

# ESAT and M-CHAT as screening instruments for autism spectrum disorders at 18 months in the general population: issues of overlap and association with clinical referrals

Karin T. Beuker · Synnve Schjølberg ·  
Kari Kveim Lie · Sophie Swinkels ·  
Nanda N. J. Rommelse · Jan K. Buitelaar

Received: 11 September 2013 / Accepted: 8 May 2014 / Published online: 28 May 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** The Modified Checklist for Autism in Toddlers (M-CHAT) and the Early Screening of Autistic Traits (ESAT) were designed to screen for autism spectrum disorders in very young children. The aim of this study was to explore proportions of children that screened positive on the ESAT or the M-CHAT and to investigate if screening positive on the ESAT and M-CHAT is associated with clinical referral by 18 months and other aspects of children's development, health, and behavior. In this study, the mothers of 12,948 18-month-old children returned a questionnaire consisting of items from the ESAT and M-CHAT, plus questions about clinical and developmental

characteristics. The M-CHAT identified more screen-positive children than the ESAT, but the ESAT was associated with more clinical referrals and tended to identify more children with medical, language, and behavioral problems. A post hoc analysis of combining the two instruments found this to be more effective than the individual instruments alone in identifying children referred to clinical services at 18 months. Further analysis at the level of single items is warranted to improve these screening instruments.

**Keywords** ASD · M-CHAT · ESAT · Early screening · Development · Infants

**Electronic supplementary material** The online version of this article (doi:10.1007/s00787-014-0561-8) contains supplementary material, which is available to authorized users.

K. T. Beuker (✉) · S. Swinkels · N. N. J. Rommelse  
Department of Psychiatry (966), Nijmegen Centre for Evidence-Based Practice, Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands  
e-mail: Karin.Beuker@radboudumc.nl

K. T. Beuker · S. Swinkels · N. N. J. Rommelse · J. K. Buitelaar  
Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands

S. Schjølberg  
Division of Mental Health, Norwegian Institute of Public Health, Oslo, Norway

K. K. Lie  
Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway

J. K. Buitelaar  
Department of Cognitive Neuroscience, Nijmegen Centre for Evidence-based Practice, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

## Introduction

Commitment to the early detection of Autism Spectrum Disorders (ASD) in children has resulted in more children being identified in the pre-school period [1]. However, there is still a gap of 18 months between parents' first concerns and the definitive diagnosis [2]. This diagnostic delay increases parental distress [3, 4] and delays treatment, which is unfortunate because there is an evidence that early intervention leads to better outcomes. Studies have shown that children with ASD make greater advances when they start treatment before 4 years of age than at an older age [5, 6], and early intervention is associated with language gains, improved social behavior, and decreased symptoms of autism [6].

Despite the importance of identifying and treating ASD early, it remains difficult to detect ASD in young children. Clinical features such as failure to develop peer relationships and lack of varied pretend play emerge over the first years of development and cannot be used as a diagnostic

criterion at an early age, and symptoms of repetitive movements, such as flapping or rocking, may be present in typically developing infants [7]. In addition, the characteristics of autism change over time and can become more obvious and severe, or more subtle and difficult to detect. The need for screening instruments to facilitate the early recognition of autism and related disabilities has long been recognized. The first instrument developed for this purpose was the Checklist for Autism in Toddlers (CHAT) [3]. Based on the Theory of Mind construct and joint attention, the CHAT was designed to screen for autism in children at 18 months of age in the general population. It includes nine parent-report items combined with five observational items that are completed by a home visitor. Researchers in the USA subsequently modified and extended the instrument to a 23-item parent-report questionnaire [8]. The Modified Checklist for Autism in Toddlers (M-CHAT) incorporates the original nine CHAT parent-report questions and additional items, and eliminates the practitioner observation component of the CHAT. Another instrument was recently developed for early screening: the Early Screening of Autistic Traits (ESAT) [9]. This questionnaire consists of 14 yes–no items and was developed as a primary screen to identify children at risk of ASD in the general population at 14 months. The ESAT is an empirically based, bottom-up constructed questionnaire, which means that it is based on the review of prominent early signs and symptoms of ASD as reported in earlier studies. The ESAT is not a top-down instrument like the CHAT key screening items that have been chosen to fit a conceptual model.

Although these screening instruments have been tested and validated in high-risk samples and in unselected samples of children of different ages, their psychometric properties as primary screening instrument for children aged 18 months in the general population have not yet been fully investigated or replicated. Since there is still a gap between parents' first concerns and clinical evaluation, it is of interest to know if a screen-positive result is related to other parental worries, behavioral and medical problems and referrals. Previous studies by Glascoe [10, 11] and Tervo [12], for example, found that parental concerns relate directly to their child's wellbeing and development. Also, parental judgment about whether or not to comply with professional recommendations did reflect a rather accurate estimate of the severity of autistic symptoms of their child [13]. However, there are some complications that add to the difficulties in early identification and intervention of children at risk. For example, health providers may minimize or dismiss parental concerns [14]. To complicate things further, previous studies found that when parental suspicions of problems or symptoms are not systematically elicited, more than 40 % of parents do not reveal them [10]. When the parents

expressed their concerns, this did not increase the likelihood of referral to diagnostic and treatment services from a primary physician [15]. Using a standardized tool or instrument improves an exchange of thoughts for parents as well as health care providers. Therefore, this study focuses on the very important stage before clinical assessment, by addressing parental concerns, behavioral and medical problems and early referral status in relation to screening outcomes at the ESAT and M-CHAT. If the ESAT or M-CHAT strongly differentiates between the degree and content of parental concerns for children being screening positive and children being screen negative, this would contribute importantly to the validity of both instruments in being able to identify children at high risk for autism spectrum disorders or other serious developmental problems at an early age. In addition, there have been no head-to-head comparisons of the two instruments in the same population.

Here, we report results from a large Norwegian prospective pregnancy cohort study that included the ESAT and M-CHAT on the 18-month questionnaire. We investigated: (1) proportions of children that screened positive for ASD on the ESAT or the M-CHAT; (2) the extent to which screening positive on the ESAT and M-CHAT is associated with clinical referral by 18 months; and (3) whether screening positive on the ESAT or M-CHAT is associated with other aspects of children's development, health, and behavior.

## Methods

### Design

This study was part of an ongoing, prospective pregnancy cohort study of more than 100,000 births in Norway, the Norwegian Mother and Child Cohort Study (MoBa) [16–18]. The MoBa cohort was initiated in 1999 and had included 100,000 pregnant women by April 2008. Pregnant women at participating hospitals were recruited to the study through a postal invitation in connection with the routine ultrasound examination offered to all pregnant women in Norway at 17–18 weeks of gestation (<http://www.fhi.no/morogbarn>). Overall, 44 % of those invited agreed to participate; informed consent was obtained from all participants. Information on the health and life style of the pregnant women, their partners, and, subsequently, their children was collected by means of questionnaires during pregnancy and post partum. The current study is based on the quality-assured data files released for research version II. The Regional Committee for Medical Research and the Norwegian Data Inspectorate approved the study. In this study, we used the data collected with the

questionnaire that mothers completed when their child was 18 months.

### Participants

The study population was a subsample of the children from the MoBa study. The two inclusion criteria were that the child became 18 months during the inclusive period 25 July 2003 to 29 March 2005, and that the mother had completed the questionnaire when the child was 18 months. Of the 16,919 children that met the first criterion, 13,015 also met the second criterion (response rate for completion of the questionnaire at 18 months was 76.6 %). Sixty-seven children were excluded because all M-CHAT and/or ESAT data were missing, leaving 12,948 questionnaires for analysis. Non-systematic missing data on items were coded as system missing, did not add to the total scores and did, therefore, not contribute to a possible screen (positive) status. Cases were included if they had at least one item of the ESAT and M-CHAT filled in. M-CHAT and ESAT data were complete (all items filled in) for 11,952 of the 12,948 children (12,102 cases on M-CHAT items and 12,666 cases on ESAT items). The mean age of the children at completion of the questionnaire was 18.53 months (SD 0.54). The sample consisted of 6,616 boys (51.1 %), 6,290 girls (48.6 %), and 42 children whose sex was not reported (0.3 %). In the total sample, 91.8 % of the children had parents who were both native Norwegian speakers, and 87.7 % had grandparents who had Norwegian as mother tongue.

In a population-based design, it is of great importance to know if the sample is a good representation of the general population. For this study, family characteristics that are related to the age of identification of ASD are of special interest, such as parents' socio-economic status [19]. In the current study, the gross income of parents was divided into income groups with a range of 99,000 NOK. The median gross income group of the sample (including child support, unemployment benefits, and other allowances) was NOK 200,000–299,000 (€27,000–40,000) for mothers and NOK 300,000–399,000 (€40,500–54,000) for fathers, which was higher than that of the Norwegian population overall, since the median income in Norway 2003 was NOK 186,500 (€25,000) for women and NOK 285,600 (€38,000) for men [20].

### Procedure

When the child was 18 months, the mother was sent a questionnaire on maternal and child health. If the mother did not return the questionnaire within 3 weeks, a written reminder was sent. All forms were scanned and the data were quality controlled and de-identified before being

given to the researchers [16]. The research team was completely separate from any clinical services that were used by parents and their children, and the obtained results from the 18-month questionnaire were not communicated to and used by general practitioners.

### Instruments

The 18-month questionnaire included questions about the child's nutrition, growth, health, illnesses, medications, development, behavior, and everyday life, and about parental health and welfare. The section on child development and behavior incorporated the M-CHAT and the ESAT. The order of the items from the two instruments was not counterbalanced or interspersed. We used a fixed format.

### M-CHAT

The M-CHAT is an extension of the Checklist for Autism in Toddlers (CHAT) [3]. The M-CHAT includes the first nine parent-report items of the CHAT and 14 additional items were created based on hypotheses from the literature, clinical instruments used to evaluate older children, and the clinical experience of the M-CHAT authors. Six M-CHAT items, identified by discriminant function analysis (DFA), are considered essential: protodeclarative pointing, response to name, interest in peers, bringing things to show parents, following a point, and imitation. Children who fail any 3 of the 23 items or 2 of the 6 critical items are considered screen positive. The psychometric properties (sensitivity, specificity, positive and negative predictive power) and reliability of the M-CHAT have been investigated in both low- and high-risk samples aged between 16 and 48 months [8, 21–24]. In these studies, the proportion of the population that screened positive ranged from 5.7 to 14 %; the positive predictive value (PPV) was low (PPV 0.058–0.11), but was increased by telephone follow-up (0.57–0.65) [21, 24]. Pandey et al. [22] reported PPVs after telephone follow-up of 0.28 and 0.61 for younger and older low-risk groups of children, respectively. A recent study by Chlebowski et al. [25] showed that the M-CHAT is an effective screening instrument for ASD when the two-step screening procedure is used including the follow-up interview to reduce the number of false positives. They also found that children with a total score of 7 or higher can be directly referred for further clinical evaluation and bypass the follow-up interview. Another recent large population study showed that screening with M-CHAT alone in a general population at 18 months is not effective in identifying the majority of children who ultimately received a diagnosis of ASD, since only a third of the children diagnosed with ASD scored above the cut-off the M-CHAT at

18 months [26]. For the present study, the M-CHAT was translated (and back translated) into Norwegian from English (the original version). All 23 M-CHAT items are represented in the 18-month questionnaire. The scoring algorithm defines screen positive as failing 3 of the 23 M-CHAT items or 2 of 6 critical items. For the current study, we did not include a telephone interview, which is consistent with the American Academy of Pediatrics' recommendations on early screening for ASD [27].

### ESAT

The ESAT was developed to identify ASD in children in the (developmental) age of 0–36 months in the general population. It consists of 14 items and covers the domains of pretend play, joint attention, interest in others, eye contact, verbal and non-verbal communication, stereotypes, preoccupations, reaction to sensory stimuli, emotional reaction, and social interaction. Items of the ESAT are answered with “yes” for typical and “no” for atypical behavior. Children with “no” answers on at least 3 of the 14 ESAT items are screen positive [9]. The ESAT has been evaluated in a two-stage screening procedure in a study involving 14-month-old children ( $N = 31,724$ ) from the general population that had been prescreened by physicians at well-baby clinics using a 4-item prescreening instrument including the first four items on the ESAT. Children who failed any of the four items were considered prescreen positive, and parents were asked to complete the 14-item ESAT. Children who screened positive on the complete 14-item ESAT were invited for further systematic psychiatric examination. Eighteen of 73 high-risk children were subsequently diagnosed with ASD. The remaining high-risk children had developmental language disorder, or mental retardation [28]. The ESAT is recently introduced as a screening instrument and available in a educational package ESAT—“screening of ASD at a young age”, containing a set of ESAT questionnaires, and a theoretical and practical manual for identifying, screening and diagnosing ASD [29].

For the present study, the ESAT was translated (and back translated) into Norwegian from Dutch (the original version). The complete instrument 14-item ESAT was included in the 18-month questionnaire. The scoring algorithm defines screen positive as failing 3 of the 14 items.

### Referral status

The main outcome was referral to developmental services as reported by the parents in the 18-month questionnaire. This included referrals to educational services, child

habilitation units, and child psychiatry services. Each service has specific functions: the educational services are for the assessment of eligibility for resources due to learning difficulties; the child habilitation units are for the medical examination and diagnostic work-up for children with neurodevelopmental disorders; and the child psychiatry services are for mental health issues and the diagnostic work-up of behavior problems. All children with disabilities are ultimately seen by the educational services. The special needs of populations of these three different services, therefore, overlap.

### Other clinical and developmental characteristics

The health and development of the children that tested positive on the M-CHAT and/or on the ESAT was investigated, using preselected items from the 18-months questionnaire. Medical problems were assessed with four items reflecting aspects of child health often associated with poor development, including worries about physical development, hearing problems, diverging head circumference, and seizures. Up to 30 % of individuals with autism spectrum disorder (ASD) have comorbid epilepsy, while the prevalence of epilepsy in all children is 2–3 % [30, 31]. Motor development was assessed with six items on gross and fine motor development. Although not being core symptoms of autism, movement disorders and delayed motor development have been associated with ASD [32–34]. Parental concerns were assessed with three items on parental “worries” about the child’s physical development and behavior. Parental concerns about their child’s behavior are associated with mental health issues [10]. However, at least one study has found that parental worries, as scored using the Parents’ Evaluation of Developmental Status (PEDS), were not predictive of M-CHAT screen positivity [23]. Social, emotional, language and communication, and joint attention were assessed with items derived from standard developmental scales such as the Ages and Stages Questionnaires (ASQ) [35]; Emotionality, Activity, Shyness, Sociability Scale (EAS) [36, 37]; Social, Communication Questionnaire [SCQ; Previously called Autism Screening questionnaire (ASQ)] [38]; NonVerbal Communication Checklist (NVCC: Schjølberg, In preparation). The source of the items is indicated in Table 2 (because Table 2 consists mainly of results, the table is placed in the result section).

In this study, all items were binomial and scored in the same direction with a score 0 for a normal answer and score 1 for an abnormal answer. In this way, the percentage of children scoring on an item reflects the percentage of children having developmental, health, and behavioral problems.

## Data analyses

First, we investigated the overlap between the M-CHAT and ESAT by comparing the proportion of children that tested positive with either instrument, using the McNemar test [39, 40]. To measure the level of agreement between the M-CHAT and ESAT, a tetrachoric correlation was calculated with the statistical program R [41] using the Polycor package version 0.7-5 [42] and an algorithm for polychoric correlations reported by Drasgow [43] and Olsson [44]. The tetrachoric correlation estimates what the correlation between binary raters would be if ratings were made on a continuous scale [45], and the trait underlying the rating is conceived to be continuous. The tetrachoric correlation is preferable to Kappa if the observed prevalence of responses in one of two available categories is low. Its value is interpreted in the same way as a Pearson correlation [45]. Cohen [46] gives the following guidelines for the interpretation of effect sizes in social sciences: a correlation coefficient of 0.10 is thought to represent a weak or small association; a correlation coefficient of 0.30 is considered a moderate correlation; and a correlation coefficient of 0.50 or larger is thought to represent a strong or large correlation. However, the criteria for the interpretation of a correlation coefficient are in some ways arbitrary and should not be followed too strictly.

Secondly, we investigated whether testing positive on the M-CHAT or ESAT was associated with referral to developmental services by 18 months, by calculating PPV and negative predictive values (NPV). The PPV reflects the proportion of children that screened positive that were referred to any of the three developmental services (educational, habilitation, child psychiatry). The NPV reflects the proportion of children that screened negative that were not referred to any of the developmental services. PPV and NPV were calculated separately for the two instruments.

Thirdly, we investigated clinical and developmental differences between the children that screened positive or negative on the M-CHAT and ESAT. The percentage of children in each group whose mothers had indicated that their children had medical, motor developmental, emotional, language and communication, or social interaction problems or showed ‘abnormal’ joint attention and play (see Table 2) was compared, using a *z* test of two independent proportions. In addition, we compared the referred and non-referred children among ESAT and M-CHAT screen-positive children on the clinical and developmental variables using Chi-square tests or Fisher’s exact test. This could give some insight into the reason why children were referred to developmental services. The statistical analyses were performed with the statistical software package SPSS 20.0.

**Table 1** Crosstabs of ESAT screen scores by M-CHAT screen scores

		M-CHAT screen		
		Negative	Positive	Total
ESAT screen	Negative	12,102	775	12,877
	Positive	20	51	71
		93.5 %	6.0 %	99.5 %
		0.2 %	0.4 %	0.5 %
Total		12,122	826	12,948
		93.6 %	6.4 %	100.0 %

McNemar  $\chi^2 = 512.080$ ,  $p < 0.01$ ;  $r_{tet} = 0.685$

## Results

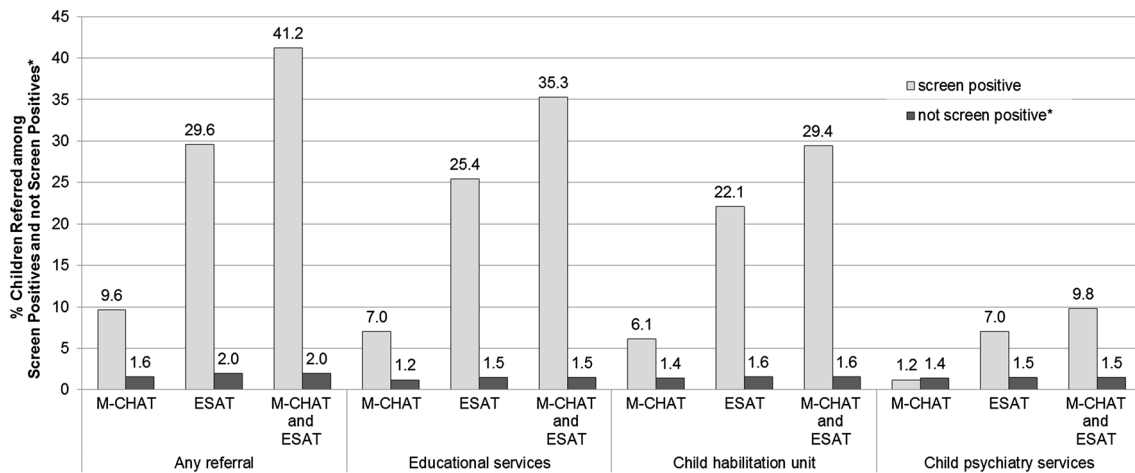
### Screen-positive scores of the M-CHAT and ESAT

Of the 12,948 children, 71 screened positive on the ESAT (0.5 %) and 826 screened positive on the M-CHAT (6.4 %) ( $p < 0.01$ ). Cross tabulation of ESAT by M-CHAT screening status showed that 93.5 % of all children screened negative on both instruments, 0.4 % screened positive on both instruments, 0.2 % screened positive on the ESAT but negative on the M-CHAT, and 6.0 % screened negative on the ESAT but positive on the M-CHAT (Table 1). The tetrachoric correlation was 0.685, indicating a strong agreement between the M-CHAT and ESAT.

ESAT and M-CHAT screen scores were the highest for the lower income groups and the lowest for the highest income groups. Although these differences were significant, partly due to the large numbers in the study, the magnitude of the differences for the ESAT and M-CHAT between the income groups was very small (maximum difference between income groups ESAT: 0.17; M-CHAT: 0.37), and the correlations between the separate income groups and the screen score of the ESAT and M-CHAT were very low ( $r_{\min-\max} = -0.012$  to 0.028). Further, no difference was found in the distribution of children referred across the income groups (father and mother), between 0.6 and 2.8 % of the children in each income group has been referred (income mother:  $\chi^2 = 9.91$ ;  $p = 0.13$ ; income father:  $\chi^2 = 8.73$ ;  $p = 0.27$ ).

### Screening status and clinical referral

In total, 184 children from the study population were referred to a developmental service. Of these 184 referred children, 79 (42.9 %) had screened positive on the M-CHAT, 21 (11.5 %) had screened positive on the ESAT, 21 (11.5 %) screened positive on both instruments and 103 (56.0 %) screened negative on both the ESAT and



**Fig. 1** Percentage of children who screened positive or negative on the M-CHAT, ESAT, or M-CHAT plus ESAT combined that was referred to developmental services \*The designation ‘Not screen positive’ refers to children that screened negative on the M-CHAT in the M-CHAT series, screened negative on the ESAT in the ESAT

M-CHAT. Reversely, 79 of 826 (9.6 %) children that tested positive on the M-CHAT were referred to developmental services, 21 of 71 (29.6 %) children that screened positive on the ESAT had been referred to developmental services, and 21 of 51 (41.2 %) children that screened positive on both instruments had been referred by 18 months. Figure 1 shows the percentages of children referred to the developmental services (children could be referred to more than one service).

The PPV of the ESAT was higher (range 0.070–0.296) than that of the M-CHAT (range 0.012–0.096) for overall referral and for referral to specific services than among M-CHAT screen positives. This difference was most marked for the category ‘Referred to child psychiatry’, where 7.0 % (5 of 71) of children that screened positive on the ESAT had been referred compared with 1.2 % (10 of 826) of the children that screened positive with the M-CHAT (Fig. 1). The PPV of screening with both instruments was consistently high for specific services, and more than 40 % for any service referral. The NPV was high for both instruments (values >0.98).

#### Other clinical and developmental concerns

For all domains examined (Table 2), a higher percentage of screen positives as compared to screen negatives was rated as abnormal. This held true for both the M-CHAT and ESAT. For example, 33.3 % of the children that screened positive on the ESAT but only 3.3 % of the children that screened negative on the ESAT showed delayed motor development. Likewise, 15.1 % of the children that screened positive on the M-CHAT but only 2.7 % of the children that screened negative on the M-CHAT showed

series, or screened negative on the M-CHAT and/or ESAT in the M-CHAT/ESAT series (all children except the children that screened positive on both instruments). Children could be referred to more than one service

delayed motor development. Two-sample *z* tests between independent proportions were performed to determine whether there was a significant difference between the percentages of children screening positive with the ESAT, M-CHAT or both the ESAT and M-CHAT on items concerning medical, motor developmental, emotional, language and communication, or social interaction problems or ‘abnormal’ joint attention and play. On nearly all items, the proportion of children with problems in the ESAT screen-positive group differed significantly from the proportion of children with problems in the M-CHAT screen-positive group. The proportion of children showing clinical and developmental problems was about twice as high among children that screened positive with the ESAT as among children that screened positive with the M-CHAT. There was no difference found between the proportion of problems between the ESAT screen-positive group and the group of children that screened positive with both the ESAT and M-CHAT (see Table 2).

In order to get insight into the clinical and developmental profiles and parental concerns among referred screen-positive children on the M-CHAT and/or ESAT, we compared those who were referred to and those who were not. Overall, a higher percentage of children who were screen positive on the M-CHAT and/or ESAT and were referred to a developmental service (R), showed motor and language problems and ‘abnormal’ joint attention and play compared to screen-positive children who were not referred (NR) ( $p < 0.05$ ). All percentages for referred and non-referred screen-positive children to any developmental service on the additional clinical and developmental variables are given in Online Resource 1. The percentage of children with behavioral, health and developmental

**Table 2** Percentage of children with additional characteristics, as rated by their parents, expressed in terms of the number of children screening positive or negative on the M-CHAT, ESAT, and M-CHAT plus ESAT combined ( $N = 12,948$ )

	1		2		3		Sign ( $p < 0.05$ )	
	M-CHAT		ESAT		Both M-CHAT and ESAT			
	Screen positive ( $N = 826$ ) (%)	Screen negative ( $N = 12,189$ )	Screen positive ( $N = 71$ ) (%)	Screen negative ( $N = 12,877$ )	Screen positive ( $N = 51$ ) (%)	Other ( $N = 12,897$ )		
<i>Medical problems</i>								
	Seizures: age 6–11 months	0.6	0.2	2.9	0.2	2.0	0.2	1 < 2
	Seizures: age 12–18 months	0.9	0.2	3.6	0.2	2.6	0.2	n.s.
	Hearing problems	6.4	1.7	15.2	1.9	15.2	1.9	1 < 2.3
	Worries: physical development	8.9	1.0	34.8	1.3	44.9	1.3	1 < 2.3
	Diverging head circumference	6.7	2.8	13.4	3.0	19.1	3.0	1 < 2.3
<i>Motor developmental problems</i>								
	Delayed motor development	15.1	2.7	33.3	3.3	45.7	3.3	1 < 2.3
ASQ	Walks rather than crawls	10.8	0.9	28.2	1.3	39.2	1.3	1 < 2.3
ASQ	Walks well and seldom falls	13.6	1.5	28.2	2.1	39.2	2.1	1 < 2.3
ASQ	Walks down stairs holding hands	20.8	6.6	35.7	7.3	49.0	7.3	1 < 2.3
ASQ	Stacks blocks	12.3	2.6	35.2	3.0	45.1	3.0	1 < 2.3
ASQ	Turns book pages	2.9	0.3	17.1	0.4	24.0	0.4	1 < 2.3
<i>Behavioral problems</i>								
	Worries: behavior	4.8	0.8	24.3	1.0	34.0	1.0	1 < 2.3
	Worries: difficult to handle	3.8	1.4	14.3	1.5	18.0	1.5	1 < 2.3
	Behavioral problems	1.1	0.1	6.1	0.1	8.7	0.1	1 < 2.3
<i>Emotional distress</i>								
EAS	Gets upset easily	23.5	18.3	36.6	18.5	35.3	18.6	1 < 2
EAS	Cries easily	10.9	6.1	21.1	6.4	19.6	6.4	1 < 2
<i>Language and communication</i>								
ASQ	Says 8 or more words	49.4	25.6	57.7	26.9	62.7	26.9	n.s.
	Delayed or aberrant language	10.8	0.6	35.8	1.1	48.9	1.1	1 < 2.3
NVCC	Uses sounds/words together with gestures	4.0	0.1	19.7	0.3	27.5	0.3	1 < 2.3
ASQ	Looks for objects when asked	16.7	1.8	36.6	2.6	49.0	2.6	1 < 2.3
<i>Social interaction</i>								
SCQ	Comforts others	47.3	24.8	56.3	26.1	60.8	26.1	n.s.
EAS	Prefers to play with others	16.4	9.0	38.6	9.3	46.0	9.3	1 < 2.3
<i>Joint attention and play</i>								
ASQ	Plays with doll or stuffed animal	9.5	1.8	23.9	2.2	29.4	2.2	1 < 2.3

**Table 2** continued

		1		2		3		Sign ( $p < 0.05$ )
		M-CHAT		ESAT		Both M-CHAT and ESAT		
		Screen positive ( $N = 826$ ) (%)	Screen negative ( $N = 12,189$ )	Screen positive ( $N = 71$ ) (%)	Screen negative ( $N = 12,877$ )	Screen positive ( $N = 51$ ) (%)	Other ( $N = 12,897$ )	
ASQ	Points to correct picture when asked	31.1	8.4	47.1	9.6	62.0	9.6	1 < 2.3
NVCC	Points distally when asked	17.4	2.0	33.8	2.8	45.1	2.8	1 < 2.3
ASQ	Comes to you needing help	13.2	1.1	34.3	1.7	47.1	1.7	1 < 2.3
ASQ	Copies activities	9.2	0.4	31.4	0.8	43.1	0.8	1 < 2.3
ASQ	Points imperative	8.2	0.1	26.8	0.5	37.3	0.5	1 < 2.3
ASQ	Gets attention by pulling hand	11.0	1.8	23.9	2.3	31.4	2.3	1 < 2.3
NVCC	Follows gaze	5.4	0.5	19.7	0.7	27.5	0.7	1 < 2.3

First column shows the name of the instrument where the item is derived from *ASQ* Ages and Stages Questionnaires [35], *EAS* Emotionality, Activity, Shyness, Sociability Scale [36, 37]; *SCQ* Social Communication Questionnaire (acronym for: Autism Screening Questionnaire) [38], *NVCC* Non-verbal Communication Checklist (Schjølberg, in preparation)

problems among referred ESAT screen positives tended to be higher than among referred M-CHAT screen positives, this seems especially the case for ‘abnormal’ joint attention and play. For the children screening positive on both the M-CHAT and ESAT, the percentages are comparable to the ESAT screen-positive group. Of special interest are the characteristics of the referred screen-positive children to child psychiatry services. There were only significant differences ( $p < 0.05$ ) between referred and non-referred children to psychiatry services among M-CHAT screen positives for behavioral problems, delayed or aberrant language, fine motor problems and imperative pointing. For the ESAT and M-CHAT/ESAT screen positives, no significant differences were found for psychiatry services, even though the percentages of these groups are higher than those of the M-CHAT screen-positives group and the percentages of problems for the referred children tended to be higher than for the non-referred children. For referrals to the Child Habituation Unit and School Psychology Services, the most marked differences between referred and non-referred children were found for motor developmental, language and joint attention problems and none or less for behavioral problems.

## Discussion

This study is one of the first to use parent-completed (without involvement of health professionals) screening instruments (ESAT and M-CHAT) in an unselected general

population. The aims of the study, which involved 18-month-old children, were to explore the differences between the proportions and the overlap thereof of children that screened positive on the ESAT or the M-CHAT, to investigate the association between screening positive on either instrument and referral to developmental services, and to establish the clinical and developmental profiles and parental concerns of children that screened positive on either or both instruments.

Results indicated a significant tenfold difference in the proportion screening positives on M-CHAT (6.4 %), an ESAT (0.5 %). These findings may suggest that the M-CHAT (without follow-up interview) is overinclusive and/or the ESAT is too conservative. The prevalence of ASD among older children is estimated around 0.9 % [47], lower than the M-CHAT screen-positive rate (6.4 %) and higher than that of the ESAT (0.5 %). The relatively high screen-positive rates of the M-CHAT in the current study are in line with previous studies (6–14 %) [21–24]. The psychometric properties of the ESAT have yet to be tested in a general unselected population.

Further analysis revealed differences in service referral between the two instruments. Almost 30 % of the children that screened positive on the ESAT had been referred to any specialist services by 18 months, which was nearly three times higher than the referral rate of children screening positive on the M-CHAT. Among children that screened positive on both instruments almost half had been referred to any services by 18 months. Of the children that screened positive on either the M-CHAT or the ESAT,



most were referred to educational services; few were referred to child psychiatry services. However, the proportion of children referred to child psychiatry services was six times higher among children that screened positive on the ESAT than among children that screened positive on the M-CHAT. This suggests that children who screen positive on the ESAT have more psychiatric problems than children that screen positive on the M-CHAT. Across all developmental services, a relatively higher percentage of screen-positive children on the ESAT compared to on the M-CHAT tended to have ‘abnormal’ joint attention and play, which is considered a primary marker for early detection of ASD. Results suggest further that the M-CHAT possibly identifies children with (milder) lower intellectual and adaptive functioning. This is in line with another large population study of MoBa, which shows that as many as 7.3 % of the children in the non-ASD group scored above cut-off on the M-CHAT 23-item criterion at 18 months, and that only a third of the children screening positive at 18 months were diagnosed with ASD at a later age [26].

The PPV of the ESAT and M-CHAT for clinical referral at 18 months was moderate for children that screened positive on both instruments, low for children that screened positive on the ESAT, and very low for children that screened positive on the M-CHAT. The PPV of the M-CHAT for clinical referral ranging from 0.012 to 0.096 was comparable with PPVs for ASD reported by Kleinman et al. [21] and Robins [24], namely, 0.11 and 0.058, respectively, for children that screened positive on the M-CHAT. The PPV of the M-CHAT in this study could be higher if we had incorporated the follow-up interview or used the suggested cut-off of 7 which bypasses the follow-up interview [25]. However, using the more conservative cut-off of 7, we would have missed some of the children with screen scores between 3 and 6, who have perhaps milder symptoms, but still have an increased risk for having developmental problems. There are no reports of the PPV of the ESAT for ASD in the general population at very young age. The low PPV we found in this study could reflect the very young age of the sample coupled with the fact that ASD is often diagnosed late.

Interestingly, it seems there were also qualitative differences in the nature of problems that were picked up by the instruments. In all domains of health and development evaluated in the exploratory analysis, a higher proportion of screen positives as compared to screen negatives, endorsed abnormality on other clinical and developmental problems, as rated by their parents. Overall, a higher proportion of children that screened positive on the ESAT than positive on the M-CHAT had these problems. Although not significantly different from ESAT screen positives, parents of children that screened positive on both instruments

reported more clinical and developmental concerns in our post hoc analysis. These children also had the highest referral rates for developmental services. This may be because the use of a combination of screening instruments means that there are more questions or items, and that sometimes more than one item addresses a certain aspect or behavior. This could increase the likelihood that parental concerns about developmental problems are addressed. Further, more than one question about a particular aspect or behavior might make parents more aware of the importance of this aspect or behavior. This suggests that investigation of the predictive value of individual items might lead to further optimizing the screening instruments.

In sum, when choosing a screening instrument for the general population, investigators need to weigh the balance between false positives and false negatives. If the goal is to identify children at risk who should be closely monitored, it might be better to select too many rather than too few children for follow-up, making the M-CHAT the preferred instrument. However, selecting children with false-positive screening results for further follow-up might cause unnecessary parental anxiety and increase costs, making the ESAT the preferred instrument. Moreover, the ESAT differentiates more than the M-CHAT between the degree and content of parental concerns for children being screening positive and children being screening negative, which strongly contributes to the validity in being able to identify children at high risk for autism spectrum disorders or other serious developmental problems. As mentioned before, identifying children at high risk for autism spectrum disorders at an early age with an standardized instrument enhances exchange of thoughts for parents as well as health care providers, reduces diagnostic delay and subsequently parental distress and enables early intervention which leads to better outcomes. Investigation of the predictive value of individual items might be warranted to be able to give a recommendation or provide a strategy for the use of different questionnaires.

## Limitations

The cross-sectional nature of the study means that we cannot establish whether, for example, parental concerns preceded referral to developmental services, or whether referral triggered symptom detection. Our data are based on parental information only. This could lead to an underestimating of the PPV of both instruments, because it is plausible that concerns increase as children get older. Parents, however, are the main and in almost all cases the sole informants about very young children’s behavior problems, and there is a body of evidence indicating the merit and validity of parent information [10–13].

Final diagnoses (ASD: yes or not; mental handicap: yes or not) are crucial for examining the psychometric properties of the screening performance. However, outcome measures as being clinically referred, presence of parental concerns, and behavioral and medical measures are very relevant and generally accepted proxy measures of “case-ness”. These proxy measures can give us better insights into the utility of the screening performance, pending more final analyses based on more specific clinical diagnosis.

As mentioned earlier, in a population-based design it is of great importance to know if the sample is a good representation of the general population. Although many contextual factors could play a role, we believe that SES stands at the basis of this comparison, since SES in itself is related to many other contextual factors and parents’ socio-economic status is found to be related to the age of identification of ASD [19]. However, Norway has relatively small differences in socio-economic status and ethnicity in the population compared to some other countries and the health- and welfare-system in Norway is well developed. In the current study, we have a sample with somewhat higher SES than the population, no relation between SES and referral status, but slightly higher scores in the lower income groups. This means that the screen scores found on the ESAT and M-CHAT could have been even slightly higher when more low SES families would have participated. This effect applies to both the instruments and counteracted rather than augmented our finding that the high scores on the M-CHAT items in particular may lead to over-identifying high-risk children. We feel justified to conclude that our results have not been biased to any relevant extent by SES factors.

### Strengths

The strength of the study is its population-based design. Moreover, the wealth of questionnaire data made it possible to investigate the screening instruments in relation to other variables related to health and development. Lastly, the longitudinal design of the MoBa mother-and-child cohort study provides the opportunity to determine the sensitivity and specificity of these screening instruments as children grow older and undergo diagnostic assessments. The longitudinal information obtained in the MoBa study might provide insight into population screening and public health requirements for an adequate instrument, and the costs of screening large populations.

### Conclusion

Our findings suggest that children who screen positive on the ESAT and the M-CHAT have different profiles in terms

of their clinical and developmental characteristics. The ESAT identified fewer children as being at risk of ASD than the M-CHAT, but a higher proportion of children that screened positive on the ESAT was referred to developmental services and the ESAT tended to identify more children with medical, emotionally, language, and behavioral problems. The M-CHAT identified more children at risk of ASD, but these children were less referred to developmental services and had fewer problems associated with other aspects of children’s development, health, and behavior than the children that screened positive with the ESAT. Since a post hoc analysis of combining the two instruments appeared to be more effective than the individual instruments alone in identifying children referred to clinical services at 18 months, further analysis at the level of single items is warranted to improve these screening instruments.

**Acknowledgments** The Norwegian Mother and Child Cohort Study is supported by the Norwegian Ministry of Health and the Ministry of Education and Research, NIH/NIEHS (Contract No. N01-ES-75558), NIH/NINDS (Grant No. 1 UO1 NS 047537-01 and Grant No. 2 UO1 NS 047537-06A1). We are grateful to all the participating families in Norway who take part in this ongoing cohort study.

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

### References

1. Chakrabarti S, Haubus C, Dugmore S, Orgill G, Devine F (2005) A model of early detection and diagnosis of autism spectrum disorder in young children. *Infants Young Child* 18(3):200–211
2. Harrington JW, Rosen L, Garnecho A (2006) Parental perceptions and use of complementary and alternative medicine practices for children with autistic spectrum disorders in private practice. *J Dev Behav Pediatr* 27(2):S156–S161
3. Baron-Cohen S, Allen J, Gillberg C (1992) Can autism be detected at 18 months—the needle, the haystack, and the chat. *Br J Psychiatry* 161:839–843
4. Schall C (2000) Family perspectives on raising a child with autism. *J Child Family Stud* 9:409–423
5. Corsello CM (2005) Early intervention in autism. *Infants Young Child* 18(2):74–85
6. Rogers SJ (1996) Brief report: early intervention in autism. *J Autism Dev Disord* 26(2):243–246
7. Leekam S, Tandos J, McConachie H, Meins E, Parkinson K, Wright C, Turner M, Arnott B, Vittorini L, Le Couteur A (2007) Repetitive behaviours in typically developing 2-year-olds. *J Child Psychol Psychiatry* 48(11):1131–1138
8. Robins DL, Fein D, Barton ML, Green JA (2001) The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disord* 31(2):131–144
9. Swinkels SHN, Dietz C, van Daalen E, Kerkhof IHGM, van Engeland H, Buitelaar JK (2006) Screening for autistic spectrum

- in children aged 14 to 15 months. I: the development of the early screening of autistic traits questionnaire (ESAT). *J Autism Dev Disord* 36(6):723–732
10. Glascoe FP (2003) Parents' evaluation of developmental status: how well do parents' concerns identify children with behavioral and emotional problems? *Clin Pediatr (Phila)* 42(2):133–138
  11. Glascoe FP (1999) The value of parents' concerns to detect and address developmental and behavioural problems. *J Paediatr Child Health* 35(1):1–8
  12. Tervo RC (2005) Parent's reports predict their child's developmental problems. *Clin Pediatr (Phila)* 44(7):601–611
  13. Dietz C, Swinkels SHN, van Daalen E, van Engeland H, Buitelaar JK (2007) Parental compliance after screening social development in toddlers. *Arch Pediatr Adolesc Med* 161(4):363–368
  14. Carter AS, Briggs-Gowan MJ, David NO (2004) Assessment of young children's social-emotional development and psychopathology: recent advances and recommendations for practice. *J Child Psychol Psychiatry* 45(1):109–134
  15. Sices L, Feudtner C, McLaughlin J, Drotar D, Williams M (2004) How do primary care physicians manage children with possible developmental delays? A national survey with an experimental design. *Pediatrics* 113(2):274–282
  16. Magnus P, Irgens LM, Haug K, Nystad W, Skjaerven R, Stoltenberg C (2006) Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol* 35(5):1146–1150
  17. Ronningen KS, Paltiel L, Meltzer HM, Nordhagen R, Lie KK, Hovengen R, Haugen M, Nystad W, Magnus P, Hoppin JA (2006) The biobank of the Norwegian mother and child cohort Study: a resource for the next 100 years. *Eur J Epidemiol* 21(8):619–625
  18. Stoltenberg C, Schjolberg S, Bresnahan M, Hornig M, Hirtz D, Dahl C, Lie KK, Reichborn-Kjennerud T, Schreuder P, Alsaker E, Oyen AS, Magnus P, Suren P, Susser E, Lipkin WI (2010) The autism birth cohort: a paradigm for gene-environment-timing research. *Mol Psychiatry* 15(7):676–680
  19. Fountain C, King MD, Bearman PS (2011) Age of diagnosis for autism: individual and community factors across 10 birth cohorts. *J Epidemiol Community Health* 65(6):503–510. doi:10.1136/jech.2009.104588
  20. Kristiansen JE, Sandnes T (2006) Women and men in Norway. What the figures say. Statistics Norway, Oslo
  21. Kleinman JM, Robins DL, Ventola PE, Pandey J, Boorstein HC, Esser EL, Wilson LB, Rosenthal MA, Sutera S, Verbalis AD, Barton M, Hodgson S, Green J, Dumont-Mathieu T, Volkmar F, Chawarska K, Klin A, Fein D (2008) The Modified Checklist for Autism in Toddlers: a follow-up study investigating the early detection of autism spectrum disorders. *J Autism Dev Disord* 38(5):827–839
  22. Pandey J, Verbalis A, Robins DL, Boorstein H, Klin A, Babitz T, Chawarska K, Volkmar F, Green J, Barton M, Fein D (2008) Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism* 12(5):513–535
  23. Pinto-Martin JA, Young LM, Mandell DS, Poghossyan L, Giarelli E, Levy SE (2008) Screening strategies for autism spectrum disorders in pediatric primary care. *J Dev Behav Pediatr* 29(5):345–350
  24. Robins DL (2008) Screening for autism spectrum disorders in primary care settings. *Autism* 12(5):537–556
  25. Chlebowski C, Robins DL, Barton ML, Fein D (2013) Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics* 131(4):e1121–e1127. doi:10.1542/peds.2012-1525
  26. Stenberg N, Bresnahan M, Gunnes N, Hirtz D, Hornig M, Lie KK, Lipkin WI, Lord C, Magnus P, Reichborn-Kjennerud T, Schjolberg S, Suren P, Susser E, Svendsen BK, von Tetzchner S, Oyen AS, Stoltenberg C (2014) Identifying children with autism spectrum disorder at 18 months in a general population sample. *Paediatr Perinat Epidemiol*. doi:10.1111/ppe.12114
  27. Johnson CP, Myers SM, American Academy of Pediatrics Council on Children With D (2007) Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 120(5):1183–1215. doi:10.1542/peds.2007-2361
  28. Dietz C, Swinkels S, van Daalen E, van Engeland H, Buitelaar JK (2006) Screening for autistic spectrum disorder in children aged 14–15 months. II: population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). Design and general findings. *J Autism Dev Disord* 36(6):713–722
  29. Buitelaar JK, van Daalen E, Dietz C, van Engeland H, van der Gaag RJ, van Steijn D, Swinkels SHN (2009) ESAT—screening van ASS op jonge leeftijd. Bohn Stafleu van Loghum, Houten
  30. Stafstrom CE, Hagerman PJ, Pessah IN (2010) Epilepsy in autism spectrum disorders. *Epilepsia* 51:78
  31. Tuchman R, Rapin I (2002) Epilepsy in autism. *Lancet Neurol* 1(6):352–358
  32. Esposito G, Venuti P (2008) Analysis of toddlers' gait after six months of independent walking to identify autism: a preliminary study. *Percept Mot Skills* 106(1):259–269
  33. Landa R, Garrett-Mayer E (2006) Development in infants with autism spectrum disorders: a prospective study. *J Child Psychol Psychiatry* 47(6):629–638
  34. Provost B, Lopez BR, Heimerl S (2007) A comparison of motor delays in young children: autism spectrum disorder, developmental delay, and developmental concerns. *J Autism Dev Disord* 37(2):321–328
  35. Squires J, Bricker D, Potter L (1997) Revision of a parent-completed developmental screening tool: ages and stages questionnaires. *J Pediatr Psychol* 22(3):313–328
  36. Buss AH, Plomin R (1984) Temperament: early developing personality traits. Lawrence Erlbaum, Hillsdale
  37. Mathiesen KS, Tambs K (1999) The EAS Temperament Questionnaire—factor structure, age trends, reliability, and stability in a Norwegian sample. *J Child Psychol Psychiatry* 40(3):431–439
  38. Berument SK, Rutter M, Lord C, Pickles A, Bailey A (1999) Autism screening questionnaire: diagnostic validity. *Br J Psychiatry* 175:444–451
  39. Houwelingen JC, Stijnen T, van Strik R (1993) Inleiding tot de medische statistiek. Bunge, Utrecht
  40. Petrie A, Sabin C (2005) Medical statistics at glance, 2nd edn. Blackwell, Malden
  41. Team RDC (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
  42. Fox J (2007) Polycor: polychoric and polyserial correlations. R Package Version 0.7-5. Available at <http://cran.r-project.org/>
  43. Drasgow F (1986) Polychoric and polyserial correlations. In: Kotz S, Johnson N (eds) The encyclopedia of statistics, vol 7. Wiley, New York, pp 68–74
  44. Olsson U (1979) Maximum likelihood estimation of the polychoric correlation-coefficient. *Psychometrika* 44(4):443–460
  45. Uebersax JS (2006) The tetrachoric and polychoric correlation coefficients. Statistical Methods for Rater Agreement web site. <http://john-uebersax.com/stat/tetra.htm>
  46. Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Erlbaum, Hillsdale
  47. Centers for Disease Control and Prevention (2009) Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, United States 2006. Surveillance summaries, December 18, 2009. *MMWR* 58(SS-10):1–24