Table 4 continued

Items	(1) AD [N = 138]	(2) PDDNOS [N = 89]	(3) Non-ASD [N = 90]	Statistics
53. Offering to share	2.0 (0.9)	1.7 (1.1)	0.2 (0.5)	$F(2, 227) = 75.4$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: NS
54. Seeking to share enjoyment with others	1.4 (0.7)	1.2 (0.8)	0.1 (0.3)	$F(2, 229) = 81.5$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: NS
A4. Lack of socioemotional reciprocity	6.0 (2.1)	4.3 (2.2)	0.7 (1.1)	$F(2, 314) = 226.5$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p < 0.001$
31. Use of other's body to communicate	1.0 (1.2)	0.8 (1.0)	0.2 (0.5)	$F(2, 273) = 226.5$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: NS
55. Offering comfort	2.1 (1.1)	1.7 (1.3)	0.0 (0.2)	$F(2, 231) = 76.5$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: NS
56. Quality of social overtures	1.7 (1.2)	1.2 (1.1)	0.1 (0.2)	$F(2, 225) = 49.9$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p = 0.02$
58. Inappropriate facial expression	0.9 (0.8)	0.4 (0.6)	0.0 (0.3)	$F(2, 293) = 48.5$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p < 0.001$
59. Appropriateness of social responses	1.7 (1.1)	1.4 (1.2)	0.2 (0.6)	$F(2, 227) = 47.1$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p < 0.001$
A. Quantitative abnormalities in eciprocal social interaction	19.9 (5.3)	14.8 (6.4)	2.3 (2.7)	$F(2, 314) = 330.6$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p < 0.001$
B1. Lack of, or delay in, spoken language and failure to compensate through gesture	4.1 (2.5)	3.0 (2.2)	0.6 (1.2)	$F(2, 314) = 79.1$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p < 0.001$
42. Pointing to express interest	1.2 (0.9)	0.9 (0.9)	0.1 (0.4)	$F(2, 227) = 38.4$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: NS

Table 4 continued

Items	(1) AD [N = 138]	(2) PDDNOS [N = 89]	(3) Non-ASD [N = 90]	Statistics
43. Nodding	0.9 (0.8)	0.4 (0.6)	0.0 (0.2)	$F(2, 314) = 33.8$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p = 0.01$ $1 > 2$: $p < 0.001$
44. Head shaking	0.8 (0.9)	0.5 (0.8)	0.1 (0.2)	$F(2, 224) = 21.3$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p = 0.003$ $1 > 2$: $p = 0.03$
45. Conventional/instrumental gesture	1.4 (1.0)	0.9 (1.0)	0.1 (0.3)	$F(2, 228) = 41.7$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p = 0.002$
B4. Lack of varied spontaneous make-believe or social imitative play	4.2 (1.8)	2.8 (2.0)	0.6 (1.1)	$F(2, 314) = 124.9$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p < 0.001$
47. Spontaneous imitation of actions	2.2 (1.1)	1.7 (1.2)	0.2 (0.6)	$F(2, 314) = 72.0$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p < 0.001$
48. Imaginative play	2.0 (1.1)	1.5 (1.1)	0.2 (0.6)	$F(2, 227) = 124.9$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p = 0.004$
61. Imitative social play	1.5 (0.9)	1.1 (1.0)	0.0 (0.1)	$F(2, 226) = 53.9$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p = 0.003$
B2(V). Relative failure to initiate or sustain conversational interchange	3.1 (1.3)	1.9 (1.6)	0.5 (1.1)	$F(2, 307) = 97.9$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p < 0.001$
34. Social verbalization/chat	1.7 (0.6)	1.4 (0.8)	0.5 (0.7)	$F(2, 314) = 67.5$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p = 0.01$
35. Reciprocal conversation	1.8 (0.7)	1.4 (0.8)	0.2 (0.6)	$F(2, 242) = 112.6$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p = 0.005$
B3(V). Stereotyped, repetitive, or idiosyncratic speech	2.9 (1.8)	2.1 (1.8)	0.9 (1.3)	$F(2, 314) = 41.2$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p < 0.001$ $1 > 2$: $p = 0.004$
33. Stereotyped utterances and delayed echolalia	1.1 (1.1)	0.6 (0.8)	0.1 (0.4)	$F(2, 257) = 30.2$, $p < 0.001$ $1 > 3$: $p < 0.001$ $2 > 3$: $p = 0.008$ $1 > 2$: $p < 0.001$

Table 4 continued

Items	(1) AD [N = 138]	(2) PDDNOS [N = 89]	(3) Non-ASD [N = 90]	Statistics
36. Inappropriate questions or statements	1.2 (0.8)	0.6 (0.7)	0.3 (0.5)	$F(2, 258) = 45.7$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p = 0.02$ 1 > 2: $p < 0.001$
37. Pronominal reversal	0.3 (0.7)	0.1 (0.4)	0.2 (0.5)	$F(2, 221) = 2.0$, $p = 0.13$ NS
38. Neologisms/idiosyncratic language	0.4 (0.7)	0.2 (0.4)	0.2 (0.4)	$F(2, 257) = 5.9$, $p = 0.003$ 1 > 3: $p = 0.02$ 2 > 3: NS 1 > 2: $p = 0.01$
BV. Qualitative abnormalities in communications, verbal subjects	14.3 (4.1)	9.7 (4.4)	2.5 (3.2)	$F(2, 260) = 210.9$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p < 0.001$ 1 > 2: $p < 0.001$
BNV. Qualitative abnormalities in communications, non-verbal subjects	12.6 (4.9)	9.0 (4.4)	2.3 (2.5)	$F(2, 51) = 21.0$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p < 0.005$ 1 > 2: $p = 0.02$
C1. Encompassing preoccupation or circumscribed pattern of interest	1.9 (1.1)	0.9 (1.0)	0.3 (0.6)	$F(2, 314) = 80.6$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p < 0.001$ 1 > 2: $p < 0.001$
67. Unusual preoccupation	1.0 (0.9)	0.4 (0.7)	0.1 (0.3)	$F(2, 303) = 40.3$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p = 0.006$ 1 > 2: $p < 0.001$
68. Circumscribed interest	1.1 (0.8)	0.5 (0.7)	0.2 (0.5)	$F(2, 294) = 40.5$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p < 0.001$ 1 > 2: $p < 0.001$
C2. Apparently compulsive adherence to non-functional routines or rituals	1.4 (1.2)	0.7 (1.1)	0.2 (0.6)	$F(2, 314) = 36.3$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p = 0.01$ 1 > 2: $p < 0.001$
39. Verbal rituals	0.8 (0.9)	0.4 (0.7)	0.1 (0.3)	$F(2, 314) = 20.6$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p = 0.03$ 1 > 2: $p = 0.004$
70. Compulsions/rituals	0.9 (1.0)	0.5 (0.9)	0.2 (0.5)	$F(2, 302) = 18.1$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: NS 1 > 2: $p = 0.002$
C3. Stereotyped and repetitive motor mannerisms	0.9 (0.9)	0.5 (0.8)	0.2 (0.6)	$F(2, 314) = 19.4$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p = 0.03$ 1 > 2: $p = 0.03$

Table 4 continued

Items	(1) AD [N = 138]	(2) PDDNOS [N = 89]	(3) Non-ASD [N = 90]	Statistics
77. Hand and finger mannerisms	0.4 (0.7)	0.2 (0.5)	0.1 (0.4)	$F(2, 302) = 9.6$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: NS 1 > 2: $p = 0.004$
78. Other complex mannerisms or stereotyped body movements	0.8 (0.9)	0.4 (0.7)	0.1 (0.4)	$F(2, 303) = 21.5$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p = 0.04$ 1 > 2: $p = 0.001$
C4. Preoccupations with part of objects or non-functional elements of material	1.4 (0.7)	0.8 (0.7)	0.3 (0.6)	$F(2, 314) = 66.5$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p < 0.001$ 1 > 2: $p < 0.001$
69. Repetitive use of objects or interest in parts of objects	1.2 (0.9)	0.5 (0.7)	0.2 (0.4)	$F(2, 303) = 59.1$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p = 0.006$ 1 > 2: $p < 0.001$
71. Unusual sensory interests	0.7 (0.7)	0.5 (0.6)	0.2 (0.5)	$F(2, 301) = 21.7$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p = 0.006$ 1 > 2: $p = 0.006$
C. Restricted, repetitive, and stereotyped patterns of behaviors	5.5 (2.4)	2.9 (2.5)	1.1 (1.8)	$F(2, 314) = 106.6$, $p < 0.001$ 1 > 3: $p < 0.001$ 2 > 3: $p < 0.001$ 1 > 2: $p < 0.001$

NS not significant

for all items, F test); the only exception was "Pronominal reversal (item 37)" ($p = 0.13$). For the post hoc analysis, the mean scores for all items, except item 37, differed significantly between the AD and non-ASD groups. In addition, the mean scores for all items differed significantly between the PDDNOS and the non-ASD groups, with the exception of "Neologism (item 38)" ($p = 0.87$, post hoc test with Bonferroni correction); "Compulsions (item 70)" ($p = 0.15$, post hoc test with Bonferroni correction); and "Hand and finger mannerisms (item 77)" ($p = 0.22$, post hoc test with Bonferroni correction).

As regards the subdomains (A1–A4, B1–B4, C1–C4), all showed a significant difference in mean scores across the three diagnostic groups using one-way ANOVA (AD vs. PDDNOS vs. non-ASD; $p < 0.001$ for all subdomains, F test; Table 4). For the post hoc analyses, the mean of all subdomain scores revealed a significant difference between the AD and non-ASD, PDDNOS and non-ASD, and AD and PDDNOS groups.

As for domains A, B (BV/BNV), and C, the mean scores for all 3 domains were significantly different across the three diagnostic groups with one-way ANOVA (AD vs. PDDNOS vs. non-ASD; $p < 0.001$ for all domains, F test; Table 4). For the post hoc analysis, the mean scores for all domains were significantly higher in the AD than in the non-ASD group ($p < 0.001$ for domains A, BV, BNV, and C, post hoc test with Bonferroni correction), and were higher in the PDDNOS than in the non-ASD group ($p < 0.001$ for domains A, BV, and C, $p = 0.005$ for domain BNV, post hoc test with Bonferroni correction). Likewise, the mean scores of all domains were significantly higher in the AD than in the PDDNOS group ($p < 0.001$ for domains A, BV, and C, $p = 0.02$ for domain BNV, post hoc test with Bonferroni correction).

Similar comparisons of mean scores of the three domains were repeated after stratification according to three age bands (<5:0 years, 5:0–9:11 years, and 10–19 years; see Appendix Table 3 in supplementary materials). For those individuals below 5 years of age, the mean scores for all domains were significantly higher in the AD (N = 11) than in the non-ASD group (N = 45)($p < 0.001$ for domain A, $p = 0.01$ for domain

BV, $p < 0.001$ for domain BNV, $p = 0.002$ for domain C, post hoc test with Bonferroni correction), and significantly higher in the PDDNOS (N = 33) than in the non-ASD group ($p < 0.001$ for domain A and BV, $p = 0.005$ for domain BNV, $p = 0.03$ for domain C, post hoc test with Bonferroni correction). However, no significant difference was found between the AD and PDDNOS groups in any of the domains ($p = 0.19$ for domain A, $p = 0.93$ for domain BV, $p = 0.33$ for domain BNV, $p = 0.29$ for domain C, post hoc test with Bonferroni correction). As for those individuals aged 5:0–9:11 years, the mean scores of all domains (A, BV, and C; note that no group comparison was conducted in domain BNV, because there was only one nonverbal subject in the non-ASD group in this age band) were significantly higher in the AD (N = 37) than in the non-ASD group (N = 28) ($p < 0.001$ for domains A, BV, and C, post hoc test with Bonferroni correction), and were significantly higher in the PDDNOS (N = 22) than in the non-ASD group ($p < 0.001$ for domains A, BV, and C, post hoc test with Bonferroni correction). Similarly, the mean scores for all three domains were significantly higher in the AD than in the PDDNOS group ($p = 0.01$ for domains A and C, $p = 0.03$ for domain BV, post hoc test with Bonferroni correction). As for those individuals aged 10–19 years, the mean scores for all three domains (A, BV, and C; no group comparison was conducted in domain BNV, because there was only one nonverbal subject in the non-ASD group in this age band) were significantly higher in the AD (N = 90) than in the non-ASD group (N = 17) ($p < 0.001$ for domains A, BV, and C, post hoc test with Bonferroni correction). Likewise, the mean scores for all domains except domain C were higher in the PDDNOS (N = 34) and non-ASD groups ($p < 0.001$ for domains A and BV, $p = 0.07$ for domain C, post hoc test with Bonferroni correction); moreover, for all domains, mean scores were also higher in the AD than in the PDDNOS group ($p = 0.002$ for domain A, $p < 0.001$ for domain BV and C, post hoc test with Bonferroni correction).

Again, the same analyses were conducted over two groups of IQ/DQ level (<70 vs. ≥70; see Appendix Table 4 in supplementary materials). For those individuals with an IQ/DQ of <70, the mean scores for all domains (A, BV/BNV, and C) were significantly higher in the AD (N = 18) than in the non-ASD (N = 8) group ($p < 0.001$ for domains A and C, $p = 0.007$ for domain BV and $p = 0.05$ for domain BNV, post hoc test with Bonferroni correction). The mean scores for domains A and BV were significantly higher in the PDDNOS (N = 9) than in the non-ASD group (N = 8), but no significant difference was found for domains BNV, and C ($p < 0.001$ for domain A, $p = 0.05$ for domain BV, $p = 0.13$ for domain BNV, $p = 0.99$ for domain C, post hoc test with Bonferroni correction). A significant difference in mean scores between the AD and PDDNOS groups was found only in domain C ($p = 0.99$ for domain A, $p = 0.08$ for domain BV, $p = 0.99$ for domain BNV, $p < 0.001$ for domain C, post hoc test with Bonferroni correction). In turn, for those individuals with an IQ/DQ of ≥70, mean scores for all domains were significantly higher in the AD (N = 120) than in the non-ASD group (N = 82) ($p < 0.001$ for domains A, BV, BNV, and C, post hoc test with Bonferroni correction), higher in the PDDNOS (N = 80) than in the non-ASD group ($p < 0.001$ for domains A, BV, C, $p = 0.002$ for domain BNV, post hoc test with Bonferroni correction), and higher in the AD than in the PDDNOS group ($p < 0.001$ for domains A, BV, C, $p = 0.01$ for domain BNV, post hoc test with Bonferroni correction).

Diagnostic Validity: Agreement with Consensus Clinical Diagnosis of AD

In our analysis of the overall diagnostic validity of the Japanese version of ADI-R, we found that across all individuals, the sensitivity, specificity, PPV, and NPV of the test were very high (92, 89, 87, and 93 %, respectively; Table 5). Similar results were also obtained for age groups 5:0–9:11 years and

Table 5 Diagnostic validity: agreement with consensus clinical diagnosis among those with algorithm diagnosis of AD

	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Consensus clinical diagnosis: *Autistic disorder* [N = 138]				
Algorithm diagnosis of AD: Domain A ≥ 10 AND (Domain BV ≥ 8 for verbal OR BNV ≥ 7 for non-verbal subjects) AND Domain C ≥ 3 AND Domain D ≥ 1 (Rutter et al. 2003)				
All individuals [N = 317]	92	89	87	93
Age: below 4:0 [N = 73]	53	92	55	92
Age: below 5:0 [N = 89]	55	92	50	93
Age: 5:0–9:11 [N = 87]	92	84	81	93
Age: 10:0 and older [N = 141]	97	90	95	94
IQ/DQ: below 70 [N = 35]	94	100	100	94
IQ/DQ: 70 and over [N = 282]	92	88	85	93

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)

older, and for IQ/DQ groupings below 70 and at 70 and above. Consistent with our initial hypotheses, the sensitivity and PPV for ages below 4 and below 5 years were both poor, i.e., between 50 and 55 %, respectively.

Overall test sensitivity, or the proportion of individuals with AD (N = 138) who were correctly categorized as having AD using ADI-R-JV, was as high as 92 %, indicating excellent clinical significance, which was also shown for the evaluation of individuals aged 5:0–9:11 and age 10 years and older, and for those individuals at either cognitive level assessed, i.e., with a score of <70 or ≥70. However, for individuals aged below 5 years, a sensitivity of 55 % was found, indicating a poor level of clinical significance.

On the other hand, among individuals without a consensus clinical diagnosis of AD (N = 179), 159 were also judged not to have AD based on ADI-R-JV algorithm diagnosis, i.e., the specificity of ADI-R-JV for correctly excluding AD was 89 % (159/179), indicating excellent clinical significance. This clinically excellent specificity was replicated for individuals in each of the three age bands, and for both IQ/DQ bands examined.

Discussion

In the present study, we reported the inter-rater reliability, discriminant validity, and diagnostic validity of the Japanese Version of ADI-R (ADI-R-JV).

Reliability of ADI-R-JV

In agreement with our hypotheses, the Kw values for all algorithm items of ADI-R-JV exceeded a value of 0.6, which was also consistent with the findings of previous studies (Hill et al. 2001; Lord et al. 1994). Furthermore, among the 42 algorithm items, all but two (items 39 and 58) showed Kw values in excess of 0.75, indicating excellent inter-rater reliability; the two exceptions showed Kw values of 0.60–0.75, indicating good inter-rater reliability.

We also investigated whether the measures for inter-rater reliability would decrease when the analysis was limited to individuals in specific age bands (Table 3). Again, the ICCs for all domains and subdomains exceeded 0.75 (excellent) among individuals aged less than 5 years, and the ICCs for all but 1 (C3) subdomain exceeded 0.75 (excellent) among individuals aged 5:0–9:11 years. Of note, ICCs can be seen as reflecting a good to excellent level of clinical significance, regardless of the age of the examinee. It is worth mentioning in this context that the ICCs became smaller in subdomains B2(V), B3(V), C1, C2, and C4 if the examinees were 10 years old or older. This tendency, i.e., smaller ICC values of the age band of 10 years and older, should first be discussed in light of the definition of inter-rater reliability, which can be easily compromised when the degree of experience and training of pairs conducting the interviews differs. When such a difference in experience occurred in the present study, compromised ICCs should have been observed irrespective of a subject's age, since the two raters were selected on a random basis from each site. Furthermore, the raters who administered ADI-R-JV were fully and equally experienced after the official training sessions. Therefore, the compromised ICCs for those subjects 10 years old and older did not seem to reflect a bias stemming from assessment skills. There is agreement between our findings and previous results showing lower scores for items under domain C than for items under domains A and B (Hill et al. 2001; Lord et al. 1994). Specifically, the inter-rater reliability of items under domain C would be particularly likely to be compromised when the examinees were older (i.e., 10 years and older), probably due to the uncertain recall of remote episodes. However, since we only obtained limited findings regarding inter-rater reliability upon assessment of adolescent subjects, elaboration on this topic remains difficult.

On the other hand, ICCs were not lower when the analysis was limited to the examination of individuals with an intellectual disability (IQ/DQ < 70), or when only males or females were included in the analysis (Table not shown). Rather, under no circumstances did we observe a Kw or ICC below 0.6 (Table 3). These findings strongly indicate the satisfactory inter-rater reliability of ADI-R-JV, i.e., the translated version appears to be as reliable as the original ADI-R in English.

Validity of ADI-R-JV

Discriminant Validity

Mean scores for three domains (A, B[BV/BNV], and C) were significantly higher in the AD group than in the PDDNOS and the non-ASD groups, indicating that the discriminant validity of ADI-R-JV was stable. Thus, our results appear to be consistent with the findings of previous pivotal studies investigating younger individuals with AD (Lord et al. 1993; Saemundsen et al. 2003), even in those with concomitant developmental delay (Gray et al. 2008).

Originally, ADI-R was designed to detect AD, not ASD (Lord et al. 1994). Therefore, in the current analysis, we expected not only that the mean scores for all domains would be higher in the AD group than in the non-ASD group, but also that they would be higher in the AD than in the PDDNOS group. These two hypotheses held true when the analysis included all study participants (N = 317).

However, the latter hypothesis (mean scores for AD > mean scores for PDDNOS) did not hold true when the analysis was limited to individuals less than 5 years of age (Appendix Table 3 in supplementary materials). Presumably, one of the main reasons for the compromised discriminability (i.e., no difference between AD and PDDNOS reflected in ADI-R-JV scores for younger individuals) was that it is difficult to differentiate AD from PDDNOS in individuals younger than 5 years of age (Turner and Stone 2007). On the other hand, the present finding may also have been due to biases; for instance, the diagnostic algorithms were prepared separately for those aged 4 years and older (based on current and past behavior) and for those younger than 4 years of age (based on current behavior). Thus, it would be possible that the discriminant validity would differ for individuals younger than 4 years old and for individuals between 4 and 5 years old. We thus analyzed a restricted sample of individuals below 4 years of age (N = 73), and found that the mean scores for domain A were 14.5 for AD, 11.4 for PDDNOS, and 3.1 for non-ASD. These results indicated that the mean was slightly higher in the AD group than in the PDDNOS group ($p = 0.051$, after Bonferroni correction), whereas the mean scores for domain BV/BNV and domain C did not reveal such differences between the AD and PDDNOS groups, suggesting that the choice of algorithm according to age may have at least partly affected the results for younger individuals.

As regards to the above results stratified by age, attention should be paid to our sample selection; among individuals below 5 years of age, 12 % had AD and 33 % had PDDNOS, whereas 64 % had AD and 24 % had PDDNOS among individuals who were 10 years old or older. These figures are consistent with differences in mean age across the three diagnostic groups shown in Table 2, and that a sample bias influenced the results. If we were to have recruited younger children with AD in the analysis, a higher level of discrimination among subgroups would likely have been observed.

Discriminant validity was also compromised for individuals with an intellectual disability (IQ/DQ < 70, see Appendix Table 4 in supplementary materials). Again, we expected that the mean scores for all domains were higher in the AD group than in both the non-ASD and PDDNOS groups. The first hypothesis (mean scores for AD > mean scores for non-ASD) held true for all domains, regardless of IQ level. However, the second hypothesis (mean scores for AD > mean scores for PDDNOS) held true only for domain C among individuals with an IQ/DQ of < 70; instead, the relationship of mean scores for PDDNOS > mean scores for non-ASD was not observed for domain C among individuals with an IQ/DQ of < 70. These results suggest that the relevance of domain C in arriving at a diagnosis of AD may differ from the relevance of domains A and B, particularly for individuals with a developmental delay. This issue has already been addressed in the literature; some authors have argued that the exclusion of domain C may improve discriminability between toddlers with and without ASD (Ventola et al. 2006). Furthermore, Lord and Jones (2012) reviewed that compared to symptoms under the social interaction and communication domains, symptoms under the repetitive behavior domain (domain C) are more heterogeneous across individuals and context-dependent, and thus caregivers may not consistently notify clinicians about domain-C symptoms. Our findings appear to be in line with the results of these previous studies. Specifically, individuals with a consensus diagnosis of AD with concomitant cognitive delay would be diagnosed as having Social Communication Disorder according to the proposed version in the DSM-5 (http://www.dsm5.org/ProposedRevision/Pages/NeurodevelopmentalDisorders.aspx), using ADI-R-JV. This issue still needs to be addressed in future studies.

Thus far, the overall discriminant validity of ADI-R-JV has been shown to be sufficient, although it appeared compromised for the assessment of younger individuals and individuals with concomitant cognitive delay. Potential biases and the limited statistical power of the present study should be noted, as these factors might have resulted in the finding of compromised discriminability among younger individuals.

As shown in Table 4, "Pronominal reversal (item 37)" showed no statistical difference among the three diagnostic groups. This finding was of interest in terms of language use, because in Japanese conversations, personal pronouns are not as frequently used as they are in English. In addition, even when personal pronouns are not used, there are no verbal conjugations in Japanese that correspond to those in Latin-derived languages. We are certain that this specific feature of the Japanese language allowed the mean scores on item 37 to remain fairly close to zero. Nevertheless, this concern did not in any way affect discriminability among domain scores, nor was diagnostic validity affected.

Diagnostic Validity

The sensitivity of ADI-R-JV with respect to correctly diagnosing autistic disorder was 92 %, indicating that the overall sensitivity of the instrument is excellent. Moreover, the algorithm's overall specificity, which was shown to be 89 %, was determined to be good. Likewise, the overall PPV and NPV were 89 and 93 %, respectively, indicating good to excellent clinical significance, consistent with our expectations. These figures were similar or even better than those obtained in a recent study using a translated version of ADI-R administered to individuals with a mean age of 10 years (Lampi et al. 2010). However, in the current study, the corresponding sensitivity decreased to 55 % (indicating

poor sensitivity; Table 5) when the analysis was limited to subjects younger than 5 years of age, suggesting that diagnostic validity was compromised in younger individuals. This finding was also consistent with our hypothesis. The compromised sensitivity for younger individuals may be rather straightforward; prior studies have been consistent with this finding, and our own results indicated compromised discriminability between AD and PDDNOS individuals below 5 years of age. However, as such compromised discriminability was not firmly upheld due to potential biases and the limited statistical power of our study sample, analysis of a larger number of individuals may have provided a higher level of sensitivity. Indeed, a recent large-scale study indicated a sensitivity for correctly diagnosing AD as high as 82.7 %, even when participants were under the age of 36 months (Risi et al. 2006). Nevertheless, it remains possible that the low level of sensitivity for those aged less than 5 years in the present study was not simply due to sample selection or the algorithm applied, but also a reflection of the difficulty of differentiating AD from PDDNOS in individuals at such young age, as was suggested by recent literature (Turner and Stone 2007).

In light of the proposed diagnosis of ASD in the forthcoming Diagnostic and Statistical Manual of Mental Disorders (version 5), research interests have increasingly focused on differentiating ASD from non-ASD individuals using ADI-R; however, there is no established cutoff for ASD in ADI-R. Attempts have been made to apply the original algorithm to ASD individuals; unfortunately, sensitivity for correctly diagnosing ASD was shown to be insufficient (Kim and Lord 2012; Risi et al. 2006). A related attempt to differentiate ASD from non-ASD individuals using ADI-R was the use of other assessment scales such as the Vineland Adaptive Behavior Scale (Sparrow et al. 1984) to improve sensitivity (Tomanik et al. 2007). Another attempt at differentiation was to relax the original, stringent algorithm for AD. For instance, in one genetic study (International Molecular Genetic Study of Autism Consortium 2001), the diagnosis of ASD was made according to ADI-R, whereby exceeding the cutoffs of three domains (A, B, C) was required for ASD diagnosis, with the exception that a score on any one of the three domains could fall one point below the threshold. We recalculated sensitivity using this relaxed criterion in the current study, resulting in an overall sensitivity of 64 %. When the same analysis was repeated for three age bands, sensitivity was 27 % for subjects aged < 5 years old, 71 % for subjects aged 5:0–9:11 years old, and 74 % for those 10 years old and older (Table not shown). At present, ADI-R-JV appears to have limited diagnostic validity with respect to detecting ASD.

Nevertheless, studies have emphasized that the use of ADOS together with ADI-R is a sensible approach, in that the combination of the two reflects consensus clinical judgments of AD as well as of ASD better than any other single instrument used alone (Le Couteur et al. 2008), even in individuals as young as 3 years old and younger (Risi et al. 2006). In this regard, evaluations of the sensitivity of both the Japanese version of ADOS and ADI-R-JV for correctly diagnosing ASD should be conducted.

It should also be noted that the sensitivity of ADI-R-JV with respect to correctly diagnosing AD among individuals with concomitant cognitive delay (IQ/DQ < 70) was 94 %, i.e., not lower than the corresponding result for individuals with an IQ of >70 (92 %); this findings was inconsistent with our expectations, as well as with a prior study (de Bildt et al. 2004). Furthermore, other studies have shown that specificity was more prone than sensitivity to be compromised when the examinee exhibited cognitive delay, and thus individuals with cognitive delay are more likely to be overdiagnosed (Lord et al. 1994; Risi et al. 2006). As regards the discrepancy with our hypothesis, the sample bias of the present study should be taken into account, because the mean IQ/DQ of individuals with AD and PDDNOS in this study was fairly high, even higher than reported in previous studies. In addition, the small number of enrolled participants with an IQ/DQ of <70 could have limited the statistical power of the study to detect any compromising effects of cognitive delay on diagnostic validity.

Limitations and Strengths

Treatment or interventions that may have affected the children enrolled in this study should also be taken into account, particularly in the assessment of diagnostic subgroups. It was a limitation of this study that we did not collect relevant data on this topic. On the other hand, ADI-R is a measure based principally on the observation of past behavior during early stages of development, and usually is employed prior to such interventions, and is not based on a patient's current status. This means that the scores we obtained were less likely to reflect intervention effects compared to the scores of instruments that assess current behaviors, such as ADOS. In addition, we observed good to excellent inter-rater reliability, discriminant validity, and diagnostic validity of ADI-R-JV even without considering treatment effects that would have been observed among clinically referred individuals. Considering that statistical tests are generally biased toward null hypotheses (no difference), an adjustment allowing for treatment effects, when examined, would increase the validity of the ADI-R-JV.

In the present study, clinically referred and control individuals were enrolled according to different protocols. If caregiver motivation to participate in this study differed

for the two groups of individuals examined, the difference may have been a substantial source of sample bias. The most likely scenario related to this issue would be that a caregiver of a control individual was highly motivated to participate in the study when there was a concern that the child may have had an undiagnosed psychiatric disorder such as ASD. Indeed, such motivation might have been reflected in high proportions of non-ASD psychiatric disorders; 2 out of 16 control individuals in the reliability study (Table 1) and 4 out of 82 control individuals had such a diagnosis. Parental education and socioeconomic status, when available, may have provided some insight into the extent of this problem, but unfortunately we did not collect such data, which might otherwise have helped to refute this scenario. However, if such a motivation to participate in the study had indeed been the case, it is likely that a number of individuals with ASD would have been detected among control individuals, yet there was not a single case of undiagnosed ASD (i.e., later detected as such) among individuals initially enrolled as controls (Table 1 and Appendix Table 2 in supplementary materials). To minimize this ambiguity, confirmatory studies will be necessary.

Consensus clinical diagnoses were obtained through clinical assessments and case reviews of all of the available information, albeit outside the context of the administration of ADI-R-JV. This approach might have led to a lack of information for optimizing the diagnosis, but it ensured the independence of the administration of the ADI-R-JV. Moreover, ADI-R-JV was administered in a blinded fashion without any reference to the clinical consensus diagnosis, which could also be considered as a strength of the present study.

When we finalized our consensus clinical diagnosis, it might have been helpful to facilitate diagnosis derived from ADOS. It may also have been helpful to adopt this protocol as an external criterion for estimating the validity of ADI-R-JV. Indeed, the Japanese translation of ADOS has been available to those who established the research reliability of ADOS (i.e., since 2010). Our research team consists of very experienced clinicians and clinical researchers, and among the 8 team members involved in establishing a consensus clinical diagnosis, 4 had already established, and 2 were planning to establish, the research reliability of the ADI-R; 3 had already established, and 3 were planning to establish, the research reliability of ADOS; and each member had participated in at least one research training session on either ADI-R or ADOS. Thus, all the team members involved in establishing a consensus clinical diagnosis were fully knowledgeable about the current diagnosis of ASD in a research setting.

Conclusions

ADI-R-JV is a reliable tool, and has sufficient ability to discriminate between individuals with AD and other diagnoses, as well as between individuals with AD and those with no psychiatric diagnosis. The sensitivity for correctly diagnosing AD was generally high (92 %), but appeared to be compromised (55 %) when the tool was used to assess children younger than 5 years of age. The specificity of ADI-R-JV was consistently high, regardless of the age and cognitive level of the examinee.

Acknowledgments Declaration of interest: None of the authors have any conflict of interest to declare. The authors thank Dr. Masahiro Oshima, M.D., for recruiting children with ASD. Funding of this study was provided by a Research Grant from the Ministry of Health, Labour and Welfare of Japan (H20-KOKORO-004: K.J. Tsuchiya, Y. Kamio), by a grant for the National Center for Child Health and Development (21S-3: K. J. Tsuchiya), by the Pfizer Health Research Fund (K. Matsumoto), and by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (A) (No. 20591396: N. Takei). These funding sources had no role in the study design; in the collection, analysis and interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication.

References

American Psychiatric Association. (2000). *Task force on DSM-IV. Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. Washington, DC: American Psychiatric Association.

Baron-Cohen, S., Scott, F. J., Allison, C., Williams, J., Bolton, P., Matthews, F. E., et al. (2009). Prevalence of autism-spectrum conditions: UK school-based population study. *British Journal of Psychiatry, 194*, 500–509.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.

Cicchetti, D. V., Lord, C., Koenig, K., Klin, A., & Volkmar, F. R. (2008). Reliability of the ADI-R: Multiple examiners evaluate a single case. *Journal of Autism and Developmental Disorder, 38*, 764–770.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127–137.

Cicchetti, D. V., Volkmar, F., Klin, A., & Showalter, D. (1995). Diagnosing autism using ICD-10 criteria: a comparison of neural networks and standard multivariate procedures. *Child Neuropsychology, 1*, 26–37.

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., et al. (1999). Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry, 40*, 719–732.

de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., et al. (2004). Interrelationship between autism diagnostic observation schedule-generic (ADOS-G), autism diagnostic interview-revised (ADI-R), and the diagnostic and statistical manual of mental disorders (DSM-IV-TR)

classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorder, 34*, 129–137.

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurements, 33*, 613–619.

Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research, 65*, 591–598.

Gilchrist, A., Green, J., Cox, A., Burton, D., Rutter, M., & Le Couteur, A. (2001). Development and current functioning in adolescents with Asperger syndrome: a comparative study. *Journal of Child Psychology and Psychiatry, 42*, 227–240.

Gray, K. M., Tonge, B. J., & Sweeney, D. J. (2008). Using the autism diagnostic interview-revised and the autism diagnostic observation schedule with young children with developmental delay: Evaluating diagnostic validity. *Journal of Autism and Developmental Disorder, 38*, 657–667.

Hill, A., Bolte, S., Petrova, G., Beltcheva, D., Tacheva, S., & Poustka, F. (2001). Stability and interpersonal agreement of the interview-based diagnosis of autism. *Psychopathology, 34*, 187–191.

International Molecular Genetic Study of Autism Consortium. (2001). A genomewide screen for autism: Strong evidence for linkage to chromosomes 2q, 7q, and 16p. *American Journal of Human Genetics, 69*, 570–581.

Kawamura, Y., Takahashi, O., & Ishii, T. (2008). Reevaluating the incidence of pervasive developmental disorders: impact of elevated rates of detection through implementation of an integrated system of screening in Toyota, Japan. *Psychiatry and Clinical Neurosciences, 62*, 152–159.

Kim, Y. S., Leventhal, B. L., Koh, Y. J., Fombonne, E., Laska, E., Lim, E. C., et al. (2011). Prevalence of autism spectrum sisorders in a total population ample. *American Journal of Psychiatry, 168*, 904–912.

Kim, S. H., & Lord, C. (2012). New autism diagnostic interview-revised algorithms for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Autism and Developmental Disorder, 42*, 82–93.

Kočovská, E., Biskupstø, R., Carina Gillberg, I., Ellefsen, A., Kampmann, H., Stora, T., et al. (2012). The rising prevalence of autism: A prospective longitudinal study in the Faroe Islands. *Journal of Autism and Developmental Disorder,*. doi: 10.1007/s10803-012-1444-9.

Koyama, T., Osada, H., Tsujii, H., & Kurita, H. (2009). Utility of the Kyoto scale of psychological development in cognitive assessment of children with pervasive developmental disorders. *Psychiatry and clinical neurosciences, 63*, 241–243.

Lampi, K. M., Sourander, A., Gissler, M., Niemela, S., Rehnstrom, K., Pulkkinen, E., et al. (2010). Validity of finnish registry-based diagnoses of autism with the ADI-R. *Acta Paediatrica, 99*, 1425–1428.

Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardised assessment instruments: The ADI-R and the ADOS. *Journal of Autism and Developmental Disorder, 38*, 362–372.

Lord, C., & Jones, R. M. (2012). Annual research review: Re-thinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 53*, 490–509.

Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry, 63*, 694–701.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., et al. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorder, 30*, 205–223.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorder, 24*, 659–685.

Lord, C., Storoschuk, S., Rutter, M., & Pickles, A. (1993). Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal, 14*, 234–252.

McGovern, C. W., & Sigman, M. (2005). Continuity and change from early childhood to adolescence in autism. *Journal of Child Psychology and Psychiatry, 46*, 401–408.

Mildenberger, K., Sitter, S., Noterdaeme, M., & Amorosa, H. (2001). The use of the ADI-R as a diagnostic tool in the differential diagnosis of children with infantile autism and children with a receptive language disorder. *European Child and Adolescent Psychiatry, 10*, 248–255.

Parner, E. T., Schendel, D. E., & Thorsen, P. (2008). Autism prevalence trends over time in Denmark: Changes in prevalence and age at diagnosis. *Archives of Pediatrics and Adolescent Medicine, 162*, 1150–1156.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*, 1094–1103.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *ADI-R: Autism diagnostic interview revised. WPS edition manual*. Los Angeles, CA: Western Psychological Services.

Saemundsen, E., Magnússon, P., Smári, J., & Sigurdardóttir, S. (2003). Autism diagnostic interview-revised and the childhood autism rating scale: Convergence and discrepancy in diagnosing autism. *Journal of Autism and Developmental Disorder, 33*, 319–328.

Sparrow, S. S., Balla, D., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales (VABS)*. Circle Pines, MN: American Guidance Service.

Tanaka Institute of Education. (1987). *Tanaka-Binet intelligence scale*. Tokyo: Taken Publisher.

Tomanik, S. S., Pearson, D. A., Loveland, K. A., Lane, D. M., & Bryant Shaw, J. (2007). Improving the reliability of autism diagnoses: Examining the utility of adaptive behavior. *Journal of Autism and Developmental Disorder, 37*, 921–928.

Turner, L. M., & Stone, W. L. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry, 48*, 793–802.

Ventola, P. E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J., et al. (2006). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorder, 36*, 839–847.

Waterhouse, L. (2008). Autism overflows: Increasing prevalence and proliferating theories. *Neuropsychology Review, 18*, 273–286.

Wechsler, D., Golombok, S., & Rust, J. (1992). *Wechsler intelligence scale for children—Third edition (WISC-IIIUK)*. London, UK: The Psychological Corporation.

Williams, J. G., Higgins, J. P., & Brayne, C. E. (2006). Systematic review of prevalence studies of autism spectrum disorders. *Archives of Disease in Childhood, 91*, 8–15.

Vaccine 30 (2012) 4292–4298

Contents lists available at SciVerse ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

The combined measles, mumps, and rubella vaccines and the total number of vaccines are not associated with development of autism spectrum disorder: The first case–control study in Asia

Yota Uno[a,b,*], Tokio Uchiyama[b,c], Michiko Kurosawa[d], Branko Aleksic[a], Norio Ozaki[a]

[a] Department of Psychiatry and Psychiatry for Parents and Children, Nagoya University Graduate School of Medicine, Nagoya, Japan
[b] Yokohama Psycho-Developmental Clinic, Yokohama, Japan
[c] Department of Faculty of Human Development, Fukushima University Graduate school, Fukushima, Japan
[d] Department of Epidemiology and Environmental Health, Juntendo University Graduate School of Medicine, Tokyo, Japan

ARTICLE INFO

Article history:
Received 9 September 2011
Received in revised form 6 January 2012
Accepted 7 January 2012
Available online 20 April 2012

Keywords:
Autism spectrum disorder
Vaccination
Measles–mumps–rubella Vaccine
Risk factors
Medical record
Case–control studies

ABSTRACT

Objective: The aim of this study was to investigate the relationship between autism spectrum disorder (ASD) and general vaccinations, including measles–mumps–rubella (MMR) vaccine, in Japanese subjects, a population with high genetic homogeneity.
Patients and methods: A case–control study was performed. Cases ($n=189$) were diagnosed with ASD, while controls ($n=224$) were volunteers from general schools, matched by sex and birth year to cases. Vaccination history and prenatal, perinatal, and neonatal factors from the Maternal and Child Health handbook, which was part of each subject's file, were examined. To determine the relationship between potential risk factors and ASD, crude odds ratios (ORs) and 95% confidence intervals (95% CIs) were calculated, and the differences in mean values of the quantitative variables between cases and controls were analyzed using an unpaired *t*-test. Moreover, MMR vaccination and the effect of the number of vaccine injections were investigated using a conditional multiple regression model.
Results: For MMR vaccination, the OR was 1.04 (95% CI, 0.65–1.68), and no significant differences were found for the other vaccines. For all of the prenatal, perinatal and neonatal factors, there were no significant differences between cases and controls. Furthermore, regarding the presence of ASD, MMR vaccination and the number of vaccine injections had ORs of 1.10 (95% CI, 0.64–1.90) and 1.10 (95% CI, 0.95–1.26), respectively, in the conditional multiple regression model; no significant differences were found.
Conclusions: In this study, there were not any convincing evidences that MMR vaccination and increasing the number of vaccine injections were associated with an increased risk of ASD in a genetically homogeneous population. Therefore, these findings indicate that there is no basis for avoiding vaccination out of concern for ASD.

1. Introduction

Autism is a life-long neurodevelopmental disorder. Its prevalence was long considered to be approximately 4 in 10,000 [1]. Due to broadening of the nosological categorization and more widespread recognition, however, in recent years the prevalence of autism spectrum disorder (ASD) [2,3], which includes autism, Asperger syndrome, and Pervasive developmental disorder not otherwise specified, has been reported at approximately 1% worldwide [4–6]. Although the pathogenesis of ASD has not yet been elucidated, genetic risk factors are strongly implicated, because the relative risk (λ_s) among siblings is greater than 20, and heritability is estimated to be as high as 38–90% [7–9]. In contrast, because the concordance rate of identical twins is not 100%, one can infer that environmental factors are also involved, and the recent increase in prevalence also indicates the involvement of various types of "novel environmental exposure". A debate has arisen over the contribution of vaccination as one environmental trigger of ASD.

The view that vaccination and ASD onset are related dates back to 1998 when the Lancet article by Wakefield et al. appeared [10] (the paper was retracted in 2010 because of ethical and methodological problems [11]). Thereafter, other published reports suggested a link between the measles–mumps–rubella vaccine

* Corresponding author at: Yokohama Psycho-Developmental Clinic, 7-7-2F Chigasaki-cyuo, Tsuzuki-ku, Yokohama-city, Kanagawa 224-0032, Japan. Tel.: +81 45 942 1077; fax: +81 45 942 1099.
E-mail addresses: yota_u@ypdc.net (Y. Uno), tokiouch@ca2.so-net.ne.jp (T. Uchiyama), mic@med.juntendo.ac.jp (M. Kurosawa), branko@med.nagoya-u.ac.jp (B. Aleksic), ozaki-n@med.nagoya-u.ac.jp (N. Ozaki).

0264-410X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.vaccine.2012.01.093

(MMR) and ASD [12–14], and concerns emerged that thimerosal, which is included in other vaccines as a preservative, and vaccination with combined vaccines might be risks for ASD onset [15–17]. Other studies, however, that examined retrospective data and rejected any such link were published in rapid succession [18–26]. For example, some reported an increase in ASD prevalence despite a decline in the MMR vaccination rate [27,28]. In Japan, only two reports have been based on a time-series design, and the results suggested no relationship between MMR and ASD [29,30]. The most prominent articles in the past have focused mainly on the results of ecologic studies, and we will discuss the few existing case–control studies [31–33]. Each study demonstrated no differences between ASD cases and controls, failing to support a conclusion that immunization using MMR increases the risk of ASD onset.

Worldwide, reports on studies of immunization with vaccines other than MMR are rare. Moreover, parents or legal guardians remain apprehensive about the perceived risk of ASD posed by vaccination [34–37]. Therefore, the purpose of this study was to investigate Japanese subjects, a genetically homogeneous population, regarding links between ASD and immunization with various vaccines, including MMR, as well as the association between ASD and the number of vaccine injections [38]. This is the first case–control study in Asia investigating links between vaccination and ASD onset. These links were examined in ASD cases and controls matched for sex and year of birth based on data found in the Maternal and Child Health (MCH) handbook. This handbook, provided to all mothers by the relevant Japanese health system institution, is a highly reliable record of early development, health, and immunization, and health professionals (e.g. public health nurses, obstetricians, and pediatricians) keep record of most of the data listed in it [39,40]. In this study, therefore, data from the MCH handbook in terms of vaccination history, as well as potential prenatal, perinatal, and neonatal risk factors, were examined.

2. Patients and methods

2.1. Study population

2.1.1. Cases (Fig. 1)

The study analyzed case data from patients of the Yokohama Psycho-Developmental Clinic (YPDC), Kanto area, Japan, which accepts only patients with suspected developmental disorders. Of the patients who initially consulted the YPDC from April 1997 (opening of the clinic) until March 2011, the cases consisted of patients who: (1) were diagnosed with ASD, and (2) had been born between April 1, 1984 and April 30, 1992, the possible time period for MMR vaccination. Subjects whose records in the MCH handbook were missing or illegible and those with a history of vaccination in another country were excluded.

2.1.1.1. Diagnosis of ASD. Patients were diagnosed based on the classifications of pervasive developmental disorders in the Diagnostic and Statistical Manual 4th edition (DSM-IV) and standardized criteria using the Diagnostic Interview for Social and Communication Disorder (DISCO) [41,42]. The DISCO is recognized as one of the best ways to obtain a reliable and valid diagnosis of ASD [43].

One of several child psychiatrists on the team met the patient's parents and used the DISCO to take the patient's developmental history. Another child psychiatrist or clinical psychologist conducted intellectual or developmental tests, such as the Psycho-Educational Profile-Revised and Wechsler Intelligence Scale for Children-Third Edition. After the interview and testing, the diagnosis was made by the team according to the DSM-IV criteria.

2.1.1.2. Period of birth. MMR vaccination in Japan was conducted under specific circumstances. It was introduced in April 1989, and only one vaccination using MMR was included in the immunization schedule. The monovalent mumps and rubella vaccines remained the optimal choice of vaccine for those who did not participate in the MMR program. However, soon after the immunization program commenced, there were several cases of aseptic meningitis, which may have been caused by the mumps vaccine [44]. As a result, in April 1993, the Japanese government ceased extensive inoculation with MMR. Therefore, children born from April 1984 to April 1992 could receive the MMR vaccination, and those children were included in the present study.

2.1.2. Controls (Fig. 1)

One to two controls were selected for each case, matched by sex and year of birth and recruited as volunteers from general schools in the Kanto area, the same area where YPDC patients reside. Consent for participation in the present study was obtained from the parents (or legal guardians) of the students. Students who had previously been recognized as having developmental problems and were already receiving care were excluded, as were those whose records in the MCH handbook were missing or illegible and those with a history of vaccination in another country.

2.2. Source of data

The vaccination history and potential prenatal, perinatal, and neonatal risk factors collected based on the MCH handbook, which was routinely attached to each patient's file, were examined. The targeted vaccines were the MMR, generally used for infants, and the individual vaccines of the same type: the diphtheria–pertussis–tetanus vaccine (DPT); the polio vaccine; the B-encephalitis vaccine; and the Bacillus of Calmette and Guerin vaccine (BCG). For DPT, Polio, and B-encephalitis, there were many subjects who received these vaccines more than once. Therefore, the times of exposure to these vaccines were counted within the period of the first three years, when ASD features first appeared. Maternal hypertension (systolic blood pressure $\geq$140 mmHg or diastolic blood pressure $\geq$90 mmHg), albuminuria or edema, and anemia were examined as prenatal factors. The birth weight, head and chest circumference, duration of labor, delivery method (normal delivery, cesarean section, and obstetrical vacuum extraction or forceps delivery), and Apgar score were examined as perinatal and neonatal factors. Hypertension, albuminuria or edema, and anemia were recorded using a two-category scale (yes/no), and the Apgar score was recorded as an ordinal variable. Duration of labor, birth weight, and head and chest circumference were handled as continuous variables. The delivery method was recorded using a two-category scale (performed/not performed) for each delivery technique.

2.3. Selection of case children and matched control children

Among the patients who initially consulted the clinic between April 1997 and March 2011, 1875 cases of ASD were identified. Of these, 89 cases were excluded because the MCH handbook was missing or the vaccination record in the handbook could not be read, and 3 were excluded because they had received MMR vaccination overseas. Of the remaining 1783 cases, 1429 were born before March 1984 or after May 1992, leaving 354 cases (males: n = 286, 80.8%) born between April 1984 and April 1992, the possible time period for MMR vaccination. The ASD group consisted of 280 subjects with Autistic disorder (79.1%), 27 subjects with Asperger disorder (7.6%), and 47 subjects with Pervasive developmental disorder not otherwise specified (13.3%).

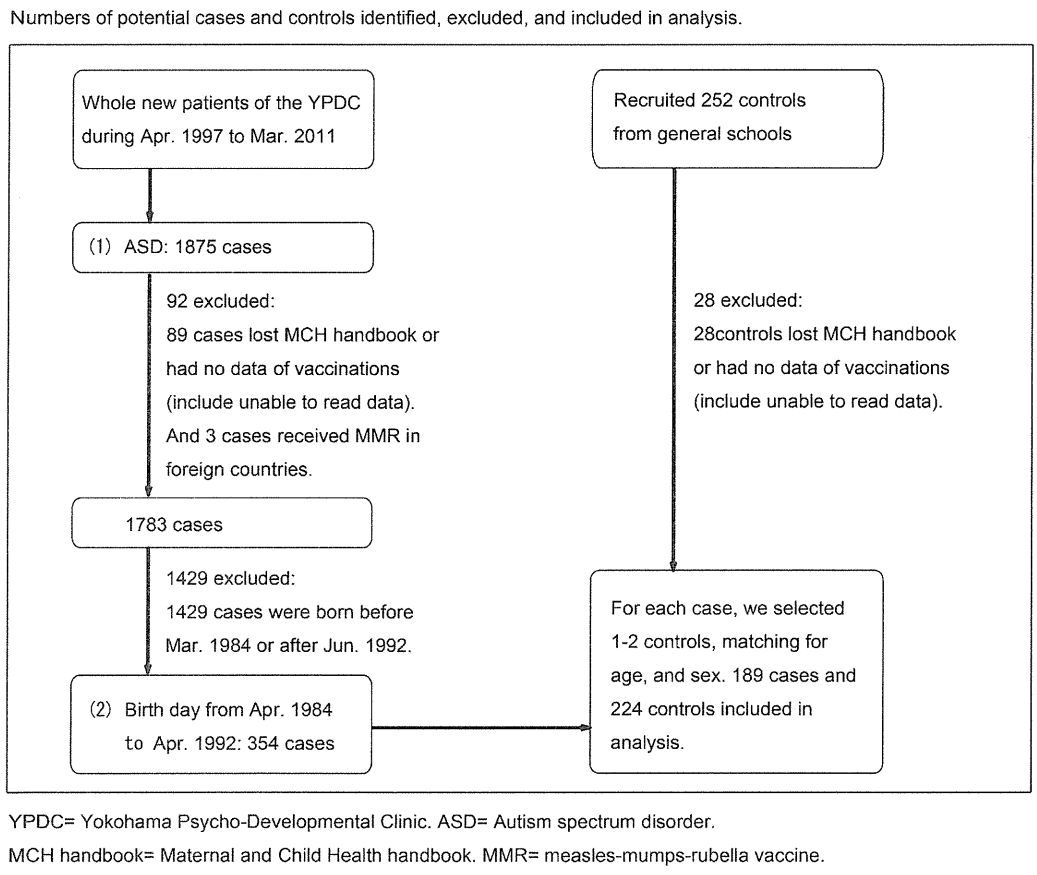

Fig. 1. Numbers of potential cases and controls identified, excluded, and included in analysis. YPDC, Yokohama Psycho-Developmental Clinic; ASD, autism spectrum disorder; MCH handbook, maternal and child health handbook; MMR, measles–mumps–rubella vaccine.

As controls, 252 subjects from the general school population were recruited into the present study. Of these, 28 cases were subsequently excluded because the MCH handbook was missing or the vaccination record could not be read. The goal was to have a matched control for each case. However, since there were not enough controls to match to all cases, 189 subjects were chosen randomly from the ASD group as a case group. The controls were individually matched to cases by age and sex. There were 189 cases, mean age 22.6 years (SD 2.2), and 224 controls, mean age 22.6 years (SD 2.2), with case-to-control ratios ranging from 1:1 to 1:2 (Fig. 1).

2.4. Statistical analysis

2.4.1. Analysis 1

Duration of labor was divided into 2 categories of normal (≤20 h) versus prolonged labor (>20 h). Because an Apgar score of less than 7 has been associated with increased ASD risk [45–48], the Apgar score was divided into 2 categories of normal (≥7 points) and low (<7 points). In order to compare the backgrounds of the cases and controls, the crude odds ratios (ORs) and 95% confidence intervals (CIs) were determined for each outcome. The relationship between ASD onset and the total number of vaccine injections was also investigated. The crude ORs and 95% CI were determined for each. The differences in the mean values of the quantitative variables between cases and controls were examined by an unpaired *t*-test. When necessary, the *t*-test was modified for unequal variances.

2.4.2. Analysis 2

Because this study was only concerned with the theoretical increase in the risk of ASD onset due to the MMR vaccine injection, a conditional logistic model was applied to evaluate the ORs of MMR vaccination after adjusting for other risk factors.

2.4.3. Analysis 3

The OR of the total number of vaccine injections after adjusting for other risk factors was evaluated with the conditional logistic model.

2.4.4. Power analysis

Power analysis was performed in accordance to general power calculation model for chi squared statistics, *t*-test, and a conditional multiple regression model. In brief, power is determined with respect to degree of freedom and predefined alpha level of the study (0.05), number of predictors (in case of a conditional multiple regression model, 4) after assuming effect size (in accordance with Cohen's criteria).

Analysis 1 was performed using SPSS 17.0 for Japan, and Analyses 2 and 3 used the HALBAU 7. ORs were considered significant when the lower 95% CI exceeded 1.0. The *t*-tests were two-sided, and significance was defined as $p < 0.05$. For power calculation G*power v3.1 was used.

2.5. Ethical considerations

This study was approved by the ethics committee at Nagoya University. All data used in this study were clinical data obtained in the course of conventional diagnosis and therapy, and cooperation in the study placed no burden on individual patients. The parents or legal guardians of all of the children in the control group provided their written, informed consent to participate. Personal information regarding subjects in this study and the resulting data

Table 1
The proportions and crude odds ratios (ORs) and 95% confidence intervals (CI) for ASD according to vaccines, prenatal factors, perinatal factors and neonatal factors.

Variable category	*n* (%)		ORs	95% CI	*p*-Value
	Cases ($n = 189$)	Controls ($n = 224$)			
Vaccines					
MMR	47 (24.9)	54 (24.1)	1.04	0.65–1.68	.86
Measles	126 (66.7)	141 (62.9)	1.18	0.77–1.80	.43
Mumps	110 (58.2)	110 (49.1)	1.44	0.96–2.17	.06
Rubella	108 (57.1)	120 (53.6)	1.16	0.77–1.74	.47
DPT	185 (97.9)	219 (97.8)	1.06	0.24–4.75	.94
Polio	184 (97.4)	221 (98.7)	0.5	0.09–2.43	.73
B-encephalitis	167 (88.4)	206 (92.0)	0.66	0.33–1.34	.22
BCG	182 (96.3)	218 (97.3)	0.72	0.21–2.42	.55
Prenatal factors					
Maternal hypertension	6 (3.2)	3 (1.3)	2.42	0.53–12.36	.20
Albuminuria, edema	18 (9.5)	19 (8.5)	1.13	0.55–2.35	.71
Anemia	59 (31.2)	69 (30.8)	1.02	0.66–1.58	.93
Perinatal and neonatal factors					
Prolonged labor (>20 h)[a]	8/169 (4.7)	9/200 (4.5)	1.06	0.36–3.06	.92
Method of delivery					
Cesarean section	21 (11.1)	24 (10.7)	1.04	0.54–2.02	.90
Obstetrical vacuum extraction or forceps delivery[a]	23/168 (13.7)	22/200 (11.0)	1.28	0.66–2.51	.43
Low Apgar score (<7)	6 (3.2)	2 (0.9)	3.64	0.66–26.39	.09

ORs, odds ratios; CI, confidence intervals; ASD, autism spectrum disorder; MMR, measles–mumps–rubella vaccines; DPT, diphtheria–pertussis–tetanus vaccines; BCG, Bacillus of Calmette and Guerin vaccine.

[a] There were 20 cases and 24 controls who did a cesarean section, and 1 case who did a cesarean section because of prolonged labor. Thus, They were excluded from population of prolonged labor and obstetrical vacuum extraction or forceps delivery.

were rendered anonymous, and analyses were performed using only quantitative data that could not be linked to any particular subject.

3. Results

3.1. Vaccination rate and time of exposure to vaccines

The vaccination rates in cases and controls were as follows: MMR, 24.9% of cases and 24.1% of controls; Measles, 66.7% and 62.9%; Mumps, 58.2% and 49.1%; Rubella, 57.1% and 53.6%; DPT, 97.9% and 97.8%; Polio, 97.4% and 98.7%; B-encephalitis, 88.4% and 92.0%, and BCG 96.3% and 97.3% (Table 1). The mean times of each vaccine injection in cases and controls were as follows: DPT, 3.8 times of cases and 3.7 times of controls; Polio, 1.9 times and 2.0 times; B-encephalitis, 1.7 times and 1.8 times (Table 2).

3.2. Analysis

3.2.1. Analysis 1

For each vaccination, the ORs of cases versus controls were as follows (no significant differences were found): MMR, 1.04 (95% CI, 0.65–1.68); Measles, 1.18 (95% CI, 0.77–1.80); Mumps, 1.44 (95% CI, 0.96–2.17); Rubella, 1.16 (95% CI, 0.77–1.74); DPT, 1.06 (95% CI, 0.24–4.75); Polio, 0.50 (95% CI, 0.09–2.43); B-encephalitis, 0.66 (95% CI, 0.33–1.34); and BCG, 0.72 (95% CI, 0.21–2.42). Maternal hypertension as a prenatal factor had an OR of 2.42 (95% CI, 0.53–12.36), but no significant difference was found between cases and controls. For the other factors as well, cases did not have significantly higher ORs than controls. As a perinatal and neonatal factor, low Apgar score and obstetrical vacuum extraction or forceps delivery had an OR of 3.64 (95% CI, 0.66–26.39) and 1.28 (95% CI, 0.66–2.51), respectively, but no significant difference was found between cases and controls. No other perinatal and neonatal factors showed significant differences between cases and controls (Table 1).

A *t*-test was performed on the mean values of the times of exposure to DPT, Polio, and B-encephalitis, birth weight and head and chest circumference between cases and controls, and no significant differences were found ($p > 0.05$ for all). The minimum number of vaccine injections was 3, and the maximum was 13. The mean (standard deviation) number of vaccine injections of cases and controls was 11.4 (1.7) and 11.4 (1.7), respectively, and there was no significant difference between cases and controls ($t = 0.07$, $p = 0.94$) (Tables 2 and 3).

3.2.2. Analysis 2

Maternal hypertension, low Apgar score, and obstetrical vacuum extraction or forceps delivery, which had higher ORs in the results of Analysis 1, were investigated as confounding factors using a conditional multiple regression model. With regard to the presence of ASD, MMR had an OR of 1.10 (95% CI, 0.64–1.90), and maternal hypertension, low Apgar score, and obstetrical vacuum extraction or forceps delivery had ORs of 4.19 (95% CI, 0.46–38.57), 2.06 (95% CI, 0.18–22.12) and 0.98 (95% CI, 0.50–1.92), respectively. There were no significant differences (Table 4).

Table 2
The comparison of the times of vaccine injection between cases and controls.

Vaccines	Cases Mean (±SD)	Controls Mean (±SD)	*p*-Value
DPT	3.8 (±0.8)	3.7 (±0.7)	.78[a]
Polio	1.9 (±0.3)	2.0 (±0.3)	.34[b]
B-encephalitis	1.7 (±0.7)	1.8 (±0.6)	.06[b]

DPT, diphtheria–pertussis–tetanus vaccines.

[a] Student's *t*-test.

[b] Welch's *t*-test.

Table 3
The comparison of quantitative variables between cases and controls.

Variables	Cases Mean (±SD)	Controls Mean (±SD)	*p*-Value
Birth weight (g)	3085.7 (±454.1)	3109.4 (±479.0)	.62[a]
Head circumference (cm)	33.5 (±2.3)	33.6 (±3.0)	.88[b]
Chest circumference (cm)	32.3 (±2.2)	32.3 (±2.7)	.90[a]
The number of vaccine injections (shots)	11.4 (±1.7)	11.4 (±1.7)	.94[a]

[a] Student's *t*-test.

[b] Welch's *t*-test.

Table 4
Odds ratios and 95% confidence intervals of MMR vaccination injection analyzed with a conditional logistic model.

Factor		ORs (95% CI)	*p*-Value
MMR vaccination injection	(−)	1	
	(+)	1.10 (0.64–1.90)	.72
Maternal hypertension	(−)	1	
	(+)	4.19 (0.46–38.57)	.21
Low Apgar score	(−)	1	
	(+)	2.06 (0.18–22.12)	.57
Obstetrical vacuum extraction or forceps delivery	(−)	1	
	(+)	0.98 (0.50–1.92)	.96

ASD, autism spectrum disorder; MMR, measles–mumps–rubella vaccines; ORs, odds ratios; 95% CI, 95% confidence intervals.

3.2.3. Analysis 3

The number of vaccine injections had an OR of 1.10 (95% CI, 0.95–1.26) in a conditional multiple regression model using the same confounding factors as for Analysis 2, maternal hypertension (OR = 3.63, 95% CI, 0.40–33.19), low Apgar score (OR = 2.14, 95% CI, 0.19–23.78), and obstetrical vacuum extraction or forceps delivery (OR = 1.02, 95% CI, 0.52–1.99), and there was no significant difference between cases and controls (Table 5).

3.2.4. Power analysis

Regarding power analysis for chi square statistics and *t*-test, when effect size is set to medium (in accordance to Cohen's criteria), both samples that are characterized in our research had more than 80% power for detecting association, respectively. However, in case, size effect is set to small, calculated power were 52% at chi square statistics and 53% at *t*-test. Similarly, regarding a conditional multiple regression model, our sample had more than 80% of power for detecting association in case of medium effect size. However in case, size effect is set to small, calculated power was 56%.

4. Discussion

The three previous case–control studies focused on the relationship between ASD and MMR. Specifically, the investigation of DeStefano et al. was based on the Metropolitan Atlanta Developmental Disabilities Surveillance Program [31]; Smeeth et al. used data from the UK General Practice Research Database [32]; and DeWilde et al. examined the association using the UK Doctors' Independent Network Database [33]. The aforementioned studies provided no epidemiological evidence for a causal association. The present study is the first case–control study in Asia investigating the relationship between a variety of vaccines including MMR and the risk of ASD onset.

Table 5
Odds ratios of one measure and 95% confidence intervals of the number of vaccine injections analyzed with a conditional logistic model.

Factor		ORs (95% CI)	*p*-Value
The number of vaccine injections	(−)	1	
	(+)	1.10[a] (0.95–1.26)	.19
Maternal hypertension	(−)	1	
	(+)	3.63 (0.40–33.19)	.25
Low Apgar score	(−)	1	
	(+)	2.14 (0.19–23.78)	.54
Obstetrical vacuum extraction or forceps delivery	(−)	1	
	(+)	1.02 (0.52–1.99)	.96

ASD, autism spectrum disorder; ORs, odds ratios; 95% CI, 95% confidence intervals.
[a] OR of the number of vaccine injections means OR of increasing one injection of vaccine.

These previous studies were conducted using relatively heterogeneous samples in terms of genetic makeup. Conversely, the Japanese population is thought to be highly homogenous on the genetic level (which gives us the opportunity to minimize the effect of population-specific risk factors that might interact with environmental exposures (i.e. immunization)), and almost all Japanese parents have an MCH handbook. The fact that highly reliable information concerning the pregnancy, perinatal, and neonatal periods is collected in the handbook was advantageous for conducting this research.

In this study, we could not find the evidence that MMR vaccination increases the risk of ASD onset. The present results support the findings from the previous case–control studies conducted in Caucasian populations. Furthermore, we could not find any evidences that other types of vaccines or a combined effect of multiple vaccines was associated with ASD onset. Therefore, this study did not support the theory that vaccinations should be avoided to reduce the risk of ASD onset. We should be more concerned about acquiring infectious diseases by avoiding vaccinations.

In the results of this study, the 95% CIs of vaccinations, especially DPT, Polio, and BCG had a wide range because of small power. The sample size was not large enough to absolutely exclude the possibility that DPT, Polio, and BCG vaccinations increased the risk of ASD onset. Additionally, there were no theories about an increase in the risk of ASD onset concerns with any single types of vaccine injection that were included in this study, other than the MMR vaccine. Then a conditional logistic model was applied not to DPT, Polio, and BCG, but to MMR which was more concerned with the risk of ASD onset. This study was limited to data from the MCH handbook, which from the viewpoint of conducting an investigation is a highly reliable vaccination data source. On one hand we believe we have obtained very reliable results. However, on the other hand, the information in this handbook does not include several factors which were known to increase the risk of ASD onset, such as parental age at birth, bleeding, birth order, previous fetal loss, maternal prenatal medication use exclusively for hypertension and maternal toxaemia which were included in this study [49], and coexisting conditions that may influence vaccinations received, for example cardiovascular disease, other physical diseases or anomaly, epilepsy, or allergy. We were not able to investigate such conditions in controls because of the nature of the data collection procedure, which involved community-based sampling. Moreover, the relationship is unclear between the time periods when ASD was diagnosed and when the child was vaccinated. It has been hypothesized that early exposure to thimerosal and immune globulin preparations influence neuropsychological deficits in children include ASD [16,50]. Additionally, it is possible that parents lost motivation regarding vaccination before ASD was diagnosed because of problems such as the child's inability to sit still or frequent tantrums. Even so, both groups showed a high vaccination rate for each of the vaccines, and because the main topic of this study was to investigate whether vaccination increases the risk of ASD onset, we believe these effects can be ignored.

In future studies on vaccination and ASD, investigations with larger sample sizes are expected, and we anticipate examining factors which were known to increase the risk of ASD onset, coexisting conditions that may influence vaccination, such as cardiovascular disease, other physical diseases or anomaly, epilepsy, or allergies, the age at ASD diagnosis and vaccines injection, and reasons why vaccinations were not performed. We also look forward to

prospective studies that include pregnancy, delivery, or even preconception factors that may be associated with ASD.

In this study, we could not find any convincing evidence that MMR vaccination and increasing the number of vaccine injections were associated with an increased risk of ASD in a genetically homogeneous population. If such an association exists, it is so rare that it could not be identified in this large regional sample. Therefore, our findings indicate there is no basis for avoiding vaccination out of concern for ASD. This study investigated the link between vaccination and the risk of ASD, but it does not guarantee the safety or efficacy of the vaccines. Adverse reactions from vaccines other than a link with ASD exist. Such adverse reactions must be studied, and safer and more effective vaccines must be developed. At one time in Japan, mumps vaccine in the MMR vaccine caused several cases of aseptic meningitis. We should continue to investigate the safety and efficacy of vaccines carefully and the biological features of ASD in greater depth to improve outcomes related to long-term function and quality of life.

Acknowledgments

A part of this study is the result of research grants from the Ministry of Health, Labor and Welfare of Japan, and "Integrated research on neuropsychiatric disorders" carried out under the Strategic Research Program for Brain Sciences by the Ministry of Education, Culture, Sports, Science and Technology of Japan. The authors would like to warmly thank all the students who took part in this research and their parents and teachers. They are grateful to Professor Nobuyuki Hamajima and Professor Yutaka Inaba for advice and for checking the statistical analyses. Special thanks are expressed to Toru Yoshikawa, Hiroshi Fujioka, Yuko Yoshida, and Yuriko Hachiya for their assistance.

References

[1] Lotter V. Childhood autism in Africa. J Child Psychol Psychiatry 1978;19:231–44.
[2] Wing L. Autistic spectrum disorders. BMJ 1996;312:327–8.
[3] Wing L. The autistic spectrum. Lancet 1997;350:1761–6.
[4] Baird G, Simonoff E, Pickles A, Chandler S, Loucas T, Meldrum D, et al. Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). Lancet 2006;368:210–5.
[5] Brugha TS, McManus S, Bankart J, Scott F, Purdon S, Smith J, et al. Epidemiology of autism spectrum disorders in adults in the community in England. Arch Gen Psychiatry 2011;68:459–65.
[6] Kim YS, Leventhal BL, Koh YJ, Fombonne E, Laska E, Lim EC, et al. Prevalence of autism spectrum disorders in a total population sample. Am J Psychiatry 2011.
[7] Levy SE, Mandell DS, Schultz RT. Autism. Lancet 2009;374:1627–38.
[8] Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 2011;68:1095–102.
[9] Lichtenstein P, Carlstrom E, Rastam M, Gillberg C, Anckarsater H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry 2010;167:1357–63.
[10] Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 1998;351:637–41.
[11] Lancet TEOT. Retraction-Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 2010;375:445.
[12] Kawashima H, Mori T, Kashiwagi Y, Takekuma K, Hoshika A, Wakefield A. Detection and sequencing of measles virus from peripheral mononuclear cells from patients with inflammatory bowel disease and autism. Dig Dis Sci 2000;45:723–9.
[13] Uhlmann V, Martin CM, Sheils O, Pilkington L, Silva I, Killalea A, et al. Potential viral pathogenic mechanism for new variant inflammatory bowel disease. Mol Pathol 2002;55:84–90.
[14] Delong G. A positive association found between autism prevalence and childhood vaccination uptake across the U.S. population. J Toxicol Environ Health A 2011;74:903–16.
[15] Joint statement of the American Academy of Pediatrics (AAP) and the United States Public Health Service (USPHS). Pediatrics 1999;104:568–9.
[16] Bernard S, Enayati A, Redwood L, Roger H, Binstock T. Autism: a novel form of mercury poisoning. Med Hypotheses 2001;56:462–71.
[17] Offit PA. Vaccines and autism revisited—the Hannah Poling case. N Engl J Med 2008;358:2089–91.
[18] Farrington CP, Miller E, Taylor B. MMR and autism: further evidence against a causal association. Vaccine 2001;19:3632–5.
[19] del Kaye JA, Mar Melero-Montes M, Jick H. Mumps, measles, and rubella vaccine and the incidence of autism recorded by general practitioners: a time trend analysis. BMJ 2001;322:460–3.
[20] Dales L, Hammer SJ, Smith NJ. Time trends in autism and in MMR immunization coverage in California. JAMA 2001;285:1183–5.
[21] Fombonne E, Chakrabarti S. No evidence for a new variant of measles–mumps–rubella-induced autism. Pediatrics 2001;108:E58.
[22] Makela A, Nuorti JP, Peltola H. Neurologic disorders after measles–mumps–rubella vaccination. Pediatrics 2002;110:957–63.
[23] Taylor B, Lingam R, Simmons A, Stowe J, Miller E, Andrews N. Autism and MMR vaccination in North London; no causal relationship. Mol Psychiatry 2002;7(Suppl. 2):S7–8.
[24] Madsen KM, Hviid A, Vestergaard M, Schendel D, Wohlfahrt J, Thorsen P, et al. A population based study of MMR vaccination and autism. New Engl J 2002;347:1477–82.
[25] Chen W, Landau S, Sham P, Fombonne E. No evidence for links between autism, MMR and measles virus. Psychol Med 2004;34:543–53.
[26] Richler J, Luyster R, Risi S, Hsu WL, Dawson G, Bernier R, et al. Is there a 'regressive phenotype' of autism spectrum disorder associated with the measles–mumps–rubella vaccine? A CPEA Study. J Autism Dev Disord 2006;36:299–316.
[27] Taylor B, Miller E, Farrington CP, Petropoulos MC, Favot-Mayaud I, Li J, et al. Autism and measles, mumps, and rubella vaccine: no epidemiological evidence for a causal association. Lancet 1999;353:2026–9.
[28] Fombonne E, Zakarian R, Bennett A, Meng L, McLean-Heywood D. Pervasive developmental disorders in Montreal, Quebec, Canada: prevalence and links with immunizations. Pediatrics 2006;118:e139–50.
[29] Honda H, Shimizu Y, Rutter M. No effect of MMR withdrawal on the incidence of autism: a total population study. J Child Psychol Psychiatry 2005;46:572–9.
[30] Uchiyama T, Kurosawa M, Inaba Y. MMR-vaccine and regression in autism spectrum disorders: negative results presented from Japan. J Autism Dev Disord 2007;37:210–7.
[31] DeStefano F, Bhasin TK, Thompson WW, Yeargin-Allsopp M, Boyle C. Age at first measles–mumps–rubella vaccination in children with autism and school-matched control subjects: a population-based study in metropolitan Atlanta. Pediatrics 2004;113:259–66.
[32] Smeeth L, Cook C, Fombonne E, Heavey L, Rodrigues LC, Smith PG, et al. MMR vaccination and pervasive developmental disorders: a case–control study. Lancet 2004;364:963–9.
[33] DeWilde S, Carey IM, Richards N, Hilton SR, Cook DG. Do children who become autistic consult more often after MMR vaccination. Br J Gen Pract 2001;51:226–7.
[34] Stephen D, Sugarman SD. Cases in vaccine court—legal battles over vaccines and autism. N Engl J Med 2007;357:1275–7.
[35] Casiday R, Cresswell T, Wilson D, Panter-Brick C. A survey of UK parental attitudes to the MMR vaccine and trust in medical authority. Vaccine 2006;24:177–84.
[36] Cassell JA, Leach M, Poltorak MS, Mercer CH, Iversen A, Fairhead JR. Is the cultural context of MMR rejection a key to an effective public health discourse. Public Health 2006;120:783–94.
[37] Hilton S, Hunt K, Petticrew M. MMR: marginalised, misrepresented and rejected? Autism: a focus group study. Arch Dis Child 2007;92:322–7.
[38] Haga H, Yamada R, Ohnishi Y, Nakamura Y, Tanaka T. Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190,562 genetic variations in the human genome. Single-nucleotide polymorphism. J Hum Genet 2002;47:605–10.
[39] Kunugi H, Nanko S, Takei N, Saito K, Murray RM, Hirose T. Perinatal complications and schizophrenia. Data from the Maternal and Child Health Handbook in Japan. J Nerv Ment Dis 1996;184:542–6.
[40] Ohara K, Tanabu S, Yoshida K, Sato Y, Shibuya H. Obstetric complications in siblings of Japanese schizophrenics: data from the Maternal and Child Health Handbook. Prog Neuropsychopharmacol Biol Psychiatry 2005;29:617–20.
[41] Wing L, Leekam SR, Libby SJ, Gould J, Larcombe M. The diagnostic interview for social and communication disorders: background, inter-rater reliability and clinical use. J Child Psychol Psychiatry 2002;43:307–25.
[42] Leekam SR, Libby SJ, Wing L, Gould J, Taylor C. The diagnostic interview for social and communication disorders: algorithms for ICD-10 childhood autism and Wing and Gould autistic spectrum disorder. J Child Psychol Psychiatry 2002;43:327–42.
[43] Nygren G, Hagberg B, Billstedt E, Skoglund A, Gillberg C, Johansson M. The Swedish version of the Diagnostic Interview for Social and Communication Disorders (DISCO-10). Psychometric properties. J Autism Dev Disord 2009;39:730–41.
[44] Sugiura A, Yamada A. Aseptic meningitis as a complication of mumps vaccination. Pediatr Infect Dis J 1991;10:209–13.
[45] Eaton WW, Mortensen PB, Thomsen PH, Frydenberg M. Obstetric complications and risk for severe psychopathology in childhood. J Autism Dev Disord 2001;31:279–85.
[46] Hultman CM, Sparen P, Cnattingius S. Perinatal risk factors for infantile autism. Epidemiology 2002;13:417–23.

[47] Glasson EJ, Bower C, Petterson B, de Klerk N, Chaney G, Hallmayer JF. Perinatal factors and the development of autism: a population study. Arch Gen Psychiatry 2004;61:618–27.
[48] Larsson HJ, Eaton WW, Madsen KM, Vestergaard M, Olesen AV, Agerbo E, et al. Risk factors for autism: perinatal factors, parental psychiatric history, and socioeconomic status. Am J Epidemiol 2005;161:916–25, discussion 26–8.
[49] Gardener H, Spiegelman D, Buka SL. Prenatal risk factors for autism: comprehensive meta-analysis. Br J Psychiatry 2009;195:7–14.
[50] Thompson WW, Price C, Goodson B, Shay DK, Benson P, Hinrichsen VL, et al. Early thimerosal exposure and neuropsychological outcomes at 7 to 10 years. N Engl J Med 2007;357:1281–92.

Research in Autism Spectrum Disorders 6 (2012) 1265–1272

Contents lists available at SciVerse ScienceDirect

Research in Autism Spectrum Disorders

Journal homepage: http://ees.elsevier.com/RASD/default.asp

Validation of an interview-based rating scale developed in Japan for pervasive developmental disorders

Hiroyuki Ito [a,*], Iori Tani [b], Ryoji Yukihiro [c], Jun Adachi [d], Koichi Hara [e], Megumi Ogasawara [f], Masahiko Inoue [g], Yoko Kamio [h], Kazuhiko Nakamura [i], Tokio Uchiyama [j], Hironobu Ichikawa [k], Toshiro Sugiyama [l], Taku Hagiwara [d], Masatsugu Tsujii [m]

[a] *Research Center for Child Mental Development, Hamamatsu University School of Medicine, Japan*
[b] *Faculty of Humanities, Tokaigakuen University, Japan*
[c] *Faculty of Human and Cultural Studies, Kyoto Gakuen University, Japan*
[d] *Hokkaido University of Education, Japan*
[e] *Faculty of Integrated Arts and Sciences, University of Tokushima, Japan*
[f] *School of Education, Tokyo Gakugei University, Japan*
[g] *Graduate School of Medical Sciences, Tottori University, Japan*
[h] *Department of Child and Adolescent Mental Health, National Center of Neurology and Psychiatry, Japan*
[i] *Department of Psychiatry and Neurology, Hamamatsu University School of Medicine, Japan*
[j] *Faculty of Human Development and Culture, Fukushima University, Japan*
[k] *Tokyo Metropolitan Children's Medical Center, Japan*
[l] *Department of Child and Adolescent Psychiatry, Hamamatsu University School of Medicine, Japan*
[m] *School of Contemporary Sociology, Chukyo University, Japan*

ARTICLE INFO

Article history:
Received 27 July 2011
Received in revised form 1 April 2012
Accepted 2 April 2012

Keywords:
PDD
Autism spectrum disorders
ADI-R
Psychometrics

ABSTRACT

The pervasive developmental disorders (PDDs) Autism Society Japan Rating Scale (PARS), an interview-based instrument for evaluating PDDs, has been developed in Japan with the aim of providing a method that (1) can be used to evaluate PDD symptoms and related support needs and (2) is simpler and easier than the currently used "gold standard" instruments such as the Autism Diagnostic Interview-Revised (ADI-R). We examined the reliability and validity of PARS on the basis of data from 572 participants (277 PDD patients and 295 nonclinical controls). Inter-rater reliability was sufficient at both the item and scale level. Factor analysis extracted four subscales, for which internal consistency was found to be high. The sub and total scores of PARS showed correlations with the domain and total scores of ADI-R, in line with theoretical prediction, indicating the convergent validity of PARS. A receiver operating characteristic analysis showed that PARS has good discriminative validity in differentiating between PDD patients and nonclinical controls, regardless of intellectual capacity. Considering that PARS can be easily implemented by professionals with appropriate knowledge regarding PDDs, PARS may be superior to the existing instruments in terms of cost performance.

1. Introduction

Over the course of many years, several instruments have been developed for the diagnosis, evaluation, and screening of pervasive development disorders (PDD). In recent years, the Autism Diagnostic Interview-Revised (ADI-R; Le Couteur et al., 1989; Lord, Rutter, & Le Couteur, 1994) has been broadly accepted as a standardized interview-based diagnostic instrument

* Corresponding author at: Research Center for Child Mental Development, Hamamatsu University School of Medicine, 1-20-1, Handayama, Higashi-ku, Hamamatsu 431-3192, Japan. Tel.: +81 53 435 2331; fax: +81 53 435 2291.
E-mail addresses: ito_hiroyuki@pd5.so-net.ne.jp, ito_h@hama-med.ac.jp (H. Ito).

1750-9467/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.rasd.2012.04.002

for PDD. The Autism Diagnostic Observational Schedule (ADOS; Lord et al., 2000, 1989) is also widely used as an observation-based diagnostic instrument. These instruments have a high level of discriminative validity with respect to the differentiation of PDD from non-PDD and are useful in reaching a definitive diagnosis; however, their implementation requires special training and significant time, leading to the development of numerous simpler evaluation scales in recent years.

The Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001), which has been broadly accepted as a screening instrument, is a unique tool that comprises a combination of questionnaires, telephone interviews, and structured follow-up interviews. Although it is a highly useful tool, its use is limited to toddlers because it was developed with the aim of early identification of PDD. In countries such as Japan and other Asian countries lacking the medical and governmental services for PDD that exist in the United States and Europe, it is believed that many people with undiagnosed PDD exist in a broad age group. In fact, Kawamura, Takahashi, and Ishii (2008) reported that in Toyota City, Japan, where a new systematic PDD screening system has been implemented, there were 11 times more detections of PDD compared with that observed in a survey done 20 years ago. However, few regions in the world have an adequate PDD detection system of this kind. Considering this, the development of a simple and practical evaluation scale that can be applied to a wide age group is an important and pressing issue.

The Autism Spectrum Screening Questionnaire (ASSQ; Ehlers, Gillberg, & Wing, 1999), Autism Screening Questionnaire (ASQ; Berument, Rutter, Lord, Pickles, & Bailey, 1999), and Social Responsiveness Scale (Constantino et al., 2003) have been developed as PDD evaluation scales that can be applied to a relatively broad age group. As all of these evaluation tools are in the format of a questionnaire that can be evaluated by parents or teachers, they have the advantage of being fairly easy to implement. However, in most cases, parents lack the specialized knowledge needed to understand PDD, so the standards for rating individual items can vary greatly depending on the individual conducting the evaluation, possibly leading to a deterioration of the reliability of evaluation results. Furthermore, though teachers generally have more PDD-related knowledge than do parents, they have less specific knowledge of each individual child; hence, their evaluations tend to be less reliable than those of parents. In practice, the sensitivity (true positive rate) and specificity (one minus false positive rate) of the ASSQ in distinguishing PDD and non-PDD was .91 and .77, respectively, for the parent evaluation and .90 and .58, respectively, for the teacher evaluation (Ehlers et al., 1999). Considering that the sensitivity and specificity of the ADI-R were 1.00 and .90, respectively (Lord et al., 1997), the level of accuracy of the ASSQ in distinguishing PDD from non-PDD was insufficient in the hands of both parents and teachers. Furthermore, in a simultaneous comparison conducted by Charman et al. (2007), sensitivity and specificity in identifying autistic spectrum disorders was .86 and .78, respectively, for the ASQ and .78 and .67, respectively, for the SRS, thereby indicating its insufficient precision in practical use.

To resolve this dilemma between accuracy and simplicity, the PDDs Autism Society Japan Rating Scale (PARS) has been developed in Japan as an instrument for evaluating PDDs (Adachi et al., 2006; Kamio et al., 2006; Tsujii et al., 2006). This scale was developed with the aim of providing an instrument that is simpler to use than the ADI-R and ADOS; is applicable to any age group, unlike the M-CHAT; and has better reliability and validity than questionnaire scales such as the ASSQ and ASQ. While PARS uses an interview format similar to ADI-R, the procedures, which are briefly summarized in the manual, can be implemented after simple training. Furthermore, because the criteria for rating each item is clearly defined in PARS, a more reliable and valid evaluation is possible than with questionnaire scales. In order to ease the rating process and shorten the evaluation time, the evaluator assigns values at three levels—none (0 points), somewhat apparent (1 point), and apparent (2 points)—for the 34 items listed as typical behavioral symptoms of PDD. This innovation ensures that the time required to implement PARS is kept to 30–90 min, depending on the interviewer's proficiency and the target's age and symptoms.

There is no international literature on the psychometric properties of PARS, although PARS is now widely used in Japan. This study examined the reliability and validity of PARS and involved a study population of 628 test subjects that included 302 people with PDD and 326 people without PDD. Specifically, we evaluated the inter-rater reliability, factor structure, internal consistency, correlation with the ADI-R, and the ability to distinguish subjects with PDD from a nonclinical sample.

2. Methods

2.1. PARS

The PARS instrument has been developed (Adachi et al., 2006; Kamio et al., 2006; Tsujii et al., 2006) and published (PARS Committee, 2008) in Japan. It involves the evaluation of PDD symptoms through a semi-structured interview conducted with a parent or family member of the subject as the target. This tool can be used to assess not only the risk of PDD but also the need for support pertaining to administrative and medical services. PARS comprises both an evaluation of symptoms when they were most pronounced during infancy (named the peak symptoms scale) and an evaluation of current symptoms (named the current symptoms scale). The former is used mainly to an assessment of PDD risk, and the latter is mainly used in assessment of actual support needs. The peak symptoms scale, which comprises 34 items, is the same for subjects of all age groups, whereas the current symptoms scale, which comprises 57 items, has 3 versions targeting different age groups: preschoolers, primary schoolers, and adolescents/adults. This study reports on data obtained from the peak symptoms scale.

The PARS peak symptoms scale comprises 34 items that describe the characteristic behavioral symptoms of PDDs during the preschooler phase. The items were selected by a panel of eight child psychiatrists and a developmental clinical psychotherapist who were specialized in autism research and clinical practice with more than 10 years of expertise. They compiled behavioral characteristics shown by children with PDD and classified them into eight categories—Interpersonal

Relationship, Communication, Restricted Interests, Stereotyped Behavior, Resistance, Hypersensitivity, Clumsiness, and other complications. From these, 34 items relating to symptoms that are specific to PDD, as well as items relating to nonspecific symptoms with high need for either clinical or administrative support, were selected. Twenty-two out of the 34 items corresponded to diagnostic features for PDD in the *Diagnostic and Statistical Manual 4th Edition, Text Revision* (DSM-IV-TR; American Psychiatric Association, 2000), and 8 corresponded to associated features. Symptoms described in the remaining four items (items 15, 27, 28, and 32) were not listed in the DSM-IV-TR, but since they are often present in PDD children seen in everyday clinical experience, they were included in the scales.

The evaluation of each item in PARS is based on a 30-page manual (PARS Committee, 2008). This manual includes detailed explanations of the questioning and rating standards for each item. For example, for item 1 of the peak symptoms scale (not making eye contact), a sample question "has the child ever had difficulty making eye contact?" is presented, and the rating standards are listed in detail: "0: made eye contact always," "1: had some difficulty making eye contact (made eye contact when requesting or showing interest in something but not otherwise; sometimes made eye contact and sometimes did not; made eye contact only with the parents but not with others)," and "2: rarely made eye contact (did not make eye contact with parents; avoided eye contact)." In this way, evaluation based on subjective criteria of the interviewer is avoided, and a more objective evaluation is possible.

2.2. Sample

The 572 subjects of the main sample comprised two broad groups: a PDD group made up of 277 subjects and a nonclinical control group made up of 295 subjects (Table 1).

Participants in the PDD group were diagnosed as having PDD or subordinate disorders based on the DSM-IV by experienced psychiatrists of medical and educational facilities in 28 areas throughout Japan. The diagnoses were made by integrating data from parental interviews; developmental and medical information; records provided by parents, other caregivers, and teachers; and direct observations of and interactions with the children. Subjects were referred to the facilities due to developmental concerns and randomly recruited for the study by examiners belonging to the facilities. Among these, 175 subjects underwent full-scale IQ tests using intelligence scales such as the Wechsler (Japanese WISC-III Publication Committee, 1998; Shinagawa, Kobayashi, Fujita, & Maekawa, 1990), Binet (Tanaka Institute for Educational Research, 2003), and K-ABC scales (Kaufman, Nadeen, & Kaufman, 1993). Of the 175 subjects, 51 were considered mentally retarded (IQ < 70), while 118 were not (IQ ≥ 70). To evaluate the correlation between PARS and the ADI-R, an ADI-R interview was additionally administered to 74 subjects (mean age = 14.0 years; SD = 3.6; range = 7–24 years; mean IQ = 86.2; SD = 24.7; range = 40–135) from the PDD group.

Table 1
Characteristics of the main sample.

	Age			IQ			Gender		
	M[a]	*SD*[b]	Range	*M*	*SD*	Range	Male	Female	Total
All age groups									
PDD[c] group	12.5	5.8	3–39	81.6	29.2	19–142	233	44	277
Without MR[d] (IQ[e] ≥ 70)	12.7	5.5	4–39	97.2	16.8	70–142	105	13	118
With MR (IQ < 70)	12.3	4.9	5–31	43.6	15.7	18–69	44	13	57
IQ unknown	12.4	6.3	3–32	–	–	–	84	18	102
Nonclinical control group	10.8	7.6	3–38	–	–	–	153	142	295
Preschoolers (age, 3–6 years)									
PDD group	5.1	1.0	3–6	74.1	24.5	22–121	27	12	39
Without MR (IQ ≥ 70)	5.4	0.8	4–6	87.7	13.6	70–121	9	5	14
With MR (IQ < 70)	5.9	0.4	5–6	47.0	17.8	22–68	3	3	6
IQ unknown	4.5	1.0	3–6	–	–	–	15	4	19
Nonclinical control group	4.8	1.0	3–6	–	–	–	69	63	132
Primary schoolers (age, 6–12 years)									
PDD group	9.9	1.8	6–12	80.9	31.9	18–140	94	15	109
Without MR (IQ ≥ 70)	10.2	1.7	7–12	99.6	16.2	71–140	46	5	51
With MR (IQ < 70)	9.2	2.0	6–12	40.5	13.5	18–65	16	5	21
IQ unknown	10.0	1.7	7–12	–	–	–	32	5	37
Nonclinical control group	9.2	1.8	6–12	–	–	–	34	33	67
Adolescents and adults (age, 12–39 years)									
PDD group	17.3	5.2	12–39	77.4	31.2	19–142	112	17	129
Without MR (IQ ≥ 70)	17.1	5.5	12–39	97.9	16.9	70–142	50	3	53
With MR (IQ < 70)	15.9	3.7	12–31	44.9	16.7	19–69	25	5	30
IQ unknown	17.8	5.0	12–32	–	–	–	37	9	46
Nonclinical control group	20.1	6.0	13–38	–	–	–	50	46	96

[a] Mean.
[b] Standard deviation.
[c] Pervasive development disorders.
[d] Mental retardation.
[e] Intelligence quotient.