
		$R^2=0.07$	$R^2=0.04$	$R^2=0.06$	
	Impact score adjusted for baseline SDQ score	0.04	0.16	0.023 ^{ns}	
		0.02, 0.07	0.11, 0.22	0.04, 0.08	
		$R^2=0.31$	$R^2=0.08$	$R^2=0.30$	

β standardized regression coefficient; CI 95 % confidence interval; R^2 = proportion of variance. All findings in bold are significant ($p < 0.05$); otherwise non-significant (ns); standardized robust regression coefficients and confidence intervals from robust maximum likelihood models are presented in each cell with R^2 as the estimate of the variance predicted

Table 6 Standardized impact at baseline as a predictor of psychiatric disorders in adjusted and unadjusted models

		Reporting source		
Predictors		Parent	Teacher	Youth
		OR	OR	OR
		95 % CI	95 % CI	95 % CI
Outcome: any disorder	Impact only	2.36	2.01	1.55
		2.16, 2.58	1.83, 2.20	1.39, 1.74
	Impact adjusted for any disorder at baseline	1.67	1.36	1.22
		1.51, 1.86	1.20, 1.52	1.07, 1.40
Outcome: emotional disorders	Impact only	1.68	1.52	1.46
		1.53, 1.85	1.35, 1.72	1.28, 1.65
	Impact adjusted for baseline emotional disorders	1.44	1.24	1.24
		1.29, 1.61	1.08, 1.43	1.07, 1.45
Outcome: externalizing disorders	Impact only	2.25	2.03	1.46
		2.05, 2.46	1.83, 2.26	1.27, 1.68
	Impact adjusted for baseline externalizing disorders	1.55	1.28	1.15 ^{ns}
		1.37, 1.75	1.10, 1.50	0.95, 1.40

OR = odds ratio; CI = 95 % confidence interval. All findings in bold are significant ($p < 0.05$); otherwise non-significant (ns); odds ratios from logistic regression models are presented

Table 7 Association between SDQ total symptom score at baseline and impact at follow-up for each informant source (only baseline and 36 month follow-up shown)

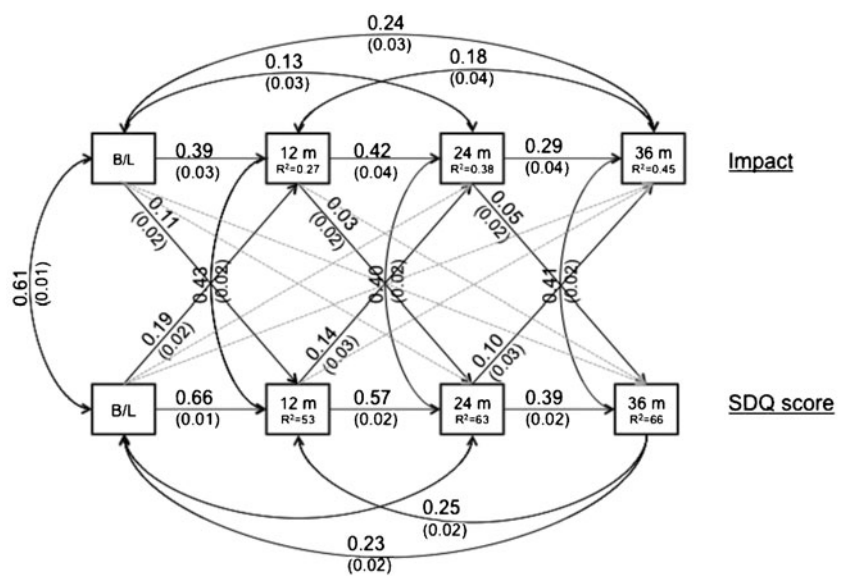
		SDQ symptom score at baseline				
		Reporting source at baseline				
Outcome: impact at 36 months follow-up	Reporting source at outcome	Predictor variables	Parent	Teacher	Youth	
			β 95 % CI R^2	β 95 % CI R^2	β 95 % CI R^2	
	Parent	SDQ score only	0.08	0.05	0.05	
			0.07, 0.08 $R^2=0.18$	0.04, 0.06 $R^2=0.09$	0.04, 0.06 $R^2=0.07$	
	Teacher	SDQ score only	0.06	0.07	0.04	
			0.05, 0.06 $R^2=0.09$	0.06, 0.08 $R^2=0.14$	0.03, 0.06 $R^2=0.05$	
	Youth	SDQ score only	0.04	0.03	0.04	
			0.03, 0.04 $R^2=0.04$	0.02, 0.03 $R^2=0.02$	0.03, 0.06 $R^2=0.05$	
		SDQ score adjusted for baseline impact	0.03	0.02	0.02	
			0.02, 0.04 $R^2=0.28$	0.01, 0.03 $R^2=0.28$	0.01, 0.03 $R^2=0.27$	
			SDQ score adjusted for baseline impact	0.03	0.04	0.03
			0.02, 0.04 $R^2=0.17$	0.03, 0.06 $R^2=0.17$	0.02, 0.04 $R^2=0.16$	
			SDQ score adjusted for baseline impact	0.03	0.02	0.04
			0.02, 0.03 $R^2=0.06$	0.01, 0.03 $R^2=0.05$	0.03, 0.05 $R^2=0.06$	

β standardized regression coefficient; CI 95 % confidence interval; R^2 = proportion of variance All findings in bold are significant ($p < 0.05$); otherwise non-significant (ns); standardized regression coefficients and confidence intervals from robust maximum likelihood models are presented in each cell with R^2 as the estimate of the variance predicted

anxiety disorders, most of the children who meet symptom criteria for a disorder experience little resultant impact (Simonoff et al. 1997). It has not yet been shown that

children who display psychiatric symptoms but are not impaired would come to significant future harm unless they were treated. Therefore, many clinicians might continue to

Fig. 2 Path analysis of the relationship between impact and SDQ total symptom score across time (all parent reported). Significant ($p < 0.05$) paths or correlations with standard errors in brackets are presented as solid straight lines or solid curved arrows respectively, dashed lines illustrate non significant associations. B/L = baseline, R^2 = proportion of variance explained



use a low measured impact to justify refraining from diagnosing or treating non-impaired children.

Our study has a number of strengths including the use of a large epidemiologic study, a longitudinal design, assessment with multiple informants, and the use of dimensional as well as categorical measures of psychopathology. However, the results should also be interpreted in the light of important limitations. Firstly, as is often the case in longitudinal studies, there was considerable attrition. While this would be most likely to bias prevalence estimates rather than the associations between disorders and other factors (Wolke et al. 2009), it has meant that those with high levels of psychopathology at baseline were less likely to participate in the study. Secondly, our measure of impact was designed to be concise and simple to use. Unlike other instruments, such as the Eyberg child behaviour inventory (Eyberg 1992), the SDQ does not ask about the impact associated with each reported symptom. Knowing which of the reported symptoms are contributing most to perceived impairment can be clinically important. Further studies, possibly also using qualitative approaches, may be required to assess the full range of experiencing the impact of psychiatric symptoms. Thirdly, our measures of concurrent validity are gathered from the same informants (parents, teachers, and youth) as our measure of impact. Ideally, these outcomes should also be gathered from external informants (such as police records). Including outcomes that are reported by sources external to the predictors avoids the problems associated with shared method variance—the inflation of estimates of association due to the use of the same informants (Campbell and Fiske 1959).

Finally, there is a certain circularity when predicting DSM psychiatric disorders as an outcome, since they contain impairment as a pre-requisite. However, we obtained a similar pattern of results when predicting to SDQ scores.

In conclusion, our findings suggest that impact can be reliably measured with a brief scale and that doing so may benefit clinical assessment. Having reliable measures of symptoms and impact may contribute in the future to investigations of why individuals with similar sets of symptoms can experience different levels of impact.

Acknowledgments The authors are grateful to Miss Acklom for her expert technical assistance; Dr Stringaris gratefully acknowledges the support by the Wellcome Trust.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Achenbach, T. M. (2012). International epidemiology of child and adolescent psychopathology I: diagnoses, dimensions and conceptual issues. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51, 1261–1272. doi:10.1016/j.jaac.2012.09.010.
- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213–232. doi:10.1037/0033-2909.101.2.213.
- Angold, A., Costello, E. J., Farmer, E. M., Burns, B. J., & Erkanli, A. (1999). Impaired but undiagnosed. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 129–137. doi:10.1097/00004583-199902000-00011.
- APA. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington: American Psychiatric Press.
- Bird, H. R., Gould, M. S., & Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research. *Journal of the American Academy of Child and Adolescent Psychiatry*, 31, 78. doi:10.1097/00004583-199201000-00012.
- Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The strengths and difficulties questionnaire: U.S. normative data and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 557. doi:10.1097/01.chi.0000159157.57075.c8.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105. doi:10.1037/h0046016.
- Costello, E. J., Angold, A., & Keeler, G. P. (1999). Adolescent outcomes of childhood disorders: the consequences of severity and impairment. *Journal of the American Academy of Child and Adolescent Psychiatry*, 38, 121–128. doi:10.1097/00004583-199902000-00010.
- Eyberg, S. M. (1992). Parent and teacher behavior inventories for the assessment of conduct problem behaviors in children. In L. VandeCreek, S. Knapp, & T. L. Jackson (Eds.), *Innovations in clinical practice: A source book* (pp. 261–270). Sarasota: Professional Resource Press.
- Ford, T., Goodman, R., & Meltzer, H. (2003). The British child and adolescent mental health survey 1999: the prevalence of DSM-IV disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42, 1203. doi:10.1097/00004583-200310000-00011.
- Ford, T., Hutchings, J., Bywater, T., Goodman, A., & Goodman, R. (2009). Strength and difficulties questionnaire added value scores: evaluating effectiveness in child mental health interventions. *The British Journal of Psychiatry*, 194, 552–558. doi:10.1192/bjp.bp.108.05.2373.
- Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337. doi:10.1097/00004583-200111000-00015.
- Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 400–403. doi:10.1097/CHI.06013e3181985068.
- 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. doi:10.1023/A:1022658222914.
- Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent

- psychopathology. *Journal of Child Psychology and Psychiatry*, 41, 645–655. doi:10.1111/j.1469-7610.2000.tb02345.x.
- Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental health of children and young people in Great Britain*. London: The Stationery Office.
- Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 221–233). California: University of California Press.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. Retrieved from <http://www.jstor.org/stable/2529310>.
- Messer, J., Goodman, R., Rowe, R., Meltzer, H., & Maughan, B. (2006). Preadolescent conduct problems in girls and boys. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 184. doi:10.1097/01.chi.0000186403.13088.d8.
- Muthen, L. K., & Muthen, B. O. (2007). Mplus software version 5.2.1.
- Parry-Langdon, N. (2008). *Three years on: Survey of the emotional development and well-being of children and young people*. Newport: Office for National Statistics.
- Pickles, A., Rowe, R., Simonoff, E., Foley, D., Rutter, M., & Silberg, J. (2001). Child psychiatric symptoms and psychosocial impairment: relationship and prognostic significance. *The British Journal of Psychiatry*, 179, 230–235. doi:10.1192/bjp.179.3.230.
- Rapee, R. M., Bogels, S. M., van der Sluis, C. M., Craske, M. G., Ollendick, T. (2012). Annual research review: conceptualising functional impairment in children and adolescents. *Journal of Child Psychology and Psychiatry*, in press.
- Rutter, M. (2011). Research review: child psychiatric diagnosis and classification: concepts, findings, challenges and potential. *Journal of Child Psychology and Psychiatry*, 52, 647–660. doi:10.1111/j.1469-7610.2011.02367.x.
- Simonoff, E., Pickles, A., Meyer, J. M., Silberg, J. L., Maes, H. H., Loeber, R., et al. (1997). The Virginia Twin Study of adolescent behavioral development. Influences of age, sex, and impairment on rates of disorder. *Archives of General Psychiatry*, 54, 801. doi:10.1001/archpsyc.1997.01830210039004.
- Ustun, B., & Kennedy, C. (2009). What is “functional impairment”? disentangling disability from clinical significance. *World Psychiatry*, 8, 82–85.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–838. Retrieved from <http://www.jstor.org/stable/1912934>.
- Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., et al. (2009). Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *The British Journal of Psychiatry*, 195, 249–256. doi:10.1192/bjp.bp.108.053751.
- World Health, O. (1994). *ICD-10: Classification of mental and behavioural disorders*. Geneva: World Health Organization.
- Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles: University of California.