ORIGINAL PAPER

New Interview and Observation Measures of the Broader Autism Phenotype: Group Differentiation

Maretha de Jonge · Jeremy Parr · Michael Rutter ·
Simon Wallace · Chantal Kemner · Anthony Bailey ·
Herman van Engeland · Andrew Pickles

Abstract To identify the broader autism phenotype (BAP), the Family History Interview subject and informant versions and an observational tool (Impression of Interviewee), were developed. This study investigated whether the instruments differentiated between parents of children with autism, and parents of children with Down syndrome (DS). The BAP scores of parents of 28 multiplex autism families were compared with parents from, 32 DS families. The BAP measures provided good group differentiation but when considered together, the subject interview did not improve group differentiation. The differentiation was better for fathers than mothers. The measures do carry an important degree of validity; whether they can differentiate the BAP from other social disorders should be tested.

Keywords Broader autism phenotype · Group differentiation · Interview measures · Observational measures · Diagnosis

M. de Jonge (✉) · C. Kemner · H. van Engeland
Department of Psychiatry, University Medical Center Utrecht,
Room A01.468, Postbox 85500, 3508 GA Utrecht,
The Netherlands
e-mail: m.v.jonge@umcutrecht.nl

M. de Jonge · C. Kemner · H. van Engeland
Brain Center Rudolf Magnus, University Medical Center
Utrecht, Utrecht, The Netherlands

Introduction

It has long been apparent that some relatives of individuals with an autism spectrum disorder (ASD) show features that are qualitatively similar to autism, but which differ in being milder and which (unlike ASD) are not associated with either epilepsy or intellectual disability (Bailey et al. 1998; Dawson et al. 2007; Losh et al. 2011; Piven et al. 1997; Szatmari et al. 2000). These characteristics in relatives have become known as the broader autism phenotype (BAP). These features have been reported in parents of children with autism associated with severe intellectual disability (Starr et al. 2001) and in family members of individuals with Asperger's syndrome (Ghaziuddin 2005; Klin et al. 2005) indicating that the family aggregation of

J. Parr · S. Wallace · A. Bailey
Department of Psychiatry, University of Oxford, Oxford, UK

J. Parr
Institute of Neuroscience, Newcastle University,
Newcastle upon Tyne, UK

the BAP extends right across the autism spectrum. Possibly, however, it may be more common in families including two or more individuals with an ASD—multiplex families (Losh et al. 2008; Szatmari et al. 2000).

Several authors suggested that the best way to identify the BAP is to rely on a combination of self-reported information, informant reported information and observational report by a clinical examiner (Dawson et al. 2007; Pickles et al. 2000; Piven et al. 1994). Therefore members of the International Molecular Study of Autism Consortium (IMGSAC) set about developing a Family History Interview in 2000, consisting of a self-report version (FHI-S) and informant report version (FHI-I) and an Impression of Interviewee rating scale (IoI) that could be used in combination to derive a dimensional measure of the BAP. The first paper in this series described the IoI schedule and its psychometric properties (Pickles et al. 2013); the second described the overall strategy and gave the reliability findings for both interviews (Parr et al., resubmitted). For subject (FHI-S) and informant interviews (FHI-I) and for the IoI these papers identified a subset of items that exhibited relatively better test–retest reliability and convergent validity. This paper assesses the construct validity of the instruments by examining the group differentiation findings for all three measures.

An essential feature of any assessment of the BAP, whether based on one or multiple measures, is that it provides good differentiation between members of families including someone with ASD and members of families without an ASD proband. That feature, in itself, is not an adequate test of validity (because the BAP measures must also be shown to differentiate BAP features from other phenomena with which they might be confused—such as schizotypy in families with a schizophrenic proband). Nevertheless, the group differentiation reported here provides a necessary first step.

The authors were part of IMGSAC and hence were familiar with the development of the measures. The sample used, however, was partially separate from the broader IMGSAC group, but was comparable in the fact that all the ASD families were multiplex (i.e. contained two or more individuals with an ASD). This characteristic was intended to reduce genetic heterogeneity and also the presence of de novo genetic mutations that might not have indexed familial risk for ASD. Comparison was made with families

M. Rutter
MRC Centre for Social, Genetic and Developmental Psychiatry, King's College London, London, UK

A. Bailey
Department of Psychiatry, University of British Columbia, Vancouver, Canada

A. Pickles
Department of Biostatistics, Institute of Psychiatry, King's College London, London, UK

with a Down syndrome (DS) child, in an attempt to control for the social burden of raising a child with a serious developmental disability. However, in order to check on this possibility we used an Impact on the Family Scale.

Methods

Sample

Families with at least two children with a clinical diagnosis of autism/ASD were recruited as part of an ongoing collaborative genetic study (by the IMGSAC). The families were recruited among the patients of the Department of Child and Adolescent Psychiatry of the University Medical Centre in Utrecht and among members of the Dutch ASD parent association. Families with children with known medical causes of autism were excluded by history, karyotype and fragile-X DNA testing. A comparison group of parents with a child with DS was recruited from the community through the Dutch DS Parent Association ('comparison families'). The study was approved by the ethics committee of the University Medical Centre Utrecht and all participants gave written informed consent.

The parents of 28 ASD multi-incidence families participated in this study. Two fathers did not want to participate, and one mother had died, leaving 53 participating parents (26 fathers, 27 mothers). Participants in the control group were included when there was no indication of an ASD in their first-degree relatives following telephone discussions with a parent. Parents of children with DS were asked whether their child had an ASD diagnosis or whether there had ever been a clinical suspicion of ASD; following their responses; from the 32 comparison families, two were excluded because the child with DS also had autistic features, resulting in a comparison group of 30 families (29 fathers and 30 mothers. One father was not willing to participate). The ASD and DS parent groups were stratified for sex, mean age, and mean IQ, measured using a short form of the Dutch version of the Wechsler Scales (Wechsler 1997; WAIS-III-NL Pearson, 2005).

Proband Assessment

All clinical diagnoses according to DSM-IV criteria (American Psychiatric Association 2000) were confirmed using two standardized diagnostic measures: the Dutch version of the Autism Diagnostic Interview Revised (ADI-R; Rutter et al. 2003) and the Dutch version of the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 1999). Level of functioning was measured with a short form of the Wechsler Scales (Wechsler 1991, 1997; WISC-III-NL, Pearson, 2005; WAIS-III-NL Pearson, 2005) the Mullen scales (Mullen 1995) or the RAVEN progressive matrices (Raven 1995, 1996).

Table 1 Characteristics of the parents of the individuals with ASDs and of the parents of the children with DS

Parents of children with	ASD			DS		
	Total group	Fathers	Mothers	Total group	Fathers	Mothers
N	53	26	27	59	29	30
Mean age parent	45.6	47.2	44.0	46.5	46.7	46.4
(SD)	(9.3)	(9.9)	(8.6)	(8.9)	(9.3)	(8.6)
Mean verbal IQ parent	105.3	101.9	108.6	102.5	106.6	98.5
(SD)	(21.7)	(22.5)	(20.8)	(23.1)	(23.9)	(22.0)
Mean non-verbal IQ parent	103.3	101.2	105.4	105.8	105.6	106.0
(SD)	(17.4)	(16.4)	(18.4)	(19.0)	(17.6)	(20.7)
Mean total IQ parent	105.9	102.5	109.1	104.1	106.7	101.7
(SD)	(20.0)	(20.2)	(19.6)	(19.7)	(20.4)	(19.1)
Mean number of children within family	2.9			2.5		
(SD)	(1.2)			(0.94)		
Mean age proband	16.4			18.7		
(SD)	(8.2)			(9.1)		
% housekeeping or no job (e.g. retired)	17			20		
% lower education[a]	30	27	33	27	24	30
% higher education[b]	70	73	67	73	76	70
% lower level employment[c]	49	46	53	39	36	44
% higher level employment[d]	51	54	47	61	64	56
% history of language delay (n)	4 (2)	8 (2)	0	2 (1)	3 (1)	0
% significant reading problems/dyslexia (n)	4 (2)	4 (1)	4 (1)	2 (1)	3 (1)	0
IFS burden score (social scale)	40.1	37.8	42.2	29.6	28.9	30.3
(SD)	(6.6)	(6.2)	(6.3)	(5.1)	(4.8)	(5.3)

[a] Highest level of education finished with a diploma: primary, secondary or lower level vocational education

[b] Level of education finished with a diploma: higher level vocational education, higher professional or university

[c] Manually skilled employers

[d] Non-manually skilled employers, employers with their own company or employers supervising several others

Families were included if both probands met criteria for autism on the ADI-R or fell at most one point short in one of the domains and met either *autism* or *autism spectrum* criteria on the ADOS. All children were at least 4 years of age and had a non-verbal intelligence of more than 35 to increase the validity of the diagnosis. See supplementary Table 1 for the characteristics of the ASD probands.

Measures and Procedure

Assessment of the BAP

Parents were interviewed about themselves using the FHI-S and about their spouse using the FHI-I. The FHI-I and FHI-S included 40 items aimed to measure behavioural characteristics in childhood and in adulthood that are qualitatively similar but more subtle than the behavior usually observed in ASD. The items covered social communicative behavior such as pragmatic and conversational qualities, friendships, demonstrativeness and response to emotional cues. Other items covered rigid or repetitive behaviours, and perfectionism. A factor analysis showed two underlying factors. The first factor consisted of social-communication items plus rigidity. The Cronbach's alpha for this set of items was high. Based on conceptual and statistical grounds, this factor was recommended to be used for characterizing the BAP through an FHI total score (Parr et al., resubmitted). The IoI is a 20 item observational measure of social functioning devised to assess the BAP (Pickles et al. 2013). A more detailed description of the development and measurement properties of the FHI and IoI instruments is provided by Parr et al. (resubmitted) and Pickles et al. (2013).

Ratings were based on detailed descriptions of behavior on a dimensional scale. Behaviors were scored as '0' (behavior does not reach scoring threshold); '1' (difficulties of the type specified, but not associated with impairment); or '2' (associated impairment). All parents were visited in their homes by two interviewers in order to carry out independent subject and informant interviews. Interviewers were not told the subjects' groups, but inevitably the

interviews revealed the group. Accordingly, for practical purposes the interviews and observations were not blind. At the end of the visit, the IoI was completed by the researcher administering the FHI-S. By means of this schedule, the observations of the interviewer during the home visit were recorded and rated with the same scoring system as the FHI. Following training, case vignettes were written and rated by investigators from the IMGSAG consortium from different sites, independently of each other and blind to the subject's family of origin. Rating inconsistencies were resolved by consensus. Parents were also assessed using the Pragmatic Rating Scale (Landa et al. 1992).

Assessment of Parental Burden

The Impact on the Family Scale (IFS) was used to measure the parental burden of raising children with a disorder and the impact on parental social life. This 27-item scale developed by Stein and Riessman (1980) to assess the burden in parents of a child with a chronic illness, has also been found to be useful in parents of children with behavioral problems (Sheeber and Johnson 1992). The scale was translated into Dutch (Hunfeld et al. 1999) and cross-cultural comparison was found to be adequate (Kolk et al. 2000). The scale measures the social/familial and financial impact of chronic disease (Stein and Jessop 2003). In this study, we present data for the social subscale only.

Statistical Analysis

All analyses were undertaken in Stata 11 (StataCorp 2009). Early analysis had shown substantial differences in rates of item endorsement and item sum scores for men and women. Analyses were therefore undertaken either for men and women separately or by accounting for sex differences by stratification or dummy variable adjustment. For each instrument the association of the proband type of each family to each item response was assessed separately for men and women using exact logistic regression as positive item endorsements were sometimes rare. The association of item sum scores with proband type was assessed using ordinal logistic regression and further analyzed using the receiver operating curve methods of Pepe and colleagues (Janes and Pepe 2008; Pepe and Longton 2005), which provided a pooled estimate of the Area-Under-Curve (AUC) after stratification by gender. This was undertaken using the comproc procedure with the tied score option. Plots of the non-parametric ROC were obtained from the procedure roccomp. For analyses that pooled mothers and fathers, in order to take account of the potentially correlated nature of these responses from the same family, all reported significance levels were obtained from Wald test

statistics calculated using the cluster-robust sandwich estimator of the parameter variance–covariance matrix (Binder 1983). For the same reason confidence intervals for pooled AUC estimates were derived by clustered bootstrap resampling (1,000 samples).

Results

Sample Description

There were no significant differences in mean age, IQ (verbal and non-verbal), level of education or level of employment between the groups, nor between the fathers or mothers of the ASD families compared with the fathers or mothers of the DS probands separately (Table 1). Most parents were in their first marriage (79 % in the ASD group and 76 % in the comparison group). Eight of the parents in the ASD group and seven of the parents in the comparison group mentioned that they had voluntarily changed a job because of difficulties with colleagues or supervisors. Two of the mothers in the ASD group had epilepsy. Parental burden of raising a child with a developmental disability was significantly higher for families with ASD probands, for both fathers $[t(53) = 6.2, p < .001]$ and mothers $[(t(55) = 7.7, p < .001]$.

Item Level Discrimination

Table 2 shows discrimination between the two groups of relatives for the items previously selected for better test–retest and convergent validity (Table 2). All estimated coefficients are in the expected direction. On the whole the FHI items discriminated among fathers more readily than among mothers, and informant reports discriminated more strongly than subject reports, such that while all items with the exception of aloofness gave significant discrimination ($p < .05$) for mothers' reports on fathers, there were no individually discriminating items for mother's reporting about themselves. For the IoI, especially among mothers, positive ratings were often too rarely given to properly assess the discriminating ability of individual items (Table 3). Nonetheless all estimated coefficients were in the expected direction and among men often significantly so.

Discrimination of Item Sum Scores

For fathers ordinal logistic regression indicated that the sum-scores were significantly associated with proband type, for the FHI childhood items ($p = .004$ subject, $p = .001$ informant) and FHI adult items (subject $p < .001$, informant $p = <.001$) and for the IoI ($p = .002$). For

Table 2 Discriminant validity of the selected Family History Interview items: logodds coefficients[a] of item score for being an ASD family parent mid-*p* values from exact logistic regression

Item	Informant adult items				Subject adult items			
	Female Coeff p value		Male Coeff p value		Female Coeff p value		Male Coeff p value	
Lack of interest in conversation	1.6	.16	1.2	.01	0.1	.47	1.8	.03
Reciprocal conversation	1.9	.08	1.5	.00	1.3	.31	1.2	.03
Pragmatics	0.9	.23	1.2	.01	0.1	.90	1.4	.01
Aloof	1.5	.10	0.8	.07	NA	NA	2.3	.00
Friendships	1.1	.21	2.9	.00	0.4	.61	1.5	.00
Affection	0.6	.31	2.1	.00	0.9	.18	0.5	.28
Intimacy	0.9	.36	0.9	.03	0.1	.95	1.5	.05
Emotional cues and responsiveness	1.6	.16	0.9	.04	0.4	.61	2.4	.01
Demononstrativeness	0.7	.21	1.3	.00	1.1	.21	0.5	.41
Social behaviour	1.7	.07	1.0	.02	0.9	.36	0.9	.09
Rigidity	1.3	.02	1.4	.01	1.4	.11	1.6	.06
	Childhood items				Childhood items			
Lack of interest in conversation	1.9	.08	1.3	.01	1.5	.10	1.8	.00
Reciprocal conversation	1.8	.04	1.1	.06	1.5	.10	2.3	.00
Aloof	1.0	.22	1.3	.00	0.1	.47	1.7	.01
Friend	1.4	.11	0.8	.07	0.9	.36	0.9	.06
Affection	2.2	.02	0.7	.13	0.4	.61	0.4	.34
Emotional cues and responsiveness	0.1	.92	0.9	.09	0.1	.86	0.5	.20
Demonstrativeness	1.1	.21	0.7	.13	0.4	.61	0.5	.34
Social behaviour	1.3	.31	0.9	.28	1.1	.16	1.6	.00
Rigidity	2.6	.00	1.9	.03	0.1	.92	0.9	.14

[a] Median unbiased estimates reported where the conditional maximum likelihood estimate was infinite. NA = too few positive ratings to compute

mothers significant discrimination was found for the sum scores from the IoI (*p* = .002) and informant FHI on the childhood items (*p* = .002) but only marginally so for informant adult items (*p* = .070) and subject childhood items (*p* = 0.066) and not for subject adult items (*p* = .182). Further analysis examined the overall scores obtained by summing over both childhood and adult items. Figure 1 shows the inverse cumulative distribution functions for the item sum-scores from each of the three instruments for the mothers and fathers of each proband type. For the FHI these findings suggest that the discrimination is better for fathers than mothers and better for the informant than the subject version. For the IoI, though the discrimination was potentially good, the range of the scores was limited, especially for women.

Figure 2 shows the nonparametric ROC curves for these instruments and shows that although the same rank order of AUCs is maintained such that IoI > informant FHI > subject FHI, the AUCs are consistently lower for females than males. The estimate of the AUCs pooled but stratified by sex were for the FHI-I 0.75 (95 % CI 0.66, 0.83), for the FHI-S 0.72 (95 % CI 0.63, 0.81) and for the IoI 0.80 (95 % CI 0.73, 0.86). For comparison, the estimate of the AUC for the Pragmatic Rating Scale was 0.57 (95 % CI 0.45, 0.68).

Cut-Offs and "prevalence"

The cut-offs for mothers that achieve most nearly equal sensitivity and specificity were FHI-I ≥ 1 (48 % sensitivity, 93 % specificity), FHI-S ≥ 2 (52 % se, 67 % sp) and IoI ≥ 2 (52 % se, 67 % sp). For fathers the equivalent cut-offs were ≥1 (81 % sensitivity, 93 % specificity), ≥7 (73 % se, 72 % sp), ≥6 (65 % se, 72 % sp). The corresponding prevalence estimates based on these cut-offs were 6, 3 and 3 % for mothers of DS probands and 48, 52 and 52 % for mothers of ASD probands. For fathers of DS probands the prevalence estimates were 7, 28 and 28 % and for fathers of ASD probands 80, 73 and 65 %.

Cut-offs that achieved the highest likelihood ratio were FHI-I ≥ 1, FHI-S ≥ 4 and IoI ≥ 6 for females and correspondingly ≥2, ≥14 and ≥13 for males, but all these cut-offs gave sensitivities for proband type of <50 %. The prevalence estimates based on these cut-offs were for

Table 3 Discriminant validity of the selected impressions of informant items: logodds coefficients[a] of item score for being an autism family parent and mid-*p* values from exact logistic regression

Item	Female		Male	
	Coeff	Mid-*p* value	Coeff	Mid-*p* value
Lack of movement/ Overactivity	NA		1.0	.22
Reduced or unusual greeting	2.02	.02	1.7	.00
Eye contact	NA		1.5	.00
Social smiling	NA		1.6	.04
Shy	NA		NA	
Facial expression	1.2	.10	2.2	.00
Unguarded/Guarded	0.1	.95	NA	
Perfectionism	NA		1.2	.10
Rigidity	0.1	.53	1.9	.03
Repeated reference to specific topics	1.5	.10	0.5	.40
Behavioural disinhibition	1.6	.04	0.7	.22

[a] Median unbiased estimates reported where the conditional maximum likelihood estimate was infinite. NA = too few positive ratings to compute

mothers of DS probands 6, 3 and 3 %, and for mothers of ASD probands 48, 41 and 22 %. For fathers of DS probands 3, 3 and 3 % and for fathers of ASD probands 54, 46 and 38 %. In the DS sample 7 relatives were identified by at least one of the three instruments. In the ASD sample 18 of the 53 relatives were identified by only 1 instrument, 12 by two and 8 by all three instruments.

Multivariate Discrimination

Discrimination using multiple instruments ordinal logistic regressions with sums-scores from each instrument were all highly significantly predicted by proband type ($p = .001$). In this sample the FHI-I and FHI-S were correlated 0.72, the FHI-S and IoI 0.57 and the FHI-I and IoI 0.45. When all three instrument sum-scores were included together in logistic regression predicting relative type, the FHI-I retained its significance ($p < .001$), as did the IoI ($p = .001$) but the FHI-S did not add to the prediction ($p = 0.981$). When included pairwise, the FHI-S contributed significantly ($p = .036$) in the presence of the FHI-I but not ($p = .118$) in the presence of the IoI. The AUC estimates from the logistic regressions without adjustment for sex were 0.73 for FHI-I alone, 0.77 for IoI alone and 0.79 with both FHI-I and IoI combined. The addition of the PRS did not significantly improve prediction ($p = 0.534$).
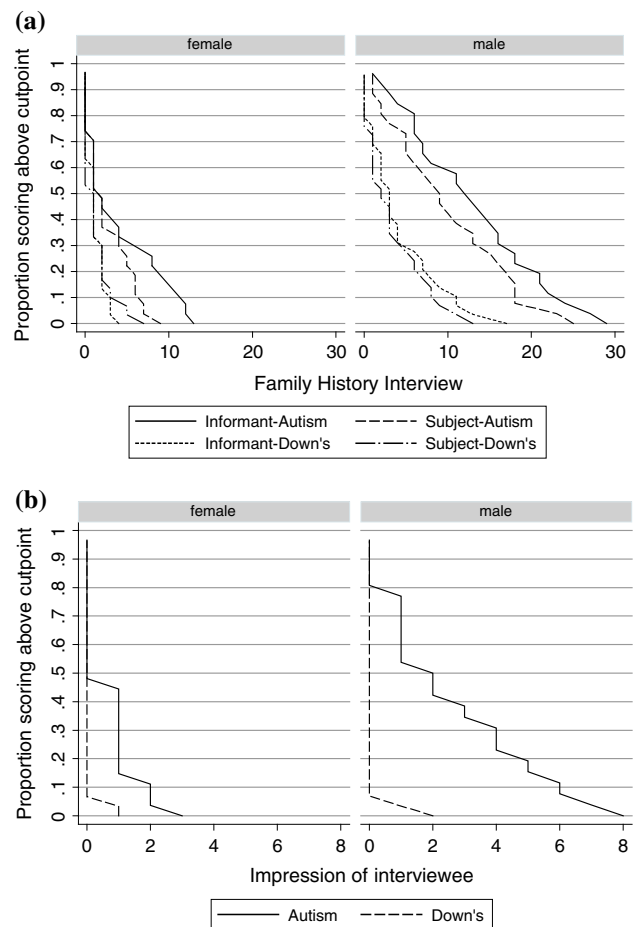

Fig. 1 **a** Inverse cumulative distribution for the item sum-scores on the FHI. **b** Inverse cumulative distribution for the item sum-scores on the IoI

Family Burden

As shown in Table 1, the social burden as measured by the IFS for fathers and mothers with a DS child was significantly less than for the parents of a child with an ASD, albeit still quite substantial. It was essential to check whether this difference could account for the greater 'BAP' scores for parents of children with an ASD. No significant correlations were found for the two groups of parents pooled (r = .17; n.s.) or in the ASD group on its own (r = −.24; n.s.).

Discussion

The BAP findings reported here are encouraging in indicating that both the interview properties (Parr et al. resubmitted) and observation measures (Pickles et al. 2013) provided a quite good differentiation between the ASD and DS families. The differentiation was best for the IoI
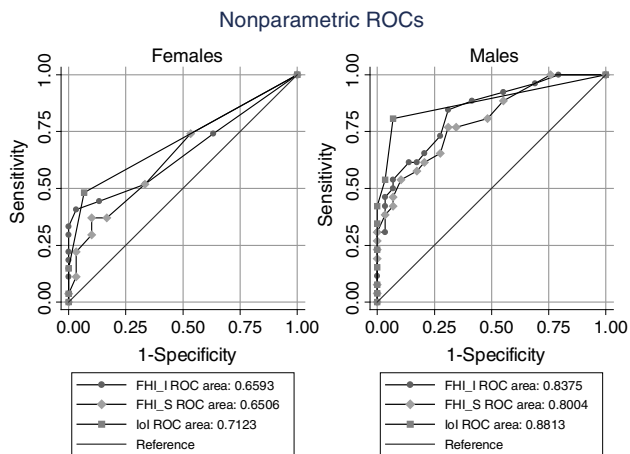
Fig. 2 Nonparametric ROC curves of the FHI and IoI

observation measure, next best for the FHI-I and not quite as good for the FHI-S, but the level of differentiation varied little among the three measures. In interpreting the findings, attention needs to be paid to the intercorrelation among the measures. The correlation between the FHI-I and the FHI-S was greatest (.72), was .57 between the FHI-S and the IoI (which were not done blind to each other), and was .45 between the FHI-I and the IoI. When the three measures were dealt with together, it was found that both the FHI-I and the IoI retained their significant group differentiation ($p < .001$), but the FHI-S did not improve group differentiation. It was also the case (see Table 2) that on the whole, FHI-I items provided a slightly better group differentiation of fathers than did the FHI-S.

The correlation between the two interview measures was not so high that one measure was redundant once the other was available. When the two interview measures were considered as a pair, the FHI-S did make a significant contribution ($p = .036$). On the other hand, despite the intercorrelation between the FHI-S and the IoI being lower (.57), the FHI-S made no significant contribution once the IoI was used. It might be considered that this finding means that the FHI-S could be dropped but there are three reasons why that would not be warranted at this stage. First, the ROC curves showed that the FHI-I and FHI-S performed just as well as each other (the AUCs being .75 and .72 respectively). Second, while the AUC for the IoI was marginally the best (.78) it should also be borne in mind that the FHI-S is the semi-structured observational setting for the IoI and the two are essentially required together. Third, the sample size of the study was too small for there to be much confidence in the robustness of small differences among measures.

Comparable questions need to be posed about the apparent differentiation superiority of the IoI. Could this

finding be due to possible bias arising from the interaction being rated unblinded to the group, although the raters of vignettes were kept blind, and by being potentially informed by the responses given during the FHI-S? Possibly, but it is noteworthy that the IoI correlated almost as highly with the independent FHI-I ($r = .45$) as with the FHI-S that provided the basis for the observation (.57). We conclude that the group differentiation findings for the IoI suggest that it may be more valid than the less encouraging retest findings might suggest (Pickles et al. 2013). That is particularly the case in the light of the finding that the pragmatic rating scale did not significantly improve prediction and had a lower AUC (0.57). It may be inferred that the IoI usefully covers a range of BAP characteristics beyond pragmatic language impairments.

An important finding was that the group differentiation was much better for fathers than for mothers. The extent to which this was the case varied across instruments but it applied to all three measures. The finding that broader phenotype scores were much higher in fathers than in mothers was to be expected. ASDs are more common in males than females (Fombonne et al. 2011) and since findings from genetic studies (Le Couteur et al. 1996) provide evidence that ASD probably reflects a quantitative trait (Constantino 2011), and findings from neurocognitive studies suggest similarities between ASD and the BAP (Losh et al. 2011) it is likely that BAP features will also be more frequent in males than females. That does not, however, mean that group differentiation should necessarily be better in males.

An issue that inevitably arises with respect to sex differences is whether the cut-offs when scores are dealt with categorically should be varied to take account of the apparent difference in ease of recognition. Thus, for example, it used to be claimed that the criteria for conduct disorder ought to be less stringent for females than males. When this possibility was systematically tested by Moffitt et al. (2001), it was clear that there was no justification for adopting different cut-offs for males and females. The sex difference was, rather, explicable on the basis of the higher frequency of neurodevelopmental risk factors in males. In our analyses of the BAP sex differences, we have explored the possible use of different cut-offs for fathers and mothers; finding that this did provide a better balance between sensitivity and specificity. We are very aware, however, that our sample size was too small for any robust estimate of categorical diagnoses of BAP. The group differentiation was mainly informative in indicating that a valid categorization based on multiple instruments should be possible, and that the differentiation was clearest when combining different measures. What would clearly help in deciding how best to diagnose a valid BAP is to have some external criterion. Such a criterion is not as yet available

but the obvious possibilities lie in biomarkers such as being used in 'baby sibling' studies (Elsabbagh and Johnson 2010), functional and structural neuroimaging or cognitive tasks (Sucksmith et al. 2011), and susceptibility genes for ASD (Lamb 2011) when they have been identified.

Our data on group differentiation were confined to a comparison between families with an ASD proband and families with a DS proband. Our choice of this comparison was made in order to have a group for which there was no expectation that parents would have a raised rate of problems in social reciprocity and in which the proband presented an increase in social burden for the families. Whether or not our use of sampling from a Dutch DS Parent Association meant that parents were more or less likely than parents in the general population to show BAP features is not known. Similarly, it is possible that the parents of children with ASD may have been especially alert to social reciprocity impairment. In that connection, it is relevant that the observational measure, which would not have that limitation, was such a good differentiator. By ensuring that vignettes were rated by researchers who were kept blind to group, we sought to eliminate rating bias. Unfortunately, however, because the vignettes were prepared by the investigators who conducted the self-report interview, this meant that they were unlikely to be blind to group, although they were not told the group of the informants. Writing vignettes and scoring by independent 'blind' raters is a strength of this study, but is also time consuming and expensive in clinical practice.

A limitation of the study was the relatively small sample size. However, item selection was very largely based on reliability analyses within a much larger sample, so the estimates are unlikely to suffer much bias towards optimistic performance characteristic of most studies of this size. We therefore believe the findings likely to generalize to larger samples. Nonetheless, there can be no assumption that all social difficulties reflect BAP features, there is now the need to test the extent to which our instruments can separate BAP features from, say, social anxiety or schizotypal disorders. In the meanwhile, our findings are encouraging in their suggestion that the measures do carry an important degree of validity.

References

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR^{TM})*. Washington, DC: Am. Psychiatr. Assoc. Press.

Bailey, A., Palferman, S., Heavey, L., & Le Couteur, A. (1998). Autism: The phenotype in relatives. *Journal of Autism and Developmental Disorders, 28*(5), 369–392.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review, 51*, 279–292.

Constantino, J. N. (2011). Autism as a quantitative trait. In D. G. Amaral, G. Dawson, & D. H. Geschwind (Eds.), *Autism spectrum disorders* (pp. 510–520). New York: Oxford University Press.

Dawson, G., Estes, A., Munson, J., Schellenberg, G., Bernier, R., & Abbott, R. (2007). Quantitative assessment of autism symptom-related traits in probands and parents: Broader phenotype autism symptom scale. *Journal of Autism and Developmental Disorders, 37*(3), 523–536.

Elsabbagh & Johnson. (2010). Getting answers from babies about autism. *Trends in Cognitive Sciences, 14*(2), 81–87.

Fombonne, E., Quirke, S., & Hagen, A. (2011). Epidemiology of pervasive developmental disorders. In D. G. Amaral, G. Dawson, & D. H. Geschwind (Eds.), *Autism spectrum disorders* (pp. 90–111). New York: Oxford University Press.

Ghaziuddin, M. (2005). A family history study of Asperger syndrome. *Journal of Autism and Developmental Disorders, 35*(2), 177–182.

Hunfeld, J. A., Tempels, A., Passchier, J., Hazebroek, F. W., & Tibboel, D. (1999). Brief report: Parental burden and grief one year after the birth of a child with a congenital anomaly. *Journal of Pediatric Psychology, 24*, 515–520.

Janes, H., & Pepe, M. S. (2008). Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: An old concept in a new setting. *American Journal of Epidemiology, 168*, 89–97.

Klin, A., Pauls, D., Schultz, R., & Volkmar, F. (2005). Three diagnostic approaches to Asperger syndrome: Implications for research. *Journal of Autism and Developmental Disorders, 35*(2), 221–234.

Kolk, A. M., Schipper, J. L., Hanewald, G. J., Casari, E. F., & Fantino, A. G. (2000). The Impact-on-Family Scale: A test of invariance across culture. *Journal of Pediatric Psychology, 25*, 323–329.

Lamb, J. A. (2011). Whole genome linkage and association analyses. In D. G. Amaral, G. Dawson, & D. H. Geschwind (Eds.), *Autism spectrum disorders* (pp. 669–689). New York: Oxford University Press.

Landa, R., Piven, J., Wzorek, M., Gayle, J. O., Chase, G. A., & Folstein, S. E. (1992). Social language use in parents in autistic individuals. *Psychological Medicine, 22*(1), 245–254.

Le Couteur, A., Bailey, A., Goode, S., Pickles, A., Robertson, S., Gottesman, I., et al. (1996). A broader phenotype of autism: The clinical spectrum in twins. *Journal of Child Psychology and Psychiatry, 37*(7), 785–801.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *ADOS. Autism diagnostic observation schedule. Manual*. Los Angeles: Western Psychological Services.

Losh, M., Adolphs, R., & Piven, J. (2011). The broad autism phenotype. In D. G. Amaral, G. Dawson, & D. H. Geschwind

(Eds.), *Autism spectrum disorders* (pp. 457–476). New York: Oxford University Press.

Losh, M., Childress, D., Lam, K., & Piven, J. (2008). Defining key features of the broad autism phenotype: A comparison across parents of multiple- and single-incidence autism families. *American Journal of Medical Genetics B: Neuropsychiatric Genetics, 147B*(4), 424–433.

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study* (p. 2001). Cambridge: Cambridge University Press.

Mullen, E. (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service Inc.

Pepe, M. S., & Longton, G. (2005). Standardizing diagnostic markers to evaluate and compare their performances. *Epidemiology, 16*, 598–603.

Pickles, A., Parr, J. R., Rutter, M. L., De Jonge, M. V., Wallace, S., Le Couteur, A. S., et al. (2013). New interview and observation measures of the Broader Autism Phenotype: Impressions of interviewee measure. *Journal of Autism and Developmental Disorders, 43*(9), 2082–2089.

Pickles, A., Starr, E., Kazak, S., Bolton, P., Papanikolaou, K., Bailey, A., et al. (2000). Variable expression of the autism broader phenotype: Findings from extended pedigrees. *Journal of Child Psychology and Psychiatry, 41*(4), 491–502.

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: Evidence from a family history study of multiple-incidence autism families. *American Journal of Psychiatry, 154*(2), 185–190.

Piven, J., Wzorek, M., Landa, R., Lainhart, J., Bolton, P., Chase, G. A., et al. (1994). Personality characteristics of the parents of autistic individuals. *Psychological Medicine, 24*(3), 783–795.

Raven, J. C. (1995). *Colored progressive matrices sets I and II*. Oxford: Oxford Psychologists Press Ltd.

Raven, J. C. (1996). *Progressive Matrices: A perceptual test of intelligence. Individual form*. Oxford: Oxford Psychologists Press Ltd.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *ADI-R. Autism diagnostic interview revised. Manual*. Los Angeles: Western Psychological Services. Dutch version, Hogrefe Amsterdam.

Sheeber, L. B., & Johnson, J. H. (1992). Applicability of the Impact on Family Scale for assessing families with behaviorally difficult children. *Psychological Reports, 71*, 155–159.

Starr, E., Berument, S. K., Pickles, A., Tomlins, M., Bailey, A., Papanikolaou, K., et al. (2001). A family genetic study of autism associated with profound mental retardation. *Journal of Autism and Developmental Disorders, 31*(1), 89–96.

StataCorp. (2009). *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.

Stein, R. E., & Jessop, D. J. (2003). The impact on family scale revisited: further psychometric data. *Journal of Developmental and Behavioral Pediatrics, 24*, 9–16.

Stein, R. K., & Riessman, C. K. (1980). The development of an Impact-on-Family Scale: Preliminary findings. *Medical Care, 1980*(28), 465–472.

Sucksmith, E., Roth, I., & Hoekstra, R. A. (2011). Autistic traits below the clinical threshold: Re-examining the broader autism phenotype in the 21st century. *Neuropsychology Review, 21*, 360–389.

Szatmari, P., MacLean, J. E., Jones, M. B., Bryson, S. E., Zwaigenbaum, L., Bartolucci, G., et al. (2000). The familial aggregation of the lesser variant in biological and nonbiological relatives of PDD probands: A family history study. *Journal of Child Psychology and Psychiatry, 41*(5), 579–586.

Wechsler D. (1991). Wechsler Intelligence Scale for Children-III-NL. 3rd ed. Amsterdam, The Netherlands: Pearson-NL 2005.

Wechsler, D. (1997). *WAIS-III, Nederlandstalige bewerking*. Amsterdam: Pearson-NL.