Autism in infancy and toddlerhood

Investigating infants at risk and enhancing early detection and intervention

Mirjam K. J. Pijl
Autism in infancy and toddlerhood
Investigating infants at risk and enhancing early detection and intervention

Mirjam Katharina Jacomina Pijl
The research described in this paper was carried out at the Donders Institute for Brain, Cognition and Behaviour at the Department for Cognitive Neuroscience of the Radboud University Medical Centre Nijmegen (Nijmegen, the Netherlands), Karakter Child and Adolescent Psychiatry University Centre (Nijmegen, the Netherlands), Baby Research Center (Nijmegen, the Netherlands), King’s College London at the Psychology Department, Institute of Psychiatry (London, United Kingdom), Centre for Brain and Cognitive Development, Birkbeck College, the University of London (London, United Kingdom), Ghent University at the Department of Experimental-Clinical and Health Psychology (Ghent, Belgium), Center of Neurodevelopmental Disorders at Karolinska Institutet (KIND) at the Department of Women’s and Children’s Health (Stockholm, Sweden), and Utrecht University at the Department of Developmental Psychology (Utrecht, the Netherlands).

The research leading to this thesis was primarily supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement n° 115300, resources of which are composed of financial contribution from the European Union’s Seventh Framework Programme (FP7/2007-2013) and EFPIA companies’ in kind contribution. The work published in this thesis was further supported by the Horizon2020 funded Marie-Curie ITN programme BRAINVIEW (nr 642996), and by the Korczak Foundation’s support for the DIANE project.

Cover design: Proefschrift All In One
Cover photo (back side): Tom Pijl
Printed by: Proefschrift All In One


© M.K.J. Pijl, Utrecht, 2018. All rights reserved.

No part of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without prior written permission of the author.
Autism in infancy and toddlerhood
Investigating infants at risk and enhancing early detection and intervention

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op woensdag 26 juni 2019
om 10.30 uur precies

door

Mirjam Katharina Jacomina Pijl
geboren 17 juli 1989
te Berkel en Rodenrijs
Promotor: Prof. dr. Jan K. Buitelaar
Copromotoren: Dr. Iris. J. Servatius-Oosterling (Karakter Kinder- en Jeugdpsychiatrie Universitair Centrum Nijmegen)
Dr. Nanda N.J. Lambregts-Rommelse
Manuscriptcommissie: Prof. dr. Carolina de Weerth
Prof. dr. Hedwig J.A. van Bakel (Tilburg University)
Prof. dr. Maretha V. de Jonge (Universiteit Leiden)
Verbluft bewonder ik het grootse van de kleinen.

Victor Hugo (1802-1885)
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1</th>
<th>General introduction</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Part I</strong></td>
<td><strong>Early risk markers of autism spectrum disorder in infancy and toddlerhood</strong></td>
<td></td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Temperament as an early risk marker for autism spectrum disorders? A longitudinal study of high-risk and low-risk infants</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td><em>Published in Journal of Autism and Developmental Disorders</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Parent-child interaction during the first year of life in infants at high- and low-risk of autism spectrum disorder</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td><em>Submitted for publication</em></td>
<td></td>
</tr>
<tr>
<td><strong>Part II</strong></td>
<td><strong>Early detection and intervention of autism spectrum disorder</strong></td>
<td></td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Sustainability of an early detection program for autism spectrum disorder over the course of 8 years</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td><em>Published in Autism</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 5</td>
<td>Does the Brief Observation of Social Communication Change (BOSCC) help moving forward in measuring change in early autism intervention studies?</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td><em>Published in Autism</em></td>
<td></td>
</tr>
<tr>
<td>Chapter 6</td>
<td>Summary and general discussion</td>
<td>113</td>
</tr>
<tr>
<td><strong>Appendices</strong></td>
<td>Nederlandse samenvatting (summary in Dutch)</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>Dankwoord (acknowledgements in Dutch)</td>
<td>162</td>
</tr>
<tr>
<td></td>
<td>Publication list</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>About the author</td>
<td>170</td>
</tr>
</tbody>
</table>
General introduction
This thesis focuses on several important issues concerning autism spectrum disorder (ASD) in infancy and toddlerhood. The introduction starts with a general description of ASD and its early clinical manifestations and etiology. Next, potential early risk markers of ASD are discussed. In addition, the importance of early detection and intervention is described and the need to improve both. This chapter concludes with the specific aims and outline of the thesis.

**Autism spectrum disorder**

Autism spectrum disorder is one of the earliest emerging neurodevelopmental conditions, affecting approximately 1% of the general population (Baird et al., 2006; Elsabbagh et al., 2012). The current descriptive diagnostic criteria for ASD include persistent deficits in social communication and interaction, restricted, repetitive patterns of behavior, interests or activities, and atypical sensory processing (see Box 1.1 | American Psychiatric Association, 2013). Although individuals diagnosed with ASD share impairments in these domains, ASD demonstrate considerable phenotypic heterogeneity, both in terms of symptom severity and course (Georgiades, Bishop, & Frazier, 2017; Georgiades, Szatmari, Boyle, et al., 2013; Wiggins, Robins, Adamson, Bakeman, & Henrich, 2012). Individuals who are diagnosed with ASD experience long-term, often lifetime, consequences (Billstedt, Gillberg, & Gillberg, 2005). The majority of individuals remain impaired to some degree in the ability to communicate and socialize. However, growing scientific evidence and expert consensus suggest that the age at intervention has a direct impact on the individuals' long-term outcome (i.e. the earlier, the better), underlining the importance to screen and diagnose as early as possible (Dawson et al., 2012; Estes, Munson, et al., 2015; Green et al., 2010; Pickles et al., 2016). Further, other problems and neurodevelopmental conditions frequently co-occur with ASD potentially leading to consider these as a differential diagnosis. These include attention deficit hyperactivity disorder (ADHD), learning disabilities, language impairments, epilepsy, motor control problems, depression, anxiety, gastrointestinal problems, and sleep disorders (Gillberg, 2010; Simonoff et al., 2008; Visser, Rommelse, Greven, & Buitelaar, 2016).
Autism spectrum disorder (ASD) is one of the earliest emerging neurodevelopmental conditions, affecting approximately 1% of the general population (Baird et al., 2006; Elsabbagh et al., 2012). The current descriptive diagnostic criteria for ASD include persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following, currently or by history (examples are illustrative, not exhaustive, see text):

1. Deficits in social-emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.
2. Deficits in nonverbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.
3. Deficits in developing, maintaining, and understanding relationships, ranging, for example, from difficulties adjusting behavior to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers.

B. Restricted, repetitive patterns of behavior, interests, or activities, as manifested by at least two of the following, currently or by history (examples are illustrative, not exhaustive; see text):

1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g. simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).
2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns or verbal nonverbal behavior (e.g. extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat food every day).
3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g. strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interest).
4. Hyper- or hyporeactivity to sensory input or unusual interests in sensory aspects of the environment (e.g. apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement).

Specify current severity: Severity is based on social communication impairments and restricted, repetitive patterns of behavior.

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level.

Box 1.1 DSM-5 diagnostic criteria for ASD

A. Persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following, currently or by history (examples are illustrative, not exhaustive, see text):

1. Deficits in social-emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.
2. Deficits in nonverbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.
3. Deficits in developing, maintaining, and understanding relationships, ranging, for example, from difficulties adjusting behavior to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers.

B. Restricted, repetitive patterns of behavior, interests, or activities, as manifested by at least two of the following, currently or by history (examples are illustrative, not exhaustive; see text):

1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g. simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases).
2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns or verbal nonverbal behavior (e.g. extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat food every day).
3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g. strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interest).
4. Hyper- or hyporeactivity to sensory input or unusual interests in sensory aspects of the environment (e.g. apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement).

Specify current severity: Severity is based on social communication impairments and restricted, repetitive patterns of behavior.

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level.
Precursors and early clinical manifestations of autism spectrum disorder in infants and toddlers

As is currently described in the DSM-5, symptoms of ASD and associated impairment of functioning must be present in the early developmental period (American Psychiatric Association, 2013). Therefore, knowledge about the early development of ASD is needed to reliably diagnose ASD. A summary of known early indicators of ASD in the first years of life is presented below.

Social communication and interaction

The first 12 months. Some research has shown that in the first 6 months of life infants who later develop ASD appear to use typical patterns of attention, gaze and affect in interaction with their parent (Rozga et al., 2011; Wan et al., 2012; Young, Merin, Rogers, & Ozonoff, 2009; Zwaigenbaum et al., 2005) or an examiner (Elsabbagh et al., 2013; Ozonoff et al., 2010). In contrast, Chawarska, Macari, and Shic (2013) found that 6-month-old infants who were subsequently diagnosed with ASD were less attentive to an examiner’s face in a naturalistic video as compared to infants with other outcomes. These contrasting results may be partly explained by a difference in methods used (live interaction vs. eye tracking). By 9 to 12 months infants diagnosed with ASD are less likely to respond to their own name (Feldman et al., 2012b; Miller et al., 2017; Nadig et al., 2007), whereas typically developing infants tend to respond to their own name around 4 to 6 months of age. Additionally, 11-month-old infants who are later diagnosed with ASD show less overall engagement during interaction with their parent as compared to controls, but they do not differ in levels of directed vocalizations, positive affect, or pointing or requesting interest (Campbell, Leezenbaum, Mahoney, Day, & Schmidt, 2015).

After the first birthday. From the first birthday onward deficits in social communication and interaction become more pronounced. Delays or atypicalities in the child’s speech and communicative development are among the most frequent initial parental concerns. By 12 to 14 months of age delays in both receptive and expressive language are observed in infants who are later diagnosed with ASD, as measured by parent-report and performance-based measures (Estes, Zwaigenbaum, et al., 2015; Landa & Garrett-Mayer, 2006; Mitchell et al., 2006; Zwaigenbaum et al., 2005). Research has shown that the divergence from typical development increases further over the second and third year of life (Estes, Zwaigenbaum, et al., 2015; Landa, Gross, Stuart, & Faherty, 2013). In contrast, not all studies have found consistent group differences in early language skills (Hudry et al., 2014; Talbott, Nelson, & Tager-Flusberg, 2015), which highlights the heterogeneous character of communication problems among children with ASD. Furthermore, toddlers with ASD show reduced attentiveness or social referencing, less gaze to faces, diminished
### Box 1.2 Case descriptions

**Thomas**

Thomas’ parents’ first worries arose when he was only 19 months old. The main parental concerns were Thomas’ delay in spoken language and being also late in reaching motor milestones. Already during his first year of life, Thomas’ mother felt that there was something special about him, but did not really express any concerns. Physiotherapy started when Thomas was 24 months old, but the well-baby clinic professional reassured them that Thomas would probably acquire comprehensible language by the age of 3. Staff at the day-care centre Thomas attended for one day a week said that he was not difficult to manage, and that although he did not play much with his peers, he played happily by himself and his speech was gradually getting better.

Just after his fourth birthday, Thomas went to primary school. In class, he often just fiddled around, and although he was compliant and seemed happy, he preferred to play alone. The concerns that Thomas’ parents had raised long time ago at the well-baby clinic were now also voiced by his teacher. His parents insisted on referral for a thorough developmental evaluation, and Thomas was diagnosed with ASD at the age of 4 years and 5 months. Thomas and his parents enrolled in Pivotal Response Training. Despite his parents’ early concerns, there had been a huge delay in the diagnosis of ASD. The delay caused parental distress about Thomas’ early development, and Thomas’ missed opportunities for stimulation of development due to rather late enrolment in a targeted intervention program.

**Maria – Thomas’ little sister**

Parents had received genetic counseling after their son Thomas had been diagnosed with ASD, so they were aware of the heightened risk of ASD and therefore closely monitored Maria’s development during the first years of her life. Her parents reported age-appropriate cognitive and language development, she could happily play together with her nephews and nieces, and was a big fan of dancing (she just couldn’t stop). Therefore, parents did not have any concerns that Maria would also develop ASD. However, her parents experienced difficulties during daily activities starting around the age of 14 months, and when she started to attend the nursery school at 36 months, the challenges became bigger. Maria covered her ears and panicked anytime she heard loud noises. Also, she did not like to be cuddled. When they discussed the challenges with the treating clinician of her older brother Thomas, they were referred for a developmental evaluation and Maria was diagnosed with ASD at 3 years and 3 months. She started to attend a special day-care centre for children with developmental problems.
eye contact, fewer directed vocalizations, limited social smiling and range of facial expressions, limited shared enjoyment, imitation deficits, and lower frequency and variety of gestures (Jones, Gliga, Bedford, Charman, & Johnson, 2014; Volkmar, Chawarska, & Klin, 2005; Zwaigenbaum, Bryson, & Garon, 2013). There is also evidence of atypicalities in the use of joint attention (i.e. triadic coordination of attention among the child, a second person, and a third entity, with a shared focus of the child and the second person on the third entity) (Bruinsma, Koegel, & Koegel, 2004; Landa, Holman, & Garrett-Mayer, 2007; Macari et al., 2012; Osterling & Dawson, 1994).

Restrictive and repetitive behaviors
The first 12 months. Although restricted and repetitive behaviors are a core feature of ASD, research on their definition, cause, and capacity for change has been relatively neglected (Leekam, Prior, & Uljarevic, 2011). The very few studies that investigated restrictive and repetitive behaviors during infancy show that unusual sensory behaviors may differentiate infants who later develop ASD from typically developing infants (Baranek, 1999; Zwaigenbaum et al., 2005). For example, infants who are later diagnosed with ASD may be more likely to show stereotyped, self-stimulatory use of objects (e.g. the infant may dangle a string of beads and wave them in front of his/her eyes) by the end of the first year (but not before) than infants with typical development (Zwaigenbaum et al., 2005). More research into restrictive and repetitive behaviors during the first year, including sensory-related behaviors, is needed. Given that restricted and repetitive behaviors are part of typical development, especially in infancy, it is important to compare trajectories of these behaviors in typical development with potential atypical trajectories in children with ASD.
After the first birthday. Previous research has indicated that early restrictive and repetitive behaviors are possibly more important for the early detection of ASD than aspects of social communication and interaction (Swinkels et al., 2006). From 12 months onward infants who are later diagnosed with ASD may display hand and finger mannerisms, inappropriate exploration or use of objects, repetitive interests or play, self-injurious behaviors, compulsive behaviors, ritual-sameness behaviors, and/or unusual sensory behaviors (i.e. hyper- or hyposensitivity to sensory stimuli) (Barber, Wetherby, & Chambers, 2012; Ben-Sasson et al., 2007; Christensen et al., 2010; Germani et al., 2014; Loh et al., 2007; Ozonoff et al., 2008; Wolff et al., 2014). However, whether these behaviors are specific precursors to ASD in the first year of life remain largely unknown and should be investigated in future work. For example, unusual sensory behaviors in infancy and toddlerhood may reflect poor emotion regulation and atypical behavioral responses as part of a trajectory toward ASD. This thesis will focus on a small part of this research area by investigating temperament (such as regulation of behavior and emotions) as a potential risk marker for ASD (see Chapter 2).

Note. These pictures illustrate a fascination with spinning objects, and a rejection of social touch.

Other markers
Although not currently included in the diagnostic features of ASD, early difficulties in the domain of motor development are common. Research has shown that infants who later develop ASD have lower fine and gross motor skills from 12 months onward based on parent-reported and experimental measures (Landa & Garrett-Mayer, 2006; Zwaigenbaum et al., 2005). In addition, abnormalities in postural control (specifically, head lag) may be detected as early as 6 months (Flanagan, Landa, Bhat, & Bauman, 2012). However, evidence for early motor atypicalities or delays is mixed. For example, Leonard, Elsabbagh, Hill, Id, and Hill (2014) did not observe any ASD-specific patterns
Chapter 1 | General introduction

Chapter 1 | General introduction

of motor delay during the first two years of life. Possibly, motor delays are a more general indicator of atypical development, emerge with age, or may be specific to a subgroup of children with ASD (Landa, Gross, Stuart, & Bauman, 2012).

The diagnostic evaluation of young children with a suspicion of ASD currently relies on these early clinical manifestations that include both the absence of typical behaviors and the presence of atypical behaviors, all of which currently start to emerge around 12 months of age. See Box 1.3 for a brief overview of early signs of ASD that can be observed during the (start of) the second year (Dietz, 2007). Although ASD can be reliably diagnosed as early as 24 months of age, in many children diagnoses are given much later, as is also reflected in the case description of Thomas. More research into the early markers of ASD, especially during the first years of life, may help to improve the early detection of ASD, which is one of the aims of this thesis.

Etiology

ASD is a complex brain development disorder, involving early atypicalities in brain structure, brain connectivity, and brain functions (Varcin & Jeste, 2017). These deviations likely result from atypical regulation of multiple ontogenetic processes. During the last decade a growing body of research indeed found a predominant role of genes (Abrahams & Geschwind, 2008; Constantino & Charman, 2016; Geschwind & State, 2015). In about 10 to 25% of individuals a monogenic alteration is the cause of ASD, such as in tuberous sclerosis (TSC) or fragile X syndrome (Bourgeron, 2015). For the remaining 75 to 90%, ASD seem to be a multi-factorial condition. Both epidemiological (i.e. twin and adoption studies) and molecular genetic studies have offered some insights into possible causes.

First, twin studies have recently shown that there is 36 to 96% concordance for ASD in monozygotic twins versus 0 to 36% concordance in dizygotic twins (Hallmayer et al., 2011; Nordenbaek, Jorgensen, Kyvik, & Bilenberg, 2014; Ronald & Hoekstra, 2011; Sandin et al., 2014). In addition, research investigating siblings of children diagnosed

<table>
<thead>
<tr>
<th>Box 1.3 Red flags or signs of ASD (Dietz, 2007)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- No babbling by 12 months</td>
</tr>
<tr>
<td>- No interest in other people by 12 months</td>
</tr>
<tr>
<td>- No smiling at others by 12 months</td>
</tr>
<tr>
<td>- No reaction when spoken to by 12 months</td>
</tr>
<tr>
<td>- No gesturing by 12 months (pointing, waving bye bye)</td>
</tr>
<tr>
<td>- No functional use of single words by 18 months</td>
</tr>
<tr>
<td>- No 2-word spontaneous phrases (not echoic) by 24 months</td>
</tr>
<tr>
<td>- Any loss of any language or social skills at any age</td>
</tr>
</tbody>
</table>
with ASD showed that 18.7% of siblings is diagnosed with ASD themselves (Ozonoff et al., 2011), whereas 19% have some traits common to ASD, but do not obtain a clinical diagnosis (Georgiades, Szatmari, Zwaigenbaum, et al., 2013). These findings suggest high heritability associated with ASD, but the mode of inheritance is still unclear. Environmental factors (see below) may also influence the expression and severity of ASD symptoms, proposing that heritability only explains 50% of the risk for ASD (Sandin et al., 2014).

Second, molecular genetic studies suggest a complex architecture of multiple genes with a crucial role for both common and rare genetic risk factors that are associated with ASD susceptibility (Bourgeron, 2015). These common and rare genetic risk factors interact with each other and with environmental risk factors. Advances in gene discovery for ASD so far has resulted mainly from the investigation of rare mutations, indicating that a growing number of these mutations linked to ASD contain genes involved in neurotransmissions or synapse formation, neuronal migration, and neuronal plasticity (such as glutamatergic and GABAergic pathways) (Bourgeron, 2015; Miles, 2011). Remarkably, many of the rare variants have also been associated with other neuropsychiatric disorders (e.g. ADHD, epilepsy, intellectual disability |Kim & State, 2014), implying a lack of specificity to ASD.

Although these findings support the strong genetic basis for ASD, environmental factors may (in correlation and interaction with genes) influence risk and phenotypic presentation of ASD. Recent publications included environmental factors such as perinatal events (e.g. low birth weight, prematurity), immunization (maternal immune activation during pregnancy), biological environmental factors (e.g. preconceptual exposure to heavy metals, air pollution), advanced paternal and maternal age, and nutrients (e.g. shortage of vitamin D) (Modabbernia, Velthorst, & Reichenberg, 2017; Visser et al., 2013).

Overall, the heterogeneous phenotype of ASD likely reflects the many different genetic factors linked to ASD, the role of an individual’s genetic background, and the influence of non-genetic environmental factors (and the interaction between them). While the field of ASD genetics is evolving, its future role in the detection and treatment of ASD remains to be established. This thesis will not investigate the genotypic basis for ASD, but will focus on potential early risk markers for ASD.

Investigating early risk markers in high-risk studies
Parents’ recollections of their children’s behavior prior to diagnosis, and analyses of home videos, indicate that abnormalities in early development are present in the first years of life (Watson et al., 2007; Werner, Dawson, Osterling, & Dinno, 2000). Although these retrospective analyses have provided valuable insight into the early signs of ASD, it may also involve shortcomings, including a bias by what parents make available for study and variability in the content of home videos (e.g. different settings). Additionally, these studies may also
lack appropriate controls; that is, parents of non-autistic children have not been asked
about similar developmental concerns. The key to overcome these issues is by prospectively
studying the interaction between parents and infants who are at high risk for developing
ASD because they have an older sibling diagnosed with ASD (Szatmari et al., 2016). Given
the risk of recurrence of 18.7% in these siblings (Ozonoff et al., 2011), the design provides
unique opportunities to examine early characteristics of ASD before diagnosis is made. A
low-risk control group, composed of children that have no family history of ASD, is followed
in parallel. This allows to compare early development between and within high-risk (HR)
and low-risk (LR) siblings with diverse developmental outcomes. Although recent efforts
have been made to apply this design (see reviews Jones et al., 2014; Varcin & Jeste, 2017),
we do not yet have a full and accurate picture of the behavioral manifestations of ASD
in the first year(s) of life. The current thesis will specifically focus on two potential risk
markers: infant temperament and characteristics of parent-infant interaction, as reported
in Chapter 2 and 3 of this thesis.

**Temperament**

Temperament can be defined as relatively stable individual differences in activity, affectivity,
attention and self-regulation that are shaped throughout development by complex
interactions between genetic, biological and environmental factors (Shiner et al., 2012).
Most temperament frameworks encompass three traits during early childhood: (1) **surgency/approach** referring to engagement with the environment, positive emotions and activity level;
(2) **negative affect/withdrawal** including negative emotions such as anger, sadness and fear; and
(3) **effortful control** referring to regulation of attention, emotions and behaviors (Putnam, Ellis,
& Rothbart, 2001). The forerunner of effortful control in infancy that differs conceptually
from this construct at older ages is labeled orienting/regulation, focusing on soothability (pace
of recovery from distress) and cuddliness (expression of enjoyment and molding of the body
to the caregiver) (Gartstein & Rothbart, 2003). Given that these temperamental traits can
be linked to neurobiological systems (White, Lamm, Helfinstein, & Fox, 2012; Whittle, Allen,
Lubman, & Yucel, 2006) and are already measurable at an early age, potentially before ASD
begins to emerge, temperament could function as a potential early risk marker. Two models
in particular support the link between early differences in temperament and a subsequent
diagnosis of ASD. First, the **spectrum or common cause model** considers temperamental
traits as features along a spectrum of psychopathology (Nigg, 2006; Shiner & Caspi, 2003;
Tackett, 2006). This model implies a shared etiology between psychopathology at the extreme
negative end of a continuum of social-communicative competences and temperamental
traits. Second, the **vulnerability or pathoplasticity models** argue for temperament as a modifier
of the development of a disorder, either by increasing the likelihood of ASD (i.e. temperament
as a risk or protective factor) or by influencing the course or severity of ASD symptoms (Nigg,
2006; Tackett, 2006). This closely relates to the **modifier model** by Peter Mundy, which
states that temperament is a predictor of differences in ASD symptom presentation (Mundy, Henderson, Inge, & Coman, 2007; Schwartz et al., 2009). The models are not mutually exclusive and may all provide cues for understanding the role of temperament in the early emergence of ASD.

Previous research has indicated that temperament and psychopathology are consistently correlated across early childhood (Nigg, 2006), and ASD research suggests that children who are later diagnosed with ASD differ from typically developing infants from 12 months onward based on generally lower levels of surgency and effortful control, and higher levels of negative affect (e.g. Garon et al., 2016; Macari, Koller, Campbell, & Chawarska, 2017). However, previous studies have only performed statistical analyses to investigate temperamental differences between groups, but did not examine whether temperament could also inform outcomes at an individual level. A critical role for future research lies in the further investigation of trajectories of temperament across early childhood and the usefulness of early temperament in the individual prediction of ASD. The study described in Chapter 2 aimed to advance this area of research and reports on differences in parent-reported temperament during the first three years of life in a sample of familial HR and LR siblings (see Box 1.4 for details on the sample).

**Parent-infant interaction**

In the earliest months and years of a child’s life social interactions provide an important learning context. During this period the interaction with parents forms a key aspect of a child’s social environment. Interpersonal exchanges with the parent shape the young child’s developing social cognition, emergent language and communication skills, attentional control, and emotion regulation. The investigation of parent-infant interactions may therefore be helpful in understanding the developmental course of early social and communicative alterations, before ASD symptoms become clinically manifest (Wallace & Rogers, 2010). In this sense, early behavioral markers of ASD may become evident by observing very early parent-child dyads. Furthermore, the investigation of parent-infant interactions may also be helpful in improving early parent-mediated interventions. It is well established in typical development that parental sensitivity, verbal stimulation of the child’s speech, and reciprocity promote the child’s social and cognitive development (Feldman, Bamberger, & Kanat-Maymon, 2013; Page, Wilhelm, Gamble, & Card, 2010). In the context of ASD, research has also emphasized the role of parent-child interactions for children’s social communicative development (Harker, Ibanez, Nguyen, Messinger, & Stone, 2016; Haven, Manangan, Sparrow, & Wilson, 2014; McDuffie & Yoder, 2010; Siller & Sigman, 2002). There is evidence for the effectiveness of early parent-mediated interventions on the subsequent social-communicative and language development of
children with ASD (Kasari, Lawton, et al., 2014; Wetherby et al., 2014), highlighting the importance of optimizing parent-infant interactions.

Research into parent-child interaction can be placed in a transactional model of development, which focuses on gene x environment (GxE) interplay, with the child’s risk status as the genetic factor and parent-child interaction as the primary social environment (Mandy & Lai, 2016). The notion that cold or aloof parenting can in any simple sense cause autism, as was proposed by the so-called ‘refrigerator mother theory’ (Bettelheim, 1967), has been fortunately discredited. Nonetheless, characteristics of the parent-infant dyad may partially modify the relationship between ASD susceptibility and the development of an ASD phenotype later on. Pre-existing neurodevelopmental differences may bias the early social environment to which the infant is exposed, in turn altering that social environment, which in turn may increase the risk of the infant developing difficulties in social interaction and communication. For example, the infant is less attentive to the parent, leading to the parent being more intrusive, further affecting the infant’s response, and so on. In addition, in line with the differential susceptibility theory, infants later diagnosed with ASD may be differentially susceptible to less optimal early social environments, such that positive interactions with their parents are especially important for them (Belsky, Bakermans-Kranenburg, & van Ijzendoorn, 2007).

As research into early parent-child interaction within groups of infants at risk of ASD is relatively new, improved knowledge about how perturbations in early parent-infant interaction arise and how they change over time may help to improve early parent-mediated interventions. The bidirectional nature of interactions can be examined to dissect the contribution of the infant and the parent for determining targets of parent-mediated interventions for ASD (Kasari, Lawton, et al., 2014; Rogers et al., 2014; Wetherby et al., 2014). In addition, parent-infant interaction research may also help to determine targets for pre-emptive intervention (i.e. ‘prevention’ to redirect developmental trajectories before the full-blown disorder becomes manifest) for infants or toddlers at-risk (Green et al., 2017; Jones, Dawson, Kelly, Estes, & Jane Webb, 2017; Watson et al., 2017). In Chapter 3 a newly developed parent-infant coding scheme was applied on a group of HR and LR siblings during two time points in their first year of life (see Box 1.4 for details on the sample).

Early signs, late identification: enhancing early detection

The identification of early reliable (risk) markers for ASD has the potential to contribute to earlier detection and diagnosis. Currently, ASD can be reliably diagnosed before 3 years of age and the diagnosis is remarkably stable in both clinical (Kleinman et al., 2008; Lord et al., 2006; van Daalen et al., 2009) and high-risk samples (Brian et al., 2015; Ozonoff et al., 2015). However, despite a trend toward earlier diagnosis, the mean age at diagnosis in daily clinical practice still lies around the late preschool years or even later.
(Begeer, Van Wijngaarden, Vreugdenhil, & Wijnker-Holmes, 2017; Daniels & Mandell, 2014). Parental concerns specifically related to ASD can be expressed around the child’s first birthday (Ozonoff et al., 2009), but there is a significant delay between the first parental report of concerns and the diagnosis of ASD. A delay of 20 to 60 months between parental suspicion and ASD diagnosis (Sivberg, 2003), and a delay of 13 months between the first evaluation and ASD diagnosis have been reported (Wiggins, Baio, & Rice, 2006). This delay in obtaining an ASD diagnosis is undesirable, because it results in later access to early intervention services, which, in turn, may affect the child’s long-term developmental outcomes (see the following paragraph for further information about the advantages of early intervention). Additionally, many parents of children with ASD experience stress related to parenting a child with a ‘different’ development (Estes et al., 2009). The length of the diagnostic process and uncertainty of diagnosis may intensify these feelings and delay the introduction of parent support services (Keenan, Dillenburger, Doherty, Byrne, & Gallagher, 2010). Furthermore, parents are more satisfied with a diagnosis when they have seen fewer professionals to get the diagnosis and when their child receives the diagnosis at a younger age (Goin-Kochel, Mackintosh, & Myers, 2006). Thus, there is an urgent need for an effective strategy that facilitates timely detection of ASD, preferably before 3 years of age.

In order to lower the age at diagnosis, strategies for the early detection of ASD have been developed, including the application of screening tools (such as the Modified Checklist for Autism in Toddlers (M-CHAT) | Robins, Fein, Barton, & Green, 2001) and training of professionals in using these tools within a specific referral protocol (Magan-Maganto et al., 2017). Promising results of early detection programs have been obtained in previous non-randomized studies (Chakrabarti, Haubus, Dugmore, Orgill, & Devine, 2005; Holzer et al., 2006; Koegel, Nefdt, Koegel, Bruinsma, & Fredeen, 2006; Mazurek, Brown, Curran, & Sohl, 2017; Swanson et al., 2014) and in one randomized controlled study (Oosterling, Wensing, et al., 2010). However, long-term effects have been only investigated by Holzer et al. (2006), who found that the positive effect of the program on the short-term was not sustained after the program ended. Additional understanding of long-term effects can be helpful to develop effective programs that will have permanent effect. The study reported in Chapter 4 focused on the long-term outcomes of a clinically relevant early detection program in the Netherlands (see Box 1.4 for details on the sample).

**From detection to intervention: enhancing early intervention**

Once substantial improvement in the earlier detection of ASD has been established, follow-up services and empirically based early intervention programs should be available for children with (or at risk of) ASD and their families. There has been increased interest in the potential of early intervention to alter developmental trajectories in children
diagnosed with (or at risk of) ASD (French & Kennedy, 2017). Because of the dynamic and plastic nature of the brain, very early interventions may alter the course of brain and behavioral development in children with ASD (Webb, Jones, Kelly, & Dawson, 2014), and recent studies have provided evidence of the long-term effectiveness of early interventions (Estes, Munson, et al., 2015; Green et al., 2017; Pickles et al., 2016). Additionally, studies focusing on pre-emptive interventions for infants or toddlers at risk also showed positive changes in both parent and child behaviors (Green et al., 2017; Jones et al., 2017; Watson et al., 2017). However, there are many wide-ranging challenges to successfully compare between intervention studies, including a diversity of treatment approaches, disparity in dose of intervention, and heterogeneity of participants (French & Kennedy, 2017). An additional challenge is the variety in outcome measures and little consensus about which measures adequately monitor change over time. This compromises conclusions regarding the most effective type and intensity of intervention for children with ASD. Also, showing true effectiveness requires generalization of effects into wider social contexts (e.g. across interaction partners) and on core ASD symptoms (Green & Garg, 2018). Therefore, there is a critical need to explore what might be an ideal battery of outcome measures, especially including an assessment of the core social communication impairments in individuals diagnosed with ASD (Oono, Honey, & McConachie, 2013).

Currently, the Autism Diagnostic Observation Schedule (ADOS |Lord et al., 2012), a measure intended for diagnostic purposes, has often been used to measure treatment response in early intervention studies (e.g. Estes, Munson, et al., 2015; Green et al., 2010). Although this measure can identify changes in core ASD symptoms over several years (Estes, Munson, et al., 2015; Gotham, Pickles, & Lord, 2012), it is less effective in detecting subtle changes over shorter periods of time (Dawson et al., 2010; Estes, Munson, et al., 2015; Shumway et al., 2012; Thurman, Manwaring, Swineford, & Farmer, 2015). To circumvent this limitation and limitations of other commonly used outcome measures, a new measure has been developed: the Brief Observation of Social Communication Change (BOSCC). The BOSCC was developed to identify changes in social communicative behaviors over relatively short periods of time by calculating subtleties in both the frequency and quality of behaviors (Grzadzinski et al., 2016). The measure can be applied on naturalistic interactions between the child and an adult (e.g. parent and child or professional and child) and can be scored with naive coders. However, before the BOSCC can be used in intervention studies, its measurement properties, that is, reliability, validity, and responsiveness to change, need to be further established. Therefore, this will be the main focus of Chapter 5 in this thesis (see Box 1.4 for details on the sample).
Aims and outline of the thesis

The overall aims of the two sections in this thesis are:
1. To investigate early risk markers in infants and toddlers at familial high and low risk of ASD;
2. To study the long-term effects of a program for early detection of ASD and to evaluate the usefulness of an instrument in measuring change during early intervention of ASD.

Data in this thesis were collected from both clinical populations and populations based on familial risks of ASD. For details on the study samples, see Box 1.4.

Part I: Early risk markers of autism spectrum disorder in infancy and toddlerhood

To increase our knowledge about early risk markers of ASD, differences in behavioral characteristics were investigated between young infants at high risk of ASD and same-aged low-risk controls. In Chapter 2 we prospectively followed familial HR and LR siblings during their first three years of life and investigated differences in parent-reported temperament at and across three longitudinal time points in early childhood between HR-ASD, HR-Atypical, HR-Typical and LR siblings. By also using machine learning algorithms, the aim was to increase our understanding of the usefulness of early temperament to predict later ASD at an individual level. In the study reported in Chapter 3, parent-infant interaction was examined in HR and LR siblings during two time points in the first year of life. A new coding scheme, the Parent-Infant/Toddler Coding of Interaction (PInTCI), was developed, with the goal to provide a measure that includes constructs that were found to predict subsequent child development and can be used across early childhood in both HR and LR groups.

Part II: Early detection and intervention of autism spectrum disorder

As only one non-randomized study has examined long-term effects of an early detection program, the study described in Chapter 4 reports on the long-term outcomes of a clinically relevant early detection program. This program was integrated in routine developmental surveillance in the Netherlands and proved to be effective immediately after implementation (Oosterling, Wensing, et al., 2010). Chapter 5 focuses on the usefulness of the newly developed BOSCC to measure outcomes in early autism intervention studies. Given the limitation of the ADOS, we investigated whether the BOSCC has greater potential for measuring change in core autistic behavior than the ADOS.

The general discussion in Chapter 6 summarizes and discusses key findings, points out limitations, suggests directions for future research, and closes with implications for clinical practice.
Box 1.4 Study samples

The British Autism Study on Infant Siblings (BASIS | Chapter 2)
This sample consisted of 199 infants (133 HR and 66 LR) who were longitudinally assessed at four time points during their first three years of life (i.e. at 8, 14, 24, and 36 months). Data were collected by the BASIS team at the Centre for Brain and Cognitive Development (London, United Kingdom). Girls and boys were relatively equally represented in both risk groups (% HR male = 48.9; % LR male = 42.4%). The HR infants had at least one older sibling with a clinical diagnosis of ASD. No known other significant conditions were present in the proband or extended family members (e.g. Fragile X syndrome, tuberous sclerosis). LR siblings were recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. There was no ASD in first-degree family members of LR siblings (as confirmed through a parent interview regarding family medical history). Diagnostic outcome was established at 36 months using information from ADOS-2, Mullen Scales of Early Learning and the Vineland Adaptive Behavior Scale-II, and the infants were categorized in HR-ASD, HR-Atypical, HR-Typical, or LR groups.

The Eurosibs Autism Research network (Dutch: Zebra-project | Chapter 3)
This sample comprised 195 infants (113 HR and 82 LR) who were assessed at 5 and/or 10 months of age. Data were collected as part of the ongoing Eurosibs project at four different sites in Europe (Belgium, United Kingdom, Sweden, the Netherlands). Girls and boys were equally represented in both risk groups (% HR male = 53.1; % LR male = 51.2). The HR infants had at least one older sibling with a community clinical diagnosis of ASD. The LR siblings had at least one older sibling with typical development and no ASD within first-degree family members (as confirmed through a parent interview regarding family medical history). Exclusion criteria for both HR and LR infants included diagnosis of epilepsy, preterm birth, and genetic syndromes clearly related to ASD in infant or proband (e.g. fragile X syndrome, tuberous sclerosis).
Diagnosis and Intervention study on Autism in the Netherlands (DIANE | Chapter 4 and 5)

The sample described in Chapter 4 included 1235 infants, toddlers, and preschoolers who were referred to Karakter Child and Adolescent Psychiatry University Center for clinical psychiatric evaluation before (n=119), during (n=531), and after (n=585) an early detection program for ASD was implemented. Of these, 38%, 47%, and 37% were newly diagnosed with ASD as compared to non-ASD diagnoses, respectively. Children with a non-ASD diagnosis had other diagnoses (including absence of a psychiatric diagnosis). All children were between 0 and 6 years of age at the time of referral, with boys overrepresented before, during, and after the program (% male = 75.6; % male = 74.8; % male = 77.1, respectively).

The study reported in Chapter 5 re-used data collected by Oosterling, Visser, et al. (2010) who investigated the efficacy of a parent training intervention involving toddlers diagnosed with ASD. The sample included 44 children who were recruited during the period 2004–2007 at Karakter Child and Adolescent Psychiatry University Centre Nijmegen, the Netherlands. All children received some form of intervention, either the treatment condition in which parental skills were promoted or care as usual. Participants had screened positive on the Early Screening of Autistic Traits Questionnaire (ESAT - Swinkels et al., 2006) and met criteria for autism according to a consensus diagnosis made by a child psychiatrist and psychologist. All children were between 19 and 41 months of age, with boys overrepresented (% male = 77.3%).
PART I

Early risk markers of autism spectrum disorder in infancy and toddlerhood
Temperament as an early risk marker for autism spectrum disorders? A longitudinal study of high-risk and low-risk infants

Pijl M.K.J., Bussu G., Charman T., Johnson M.H., Jones E.J.H., Pasco G., Oosterling I.J., Rommelse N.J., Buitelaar J.K., and The BASIS Team*


1 Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, the Netherlands
2 Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, the Netherlands
3 King’s College London, Psychology Department, Institute of Psychiatry, Psychology & Neuroscience, London, United Kingdom
4 Centre for Brain and Cognitive Development, Birkbeck College, University of London, United Kingdom
5 Psychology Department, University of Cambridge, United Kingdom
6 Radboud University Medical Centre, Department of Psychiatry, Nijmegen, the Netherlands

Abstract

To investigate temperament as an early risk marker for autism spectrum disorder (ASD), we examined parent-reported temperament for high-risk (HR, n=170) and low-risk (LR, n=77) siblings at 8, 14, and 24 months. Diagnostic outcomes, categorized as HR-ASD, HR-Atypical, and HR-Typical, were established at 36 months using information from ADOS-2, Mullen Scales of Early Learning and the Vineland Adaptive Behavior Scale-II. Group-based analyses showed linear risk gradients, with more atypical temperament for HR-ASD, followed by HR-Atypical, HR-Typical, and LR siblings. Temperament differed significantly between outcome groups (0.03 ≤ $\eta_p^2$ ≤ 0.34). Machine learning analyses showed that, at an individual level, HR-ASD siblings could not be identified accurately, whereas HR infants without ASD could. Our results emphasize the discrepancy between group-based and individual-based predictions and suggest that while temperament does not facilitate early identification of ASD at an individual level, it may help identify HR infants who do not develop ASD.
Temperament can be defined as relatively stable individual differences in activity, affectivity, attention, and self-regulation that are shaped throughout development by complex interactions between genetic, biological, and environmental factors (Shiner et al., 2012). Given that temperament traits can be linked to neurobiological systems (White et al., 2012; Whittle et al., 2006) and are already measurable at an early age, potentially before psychopathology begins to emerge, temperament could function as a potential risk marker of later psychopathology (Fox, 2004; Nigg, 2006; Perez-Edgar & Fox, 2005). The aim of this study was to investigate temperament as an early risk marker for autism spectrum disorders (ASD) in the high-risk (HR) younger siblings of children diagnosed with ASD and low-risk (LR) controls. Research has shown that 18.7% of HR siblings are diagnosed with ASD themselves (Ozonoff et al., 2011), and that 19% of HR siblings have some traits common to ASD, but not sufficient to warrant a clinical diagnosis (Georgiades, Szatmari, Zwaigenbaum, et al., 2013). By applying a HR design, shared and unique characteristics of temperament between and within familial HR siblings (diagnosed with ASD, atypically developing, or typically developing) and LR siblings can be studied to reveal possible early predictors of later ASD or atypical development.

Most temperament frameworks encompass three traits during early childhood: (1) surgency/approach referring to engagement with the environment, positive emotions, and activity level; (2) negative affect/withdrawal including negative emotions such as anger, sadness, and fear; and (3) effortful control referring to regulation of attention, emotions, and behaviors (Putnam et al., 2001). In infancy, effortful control is described as orienting/regulation, focusing on soothability (pace of recovery from distress) and cuddliness (expression of enjoyment and molding of the body to the caregiver) (Gartstein & Rothbart, 2003). In the current study, we refer to this construct as effortful control in both infancy and toddlerhood.

Previous research has revealed that these three broader traits can differentiate children with ASD from others from 12 months onward (see Table 2.1). First, low levels of the trait surgency (i.e. approach behaviors, positive affect, and activity level) have been associated with later ASD (Del Rosario, Gillespie-Lynch, Johnson, Sigman, & Hutman, 2014; Garon et al., 2009; Garon et al., 2016; Macari et al., 2017; Zwaigenbaum et al., 2005). However, findings up to one year are discrepant, showing that HR siblings that develop ASD have higher levels of surgency than high-risk siblings who do not develop ASD (Clifford et al., 2013; Del Rosario et al., 2014). This discrepancy suggests that temperamental patterns change with development, but could also reflect differences in the applied construct of surgency across age and as used in different temperament measures. In-depth examination at a dimensional level showed contrasting patterns for activity levels, with lower levels of activity being seen in infants with (or at risk of) ASD during the first year (Bolton, Golding,
Emond, & Steer, 2012; Del Rosario et al., 2014; Zwaigenbaum et al., 2005), followed by elevated levels of activity around the second year (Bolton et al., 2012; Garon et al., 2009). Second, higher levels of the temperament trait negative affect have been consistently associated with ASD from 12 months onward (Bolton et al., 2012; Clifford et al., 2013; Garon et al., 2009; Garon et al., 2016; Macari et al., 2017; Zwaigenbaum et al., 2005). Lastly, children with ASD have more self-regulatory difficulties (effortful control) from around the first birthday onward (Bolton et al., 2012; Clifford et al., 2013; Garon et al., 2009; Garon et al., 2016; Gomez & Baird, 2005; Macari et al., 2017; Zwaigenbaum et al., 2005). However, Del Rosario et al. (2014) did not find any differences in negative affect or effortful control between HR-ASD and LR siblings during early childhood, which could be due to the use of different instruments to assess temperament. See Table 2.1 for a detailed overview of the abovementioned studies focusing on temperament traits in ASD.

Most of the abovementioned studies focused on differences in distinct temperament traits at separate time points (e.g. the level of surgency at 12 months) and did not integrate findings across various traits and time. To the best of our knowledge, only two studies investigated the time course of temperament in young children at risk of ASD (Del Rosario et al., 2014; Garon et al., 2016). The investigation of trajectories of temperament across multiple time points is potentially more informative than measures of temperament at single time points, because it provides information about the structure of change across early childhood. In addition, investigating the integration of different temperament traits at different time points could help to combine complementary information across traits. Furthermore, while previous studies investigated temperamental differences between groups, they did not examine whether temperament provides information about individual outcomes. Although findings on group differences are valuable in terms of finding relevant biomarkers for ASD, there is often substantial overlap between groups in individual variation, making prediction for individual infants difficult. To fully judge whether temperament is useful in the early prediction of ASD, analyses at an individual level are needed.

The current study prospectively followed familial HR and LR siblings during their first 3 years of life, with the aim of observing differences in temperament between outcome groups. For these outcome groups, the HR group was divided into HR-ASD (HR siblings subsequently diagnosed with ASD at 36 months), HR-Atypical (HR siblings not diagnosed with ASD, but with some evidence of atypical development) and HR-Typical siblings (HR siblings with typical development). The objectives were 1) to investigate group differences in early temperament at and across multiple time points between HR-ASD, HR-Atypical, HR-Typical, and LR siblings, and 2) to examine whether temperament (both single traits and profiles) during the first 2
years of life (both separate time points and trajectories) can help to predict ASD and atypical development at 36 months at an individual level. For the latter objective we extended previous work by using machine learning algorithms to combine complementary information about different temperament factors in order to identify the best predictive combination of factors. We expected that trajectories of temperament would differentiate between outcome groups and that the integration of different domains of temperament measured at different time points would improve the prediction of ASD in an individual as compared to prediction based on a single domain and/or time point. Further, based on their risk status, we hypothesized that HR-ASD would show the most ‘atypical’ temperament (i.e. low surgency, high negative affect, low effortful control), followed by HR-Atypical siblings, HR-Typical siblings, and LR siblings.

Table 2.1. Summary of findings on the three temperament traits and/or dimensions related to the traits in infants and toddlers with (or at risk of) ASD.

<table>
<thead>
<tr>
<th>Study</th>
<th>Participant description (N)</th>
<th>0-11 months</th>
<th>1-2 years</th>
<th>2-3 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SU</td>
<td>NA</td>
<td>EC</td>
</tr>
<tr>
<td>Clifford et al., 2013</td>
<td>HR-ASD (17), HR-Atypical (12), HR-Typical (24), LR (50)</td>
<td>↑</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Del Rosario et al., 2014</td>
<td>HR-ASD (10-16), HR-non ASD (7-27)</td>
<td>↑</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Garon et al., 2009</td>
<td>HR-ASD (34), HR-non ASD (104), LR (73)</td>
<td>↑</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Garon et al., 2016</td>
<td>HR-ASD (98), HR-non ASD (285), LR (162)</td>
<td>↑</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Zwaigenbaum et al., 2005</td>
<td>HR-ASD (19), HR-non ASD (46), LR (23)</td>
<td>↑</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Gomez &amp; Baird, 2005</td>
<td>ASD (65), TD (120)</td>
<td>↑</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Bolton et al., 2012</td>
<td>ASD (85), non-ASD (13885)</td>
<td>↑</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Macari et al., 2017</td>
<td>ASD (165), DD (58), TD (92)</td>
<td>↑</td>
<td>ns</td>
<td>ns</td>
</tr>
</tbody>
</table>

Notes. ASD = infants or toddlers diagnosed with autism spectrum disorders; DD = developmentally delayed infants or toddlers; EC = Effortful Control; HR-ASD = at-risk siblings subsequently diagnosed with ASD; HR-Atypical = at-risk siblings not diagnosed with ASD, but following an atypical development; HR-Typical = at-risk siblings following a typical development; LR = low-risk controls; NA = Negative Affect; SU = Surgency; TD = typically developing infants or toddlers.

Marked cells indicate findings based on the temperament trait’s composite score instead of findings based on dimensions or constructs related to the broader trait. Dimensions or constructs that could not be related to one of the three traits were not included in this table. Empty cells not investigated.

1 HR-ASD as compared to HR-Typical; 2 HR-ASD as compared to LR; 3 HR-ASD as compared to LR and HR-Atypical; 4 HR-ASD as compared to HR non-ASD and LR; 5 HR-ASD as compared to HR non-ASD; 6 HR (HR-ASD and HR non-ASD) as compared to LR; 7 Findings reported here are controlled for gender; 8 ASD as compared to both DD and TD; 9 ASD as compared to TD.
Chapter 2 | Temperament as an early risk marker for ASD

Methods

Participants and procedure
As part of the British Autism Study on Infant Siblings (BASIS: www.basisnetwork.org), 247 infants (170 HR and 77 LR) were assessed at four time points during their first three years of life. Data for 104 infants were collected during the first phase of the longitudinal study, which were also reported by Clifford et al. (2013). Ethical approval was given by NHS NRES London RC (06/MRE02/73, 08/H0718/76), and one or both parents gave informed consent. Most of the infants were born full-term (i.e. N=236 were born ≥36 weeks, N=11 were born between 32 and 36 weeks) and none of them had known medical or developmental conditions at the time they were enrolled. The HR infants had at least one older sibling with a clinical diagnosis of ASD [hereafter: ‘proband’], confirmed in most cases by expert clinicians using information from the Development and Wellbeing Assessment (DAWBA - Goodman, Ford, Richards, Gatward, & Meltzer, 2000) and the Social Communication Questionnaire (SCQ - Rutter, Bailey, & Lord, 2003). No known other significant conditions were present in the proband or extended family members (e.g. Fragile X syndrome, tuberous sclerosis). LR siblings were recruited from a volunteer database at the Birkbeck Centre for Brain and Cognitive Development. There was no ASD in first-degree family members of LR siblings (as confirmed through a parent interview regarding family medical history).

Of 247 siblings recruited, data for 33 HR and 9 LR siblings were excluded from the current study because of a substantial amount of missing data. Further information about this exclusion criterion is presented in the Measures section. We also excluded infants with no information about outcome status (N=4 HR, N=2 LR). The final sample comprised 133 HR infants (65 male; 48.9%) and 66 LR infants (28 male; 42.4%). All infants were examined at approximately 8 months (mean=8.4, SD=1.3, hereafter 8 months), 14 months (mean=14.8, SD=1.4, hereafter 14 months), around their second birthday (mean=25.4, SD=1.9, hereafter 24 months), and around their third birthday (mean=38.6, SD=2.2, hereafter 36 months).

Measures

Infant and toddler temperament. Two measures of temperament, appropriate to the child’s age, were administered. Parents completed the Infant Behavior Questionnaire-Revised (IBQ-R - Gartstein & Rothbart, 2003) at the 8- and 14-month visits, and the Early Childhood Behavior Questionnaire (ECBQ - Putnam, Gartstein, & Rothbart, 2006) at the 24-month visit. Both measures are reliable and well-validated parent-reported questionnaires that are scored on a Likert scale ranging from 1 (never) to 7 (always). The IBQ-R was designed to assess temperament in the first year of life and
contains 14 dimensions based on 184 items. The ECBQ was developed for children aged 18 to 36 months and consists of 18 dimensions based on 201 items. Three broad factors can be identified with both the IBQ-R and the ECBQ: Surgency, Negative Affect, and Effortful Control (labeled ‘Orienting’ on the IBQ-R). Of note, although both the IBQ-R and ECBQ provide a similar 3-factor model, the loading on the factors is different. See Putnam et al. (2001) for a discussion of this structure of temperament.

To ensure the validity of the temperament measures, dimensions were only calculated if data on ≥70% of items were available. Similarly, factors were only computed if ≥70% of dimension scores were available. Given that this study focused on longitudinal trajectories of temperament at 8, 14, and 24 months, participants were only included if data on ≥70% of the factors were available across the three time points.

Table 2.2. Sample characterization (means and standard deviations) for low-risk siblings and subgroups of high-risk siblings.

<table>
<thead>
<tr>
<th></th>
<th>HR-ASD (N=24)</th>
<th>HR-Atypical (N=34)</th>
<th>HR-Typical (N=75)</th>
<th>LR (N=66)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex (% male)</strong></td>
<td></td>
<td></td>
<td></td>
<td>42.4°</td>
</tr>
<tr>
<td>8 months</td>
<td>75°</td>
<td>47.1</td>
<td>41.3 b</td>
<td></td>
</tr>
<tr>
<td>14 months</td>
<td>8.3 (1.4)</td>
<td>8.6 (1.0)</td>
<td>8.5 (1.3)</td>
<td>8.3 (1.4)</td>
</tr>
<tr>
<td>24 months</td>
<td>14.8 (1.6)</td>
<td>14.7 (1.4)</td>
<td>14.9 (1.3)</td>
<td>14.7 (1.3)</td>
</tr>
<tr>
<td>36 months</td>
<td>25.4 (2.8)</td>
<td>25.4 (2.1)</td>
<td>26.0 (1.9) °</td>
<td>24.7 (1.0)</td>
</tr>
<tr>
<td></td>
<td>38.0 (2.0)</td>
<td>38.0 (2.8)</td>
<td>38.5 (1.8)</td>
<td>38.4 (2.7)</td>
</tr>
<tr>
<td><strong>MSEL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 months</td>
<td>98.0 (15.5) a</td>
<td>100.0 (13.8)</td>
<td>106.3 (15.8)</td>
<td>107.7 (12.6)</td>
</tr>
<tr>
<td>14 months</td>
<td>89.8 (17.3) a</td>
<td>96.5 (14.0) a</td>
<td>99.8 (14.6) a</td>
<td>106.0 (15.0)</td>
</tr>
<tr>
<td>24 months</td>
<td>94.5 (24.8) a</td>
<td>99.2 (21.8) a</td>
<td>104.9 (15.9) a</td>
<td>115.4 (14.2)</td>
</tr>
<tr>
<td>36 months</td>
<td>98.0 (26.7) a</td>
<td>95.9 (24.4) a</td>
<td>115.1 (15.5) a</td>
<td>118.1 (15.0)</td>
</tr>
<tr>
<td><strong>ADOS severity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36 months</td>
<td>5.1 (3.0) a</td>
<td>5.1 (2.2) a</td>
<td>1.5 (0.9) b</td>
<td>2.5 (1.8)</td>
</tr>
</tbody>
</table>

Notes. Superscripted letters that differ from other superscripted letters indicate significant differences across groups for the given measure (p≤0.05). Values without superscript letters indicate no significant differences from another group.

1 Mullen Scales of Early Learning (Mullen, 1995) Early Learning Composite Standard Score
2 Autism Diagnostic Observation Schedule-2 (ADOS-2 | Lord et al., 2012)
3 ADOS-2 calibrated severity score (Gotham, Pickles, & Lord, 2009)
Outcome characterization. At the 36-month visit, various clinical research measures were used to characterize the outcome of the HR siblings. The Autism Diagnostic Observation Schedule (ADOS-2 |Lord et al., 2012), the Autism Diagnostic Interview (ADI-R |Rutter, Le Couteur, & Lord, 2003), and the SCQ (Rutter, Bailey, et al., 2003) were used to obtain information about ASD symptomatology. In addition, the Mullen Scales of Early Learning (Mullen, 1995) and the Vineland Adaptive Behavior Scale-II (Sparrow, Balla., Cicchetti, & Doll, 2005) were assessed to gather information about the child’s development and adaptive functioning, respectively. Experienced clinical researchers (TC, GP) reviewed the outcomes of each HR sibling. Consensus ICD-10 or DSM-5 criteria were used to ascertain ASD diagnostic outcome. Among the 133 HR siblings enrolled in this study, 24 HR siblings met criteria for ASD [hereafter: ‘HR-ASD’] and 34 HR siblings did not meet criteria for ASD, but scored above the ASD threshold on the ADOS and/or ADI-R and/or scored >1.5 SD below the population mean on the MSEL receptive language, expressive language, and/or early learning composite score [hereafter: ‘HR-Atypical’]. The remaining 75 HR siblings were considered to be developing typically [hereafter: ‘HR-Typical’]. No formal research diagnoses were assigned to the LR group, but none of the LR infants had a community clinical ASD diagnosis. See Table 2.2 for detailed demographics of the included participants.

Statistical analyses
Multiple imputation with the expectation maximization algorithm was used to account for missing data (Tabachnik & Fidell, 2001). In addition, a Van der Waerden transformation was applied to data for temperament factors, which transforms raw scores into z-scores corresponding to the estimated cumulative proportion of the distribution analogous to a particular rank (using Statistical Package for the Social Sciences [SPSS] version 22).

Group-based analyses. MANCOVAs were used to investigate whether a risk gradient was present in polynomial group contrasts at separate time points. The outcome groups were ranked as follows: 1=HR-ASD, 2=HR-Atypical, 3=HR-Typical, and 4=LR, assuming that polynomial group contrasts would indicate linear risk gradients for atypical temperament (HR-ASD > HR-Atypical > HR-Typical > LR). Analyses were performed for each temperament trait separately, including group as independent variable and temperament at three time points as dependent variables (e.g. surgency at 8, 14, and 24 months). Sex was differently distributed across groups (with more males than females in the HR-ASD group), and age at intake was variable (between 5 and 11 months), introducing potential noise in results due to different starting ages. Therefore, sex and age at the first visit were included as covariates.
In post-hoc analyses, pair wise group contrasts were examined across time by performing two-way mixed ANCOVAs and paired sample t-tests, resulting in six pair wise comparisons (i.e. HR-ASD vs. HR-Atypical, HR-ASD vs. HR-Typical, HR-ASD vs. LR, HR-Atypical vs. HR-Typical, HR-Atypical vs. LR, HR-Typical vs. LR). The effect of group (e.g. HR-ASD, LR), time (8, 14, 24 months), and the interaction effect group x time on trajectories of a temperament trait was investigated, while controlling for sex and age at the first visit. A correction for multiple comparisons was applied for the post-hoc analyses, using the false discovery rate controlling procedure with a q-value of 0.05 (Benjamini & Hochberg, 1995). If Mauchly’s test indicated that the assumption of sphericity had been violated, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Following Cohen’s guidelines (Cohen, 1988), effect sizes were defined in terms of the percentage of variance explained: 1, 9 and 25% were used to define small, medium, and large effects (these percentages translate into \( \eta^2 \)-values of 0.01, 0.06 and 0.14). Analyses contrasting the HR group (without a differentiation based on 36-month outcome) and the LR controls are described in Supplemental Material.

Classifier Analyses – from group-based to individual analysis. As a next step, we investigated how temperament factors at 8, 14, and 24 months related to atypical development, and more specifically ASD, at an individual level among infants in the HR group. To this end, we performed confounder-corrected support vector machine classification with 40% holdout cross-validation repeated 10 times using custom made scripts implemented in Matlab R2016b (MATLAB 9.1, The MathWorks Inc., Natick, MA, 2016). We addressed two distinct binary classification problems: distinguishing HR-ASD from HR-Atypical and HR-Typical; and distinguishing HR-ASD and HR-Atypical together from HR-Typical. In fact, while the most clinically relevant question is to distinguish HR-ASD siblings from their peers at an early age, distinguishing HR-ASD and HR-Atypical together from HR-Typical is also clinically relevant and potentially useful for early intervention. Sex and age at the first visit were included as covariates, and findings were corrected for inverse probability weighting. Features for the classifiers consisted of temperament factors (surgency, negative affect, effortful control, and all their combinations) from different time points (8 months, 14 months, and 24 months). To exploit the longitudinal information on developmental dynamics, the intercept and slope of the developmental trajectories on single measures between 8 and 24 months were also used as features for the classifiers. Trajectories were computed for single individuals by linear regression modeling using the lme4 software package on R (Bates, Machler, Bolker, & Walker, 2015). A total of 28 classifiers were compared to find the best predictor of HR-ASD and HR-ASD+HR-Atypical at 36 months (see Supplemental Material for details).
For each classifier, the area under the curve (AUC) was computed to determine the best classifier, and we evaluated the classifier performance via sensitivity, specificity, accuracy, negative predictive value (NPV – i.e. true negative over negative predicted cases), and positive predictive value (PPV – i.e. true positive over positive predicted cases). 95% confidence intervals (CI) for each metric were computed using bootstrap with n=1000 repetitions for each cross-validation fold, then averaging over folds (n=10000 in total). The p-value of AUC was computed for each classifier through a shuffle test (n=10000 total repetitions; n=1000 repetitions for each classification fold) to test the significance of classification performance. Performance metrics are reported only when the performance was significantly different from chance level.

For both classifications (HR-ASD vs. HR-Atypical + HR-Typical | HR-ASD + HR-Atypical vs. HR-Typical), the best predicting classifier at each time point was selected based on the AUC. A nonparametric Friedman test was performed on classifier performance metrics (i.e. AUC) at each time point separately to test for significant differences in performance between distinct classifiers. If the Friedman test was significant, post-hoc paired Wilcoxon rank sum tests were performed between the classifier of interest (i.e. the one with highest AUC) and all other classifiers. Bonferroni correction was applied to avoid biasing effects due to multiple comparisons. In addition, differences in performance of the best classifiers across time points were tested by a two-sided Wilcoxon rank sum test.

Results

Temperament differences between groups
Surgency. A polynomial group contrast indicated a linear risk gradient to be present at 14 months of age (Contrast Estimate [CE]=0.40, p=0.02), implying that LR siblings had the highest levels of surgency (i.e. approach behaviors, positive affect, activity level), followed by HR-Typical siblings, HR-Atypical siblings, and HR-ASD siblings. No significant gradient was present at 8 or 24 months (CE=-0.08, p=0.64; CE=0.27, p=0.10, respectively).

Two-way mixed ANCOVAs examining pair wise group contrasts revealed a significant group x time effect for the comparison between HR-ASD and HR-Typical siblings (F(1.77, 168.07)=3.67, p<0.05, η²=0.04; see Figure 2.1), as well as between HR-ASD and LR siblings (F(1.86, 160.06)=3.98, p<0.05, η²=0.04). Post-hoc tests revealed that both interaction effects were driven by a group x time effect between 8 and 14 months of age (F(1, 95)=6.69, p<0.05, η²=0.07; F(1, 86)=9.79, p<0.01, η²=0.10, respectively), with HR-ASD siblings showing diverging levels of surgency from 8 to 14 months compared with HR-Typical and LR siblings (paired sample t-tests for each group...
were non-significant). In addition, for the comparison between HR-ASD and LR siblings a significant main effect of group was found between the 14- and 24-month time point \((F(1, 86)=4.89, \ p<0.05, \ \eta^2=0.05)\), indicating stable lower levels of surgency in the HR-ASD group than in the LR group between 14 and 24 months of age.

**Negative Affect.** A polynomial group contrast indicated a linear risk gradient to be present at 8, 14, and 24 months \((CE=-0.46, \ p=0.004; \ CE=-0.38, \ p=0.02; \ CE=-0.69, \ p<0.001, \) respectively), suggesting that HR-ASD siblings showed the highest levels of negative affect, followed by HR-Atypical siblings, HR-Typical siblings and LR siblings. A two-way mixed ANCOVA revealed significant main group effects for HR-ASD vs. HR-Typical \((F(1, 95)=7.47, \ p<0.01, \ \eta^2=0.07; \) see Figure 2.2), HR-ASD vs. LR \((F(1, 86)=15.57, \ p<0.001, \ \eta^2=0.15)\), and HR-Typical vs. LR siblings \((F(1, 137)=6.49, \ p<0.05, \ \eta^2=0.05)\). These effects indicate that, independent of age, HR-ASD siblings had developmentally stable higher levels of negative affect than HR-Typical and LR siblings, and that HR-Typical siblings had stable higher levels of negative affect than LR siblings.

**Effortful Control.** A polynomial group contrast indicated a linear risk gradient to be present at 14 and 24 months \((CE=0.69, \ p<0.001; \ CE=0.84, \ p<0.001, \) respectively), showing that LR siblings had the highest levels of effortful control, followed by HR-Typical siblings, HR-Atypical siblings and HR-ASD siblings. No significant gradient was present at 8 months of age \((CE=0.21, \ p=0.20)\).

A two-way mixed ANCOVA showed significant group x time interaction effects for the comparisons between HR-ASD and HR-Typical siblings \((F(1.85, 175.79)=6.95, \ p<0.01, \ \eta^2=0.07; \) see Figure 2.3), and between HR-ASD and LR siblings \((F(2, 172)=8.41, \ p<0.001, \ \eta^2=0.09)\). Post-hoc tests revealed that the interaction effects were driven by the 8- to 14-month trajectory \((F(1, 95)=8.53, \ p<0.01, \ \eta^2=0.08; \) \(F(1, 86)=12.69, \ p<0.01, \ \eta^2=0.13, \) respectively), showing that the level of effortful control decreased in HR-ASD siblings from 8 to 14 months \((t(23)=2.85, \ p=0.009)\) relative to the static levels of effortful control seen in HR-Typical \((t(74)=1.08, \ p=0.28)\) and LR \((t(65)=-1.03, \ p=0.31)\) siblings. Between 14 and 24 months of age, significant main effects of group were found \((F(1, 86)=18.90, \ p<0.001, \ \eta^2=0.17; \) \(F(1, 86)=44.22, \ p<0.001, \ \eta^2=0.34, \) respectively), suggesting that HR-ASD siblings had stable lower levels of effortful control than HR-Typical and LR siblings. Furthermore, significant main group effects were found between HR-ASD vs. HR-Atypical \((F(1, 54)=6.28, \ p<0.05, \ \eta^2=0.10)\), HR-Typical vs. LR \((F(1, 137)=4.31, \ p<0.05, \ \eta^2=0.03)\), and HR-Atypical vs. LR \((F(1, 96)=5.19, \ p<0.05, \ \eta^2=0.05)\) siblings. These results showed that HR-ASD siblings had developmentally stable lower levels of effortful control than HR-Atypical siblings, and that LR controls had higher levels of effortful control than both HR-Typical and HR-Atypical siblings.
A significant group x time effect was found for the comparison between HR-ASD and HR-Typical siblings, and between HR-ASD and LR siblings. Post-hoc tests revealed that both interaction effects were driven by a group x time effect between 8 and 14 months of age. Additionally, for the comparison between HR-ASD and LR siblings a significant main effect of group was found between the 14- and the 24-month time point. For details see text.

Figure 2.1. Estimated Means for Surgency by Diagnostic Group and Time controlled for Sex and Age at Start.

Significant main effects of group were found for HR-ASD vs. HR-Typical, HR-ASD vs. LR, and HR-Typical vs. LR siblings. For details see text.

Figure 2.2. Estimated Means for Negative Affect by Diagnostic Group and Time controlled for Sex and Age at Start.
Legend: This figure shows the estimated means for effortful control with error bars representing standard errors. Significant group x time effects were found for the comparisons between HR-ASD and HR-Typical siblings, and between HR-ASD and LR siblings. Post-hoc tests revealed that both interaction effects were driven by a group x time effect between 8 and 14 months of age. Significant main effects of group were found between the 14- and 24-month time point for both comparisons. Additionally, significant main effects of group were found for HR-ASD vs. HR-Atypical, HR-Typical vs. LR, HR-Atypical vs. LR siblings. For details see text.

**Figure 2.3.** Estimated Means for Effortful Control by Diagnostic Group and Time controlled for Sex and Age at Start.

**Individual prediction of HR clinical outcome**

Classification of HR-ASD among HR siblings was significantly different from chance level using measures from 14 months onward. In contrast, classification of HR-ASD and HR-Atypical together from HR-Typical was not significantly different from chance level at any of the time points, with only marginal significance at 24 months. See Table 2.3 and 2.4 for an overview of the performance metrics of classifiers that were significantly different from chance level for the two classifications (i.e. HR-ASD vs. HR-Atypical + HR-Typical, and HR-ASD+HR-Atypical vs. HR-Typical). Detailed statistics can be found in the Supplemental Material.

To evaluate which combination of temperament factors best predicted ASD at different time points, we compared the performance of the different classifiers at the separate time points, based on the AUC. The combination of all factors at 24 months provided the most promising classifier to predict ASD among HR siblings ($p=0.02$; mean [CI]: AUC=72% [57% to 83%]; sensitivity=85% [61% to 99%], specificity=58% [43% to 73%], accuracy=63% [49% to 75%], PPV=30% [13% to 49%], NPV=95% [86% to 100%]). However, the predictive performance was
not significantly different from that of effortful control \( (z=-0.51, p=0.61) \) and its combination with other factors at 24 months \( (z=-1.58, p=0.11) \). Furthermore, effortful control had the highest predictive power at 14 months \( (\text{AUC}=64\%) \), and when using the developmental trajectory between 8 and 24 months as feature for the classifiers, the integration of scores from effortful control and negative affect provided the classifier with the highest AUC \( (\text{AUC}=68\%) \). After Bonferroni correction for multiple comparisons \( (\alpha_{\text{Bonferroni}}=0.017) \), the difference in classification performance between the combined factors at 24 months and effortful control at 14 months was not significant \( (\text{Wilcoxon } z=-2.14, p=0.032) \), and the same applies to the difference in classification performance between the combined factors at 24 months and the combined longitudinal trajectories of effortful control and negative affect \( (\text{Wilcoxon } z=-1.86, p=0.063) \).

For classification of HR-ASD plus HR-Atypical from HR-Typical, the integration of effortful control and negative affect at 24 months provided the highest AUC \( (p=0.056; \text{mean [CI]}: \text{AUC}=61\% [48\% to 74\%]; \text{sensitivity}=60\% [39\% to 79\%], \text{specificity}=62\% [45\% to 79\%], \text{accuracy}=61\% [48\% to 74\%], \text{PPV}=55\% [35\% to 75\%], \text{NPV}=68\% [50\% to 85\%]). \) Since performance was not significantly different from chance level, classifier comparison was not performed.

Overall, even though effortful control and a combination of the temperament factors at 24 months predicted ASD outcome at a moderate level \( (\text{AUC}=71\%; \text{AUC}=72\%, \text{respectively}) \), its positive predictive value for ASD was low and none of the classifiers adequately predicted broader atypical development at 36 months.
### Table 2.3. HR-ASD vs. HR-Typical + HR-Atypical

Performance metrics of temperament factors for classifying HR siblings who later develop ASD from their peers.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>p</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effortful control</td>
<td>0.047</td>
<td>64.4</td>
<td>[46.7, 80.3]</td>
<td>69.6</td>
<td>[37.4, 97.8]</td>
<td>59.2</td>
<td>[44.5, 73.4]</td>
</tr>
<tr>
<td>24 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effortful control</td>
<td>0.006</td>
<td>71.4</td>
<td>[57.0, 82.6]</td>
<td>88.0</td>
<td>[66.2, 100]</td>
<td>54.8</td>
<td>[39.8, 69.0]</td>
</tr>
<tr>
<td>Surgency + Effortful control</td>
<td>0.021</td>
<td>66.7</td>
<td>[49.6, 80.8]</td>
<td>79.3</td>
<td>[49.3, 100]</td>
<td>54.1</td>
<td>[39.1, 68.8]</td>
</tr>
<tr>
<td>Effortful control + Negative affect</td>
<td>0.031</td>
<td>69.9</td>
<td>[55.0, 82.6]</td>
<td>82.6</td>
<td>[57.0, 98.9]</td>
<td>57.3</td>
<td>[42.4, 71.8]</td>
</tr>
<tr>
<td>All factors</td>
<td>0.020</td>
<td>71.5</td>
<td>[57.1, 83.5]</td>
<td>84.8</td>
<td>[60.5, 98.9]</td>
<td>58.2</td>
<td>[43.4, 72.6]</td>
</tr>
<tr>
<td>Longitudinal trajectory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effortful control</td>
<td>0.042</td>
<td>66.4</td>
<td>[49.2, 80.3]</td>
<td>79.3</td>
<td>[48.9, 100]</td>
<td>53.5</td>
<td>[38.9, 67.9]</td>
</tr>
<tr>
<td>Effortful control + Negative affect</td>
<td>0.013</td>
<td>67.6</td>
<td>[50.5, 81.7]</td>
<td>77.3</td>
<td>[47.3, 99.1]</td>
<td>57.8</td>
<td>[42.8, 72.1]</td>
</tr>
<tr>
<td>All factors</td>
<td>0.048</td>
<td>65.1</td>
<td>[48.1, 79.3]</td>
<td>75.2</td>
<td>[44.8, 97.9]</td>
<td>55.0</td>
<td>[40.2, 69.5]</td>
</tr>
</tbody>
</table>

Notes. All classifiers reported in this table significantly differed from prediction at chance level (shuffle test p<0.05). All metrics are reported as mean [95% confidence interval]. 95% confidence interval was computed using bootstrap. The classifiers are divided based on the data used as features: data collected at 14 months, data collected at 24 months, intercept and slope of the longitudinal trajectory between 8 and 24 months at the individual level. The abbreviations: AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.
Table 2.4. HR-ASD + HR-Atypical vs. HR-Typical. Performance metrics of temperament factors for classifying the HR atypical group as a whole (including atypically developing siblings and those who later develop ASD) from typically developing siblings.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>p</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgency + Effortful</td>
<td>0.058</td>
<td>60.6</td>
<td>63.7</td>
<td>57.6</td>
<td>60.2</td>
<td>52.7</td>
<td>68.2</td>
</tr>
<tr>
<td>control</td>
<td></td>
<td>[47.0, 73.5]</td>
<td>[43.0, 82.9]</td>
<td>[39.6, 74.7]</td>
<td>[46.8, 72.8]</td>
<td>[34.2, 70.9]</td>
<td>[49.1, 85.4]</td>
</tr>
<tr>
<td>Effortful control +</td>
<td>0.056</td>
<td>61.1</td>
<td>59.8</td>
<td>62.4</td>
<td>61.3</td>
<td>55.0</td>
<td>67.9</td>
</tr>
<tr>
<td>Negative affect</td>
<td></td>
<td>[48.0, 74.1]</td>
<td>[39.2, 79.2]</td>
<td>[45.4, 78.9]</td>
<td>[48.3, 74.3]</td>
<td>[34.2, 70.9]</td>
<td>[49.9, 84.7]</td>
</tr>
<tr>
<td>All factors</td>
<td>0.051</td>
<td>60.0</td>
<td>58.8</td>
<td>61.2</td>
<td>60.2</td>
<td>53.2</td>
<td>66.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[46.3, 72.9]</td>
<td>[37.7, 78.7]</td>
<td>[43.7, 77.6]</td>
<td>[47.0, 72.6]</td>
<td>[33.0, 72.0]</td>
<td>[48.5, 83.7]</td>
</tr>
</tbody>
</table>

Notes. None of the classifiers performed significantly different from chance level (shuffle test p<0.05). Here we report classifiers performing marginally different from random. All metrics are reported as mean [95% confidence interval]. 95% confidence interval was computed using bootstrap. Abbreviations: AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.
Discussion

The current study is the first to examine differences in temperament at and across three time points in early childhood between outcome groups (i.e. HR-ASD, HR-Atypical, HR-Typical and LR siblings), and to investigate temperament at an individual level. At a group level, our findings revealed positive linear risk gradients for surgency at 14 months, and effortful control at 14 and 24 months, and negative linear risk gradients for negative affect at 8, 14, and 24 months. This indicates that temperament in early childhood was more atypical in HR-ASD siblings, followed by HR-Atypical siblings, HR-Typical siblings, and LR controls. Post-hoc pair wise comparisons indicated differences in early temperament between the outcome groups. However, the effect sizes were generally medium, especially regarding differences within the HR group. Machine learning analyses using temperament traits during infancy (i.e. 8 months) did not accurately predict ASD at 36 months at an individual level. From 14 months onward, effortful control (or its combination with other traits) had the highest predictive power for ASD as compared to other temperament traits and combinations, with a high negative predictive value, but with a positive predictive value that was far from being clinically useful. Neither the separate temperament traits nor a combination of traits was able to accurately predict broader atypical development (i.e. HR-ASD and HR-Atypical). Thus, although differences in temperament traits can be detected in infancy at a group level, this difference does not necessarily translate into an acceptably accurate prediction of ASD in the individual infant.

Temperament differences between HR subgroups and LR controls

At a group level, our findings showed that HR siblings with or without a subsequent ASD diagnosis could be distinguished from LR controls based on higher levels of negative affect and lower levels of effortful control (with the exception of HR-Atypical siblings regarding negative affect). These findings replicate and extend previous research (Garon et al., 2016), showing that young siblings at risk of ASD, regardless of whether they develop ASD or not, tend to use more negative emotions and have more difficulties regulating attention, emotions, and behaviors than do LR controls. Furthermore, we found that the pattern of surgency from 8 to 14 months and levels of surgency thereafter were different between HR-ASD and LR siblings, whereas levels of surgency in the HR-Typical and HR-Atypical siblings did not differ from those of the LR group. As to be expected, this suggests that, on average, low levels of approach and positive emotions are specifically associated with the development of ASD. Differences in surgency levels across time may be explained by the multi-dimensional nature of the factor surgency (Gartstein & Rothbart, 2003; Putnam et
Future research may use a dimensional or item level approach to delineate the underlying mechanisms and to enable comparison of findings between studies.

**Temperament differences within at-risk siblings**

Within the HR group, temperament traits distinguished HR-ASD siblings from HR siblings without a clinical diagnosis, suggesting the presence of more temperamental challenges early in life of children with subsequent ASD. Interestingly, higher levels of negative affect were already present from 8 months onward in the HR-ASD siblings, whereas effortful control started to distinguish between the groups from 14 months onward. These findings, combined with those of a recent study examining temperament trajectories from 12 months onward (Garon et al., 2016), may indicate that early affective behaviors play an important role in the subsequent regulation of attention, emotions, and behaviors. Garon et al. (2016) found that affective components of temperament at 12 months predicted regulatory behaviors at 24 months in both HR and LR infants, and that regulatory behaviors in turn predicted ASD symptoms at 36 months in the HR sample. Future investigation of the associations between temperament traits in different outcome groups is needed, including the assessment of temperament during the first year of life.

**Temperament as a potential early risk marker**

The idea that temperament may be an early risk marker is in accordance with the spectrum theory (Tackett, 2006), which holds that there is a shared etiology between psychopathology at the extreme negative end of a continuum of social-communicative competences and temperament traits. A study of monozygotic and dizygotic adult twins supported this idea by showing that ASD and most temperament traits share common genetic and environmental etiological factors (Picardi et al., 2015). Temperament may be a fruitful risk marker that could help differentiate between groups of children on different developmental pathways.

Nonetheless, the use of temperament traits as an early risk marker is constrained by two findings. First, identification of ASD at an individual level on the basis of temperament traits had low positive predictive value and specificity. This indicates that based on (combinations of) temperament traits a substantial number of HR siblings would be falsely classified as HR-ASD at 36 months (i.e. false positives). However, the high negative predictive values indicate that temperament traits can accurately predict which infants are not going to develop ASD in all likelihood. This has still clinical value, especially for the selection of infants who might need early intervention. In other words, results at the individual level suggest that while low levels of effortful control do not predict ASD development, high levels of effortful control accurately predict typical development. The predictive value of effortful control for
non-ASD development is in line with the view that effortful control, as a measure of executive function, might promote resilience, such that infants with higher levels of effortful control may be better able to compensate for atypicalities that lead to ASD outcome (Johnson, 2012). However, our results highlight the difficulties of translating findings from a group to an individual level. In fact, there is often substantial overlap between groups in individual variation, making it more difficult to make predictions for individual infants. Instead of a risk marker for ASD, variation in temperament may therefore function as a *stratification marker* that allows to classify individuals with ASD into biologically more homogeneous subtypes (Loth et al., 2017). In this way, temperament may help to unravel the heterogeneous character of ASD. Importantly, the extent to which atypical temperament reflects brain alterations that predispose a child to developing ASD and/or are shared between atypical temperament and ASD need to be investigated. Additionally, future work should investigate the integration of clinical (e.g. MSEL, VABS, AOSI) and biological (e.g. eye tracking, functional imaging) measures, to improve the positive predictive value for the clinical diagnosis of ASD at an individual level (Bussu et al., 2018), and to investigate the additional value of temperament. Second, the differences found in this study mainly started to emerge around the first birthday (at both group and individual levels), which is also when behaviors related to ASD start to emerge (Ozonoff et al., 2010; Wan et al., 2013). This makes it important to ascertain whether temperament measures actually assess characteristics of temperament, or whether they just pick up the emergence of ASD symptoms. Future research should further investigate the conceptual nature of temperament measures by examining the structure of traits in different outcome groups and in relation to ASD severity.

**Limitations and Future Directions**

Particular strengths of this study are its longitudinal design, which allowed the assessment of temperament trajectories across early childhood, and the differentiation between siblings based on their diagnostic status at 36 months of age. A limitation is that temperament was assessed on the basis of parent-reported measures and not on observational measures of temperament (e.g. Lab-TAB; Gagne, Van Hulle, Aksan, Essex, & Goldsmith, 2011). It will therefore be essential to demonstrate convergence between the parent-reported IBQ-R and ECBQ and indicators of temperament based on standardized laboratory or home assessments. Nonetheless, evidence of convergent validity between a preliminary version of the IBQ and home observations of infant temperament implies that parents’ familiarity with a child’s behavior may make them the best possible source of reliable information (Rothbart, 1986). In addition, although early temperament has been found to be fairly stable in the general population (Casalin, Luyten, Vliegen, & Meurs, 2012), this study showed
that this may not apply similarly to young children at risk of ASD. There is a lack of research into the stability of temperament in children at risk of ASD, so the time x group interaction effects that were found in this study should be interpreted with caution. Future research should focus on the stability of temperament in children (at risk of or) diagnosed with ASD. Finally, given that temperament is the result of complex interactions between genetic, biological, and environmental factors (Shiner et al., 2012), the role of the environment, such as the child’s family, should also be considered in temperament research. Previous research has shown that the quality of parenting interacts with individual differences in genetic variation to influence temperament traits (Sheese, Voelker, Rothbart, & Posner, 2007; Voelker, Sheese, Rothbart, & Posner, 2009).

Conclusions
Taken together, our longitudinal study identified differences in early temperament traits between HR and LR siblings as well as between the different outcome subgroups among HR children, as most clearly demonstrated by differences in negative affect from 8 months onward and effortful control from 14 months onward. Our results underscore the complexity of translating findings from a group to an individual level, as findings did not accurately predict ASD at an individual level. From a clinical perspective, our results indicate that temperament traits may provide useful information about which HR infants are less likely to develop ASD but are not useful in predicting which HR infants will develop ASD or an atypical outcome. Future studies should increase our understanding of the role of temperament when it comes to individualizing interventions. Knowledge about temperament traits that influence adaptive functioning might help to improve the benefit of interventions in young children at risk of ASD.
Supplemental Material

HR versus LR comparison
Two-way mixed ANCOVAs were performed for each temperament trait, including the effect of group (HR, LR), time (8, 14, 24 months), and the interaction effect group x time, while controlling for sex and age at the first visit. A correction for multiple comparisons was applied for the post-hoc analyses, using the false discovery rate (FDR) controlling procedure with a q-value of 0.05 (Benjamini & Hochberg, 1995). If Mauchly’s test indicated that the assumption of sphericity had been violated, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. Results revealed no significant interaction or main effects for surgency ($p>0.05$), suggesting that HR siblings as a group do not differ from LR controls in levels of approach and positive emotions during their first years of life. Significant main group effects were found for both negative affect ($F(1, 195)=11.01, p<0.05$) and effortful control ($F(1, 195)=10.92, p<0.05$), indicating that HR siblings had developmentally stable higher levels of negative affect and lower levels of effortful control than LR siblings. See Figures S2.1-S2.3 for the trajectories of each temperament trait by risk group and time.

Figure S2.1. Estimated Means for Surgency by Risk Group and Time controlled for Sex and Age at Start
Figure S2.2. Estimated Means for Negative Affect by Risk Group and Time controlled for Sex and Age at Start

Figure S2.3. Estimated Means for Effortful Control by Risk Group and Time controlled for Sex and Age at Start
Classifiers

To predict autism at pre-diagnostic ages, we performed a classifier analysis using scores from temperament factors as features. Seven classifiers were built based on different features: 1) surgency; 2) negative affect; 3) effortful control; 4) surgency + negative affect; 5) surgency + effortful control; 6) effortful control + negative affect; 7) surgency + negative affect + effortful control. Each of these seven classifiers was tested using the intercept and slope of the linear developmental trajectories between 8 and 24 months, and using cross-sectional measures at: 1) 8 months; 2) 14 months; 3) 24 months. Thus, a total of 28 classifiers have been tested to predict HR-ASD vs. HR-Typical + HR-Atypical, and HR-ASD + HR-Atypical vs. HR-Typical.

Figure S2.4. Prediction of ASD clinical outcome: AUC. In this figure the area under the curve (AUC, %) is reported for different classifiers based on temperament factors (surgency, negative affect, effortful control) and their combinations at different time points (developmental trajectory between 8 and 24 months). Performance is computed for classification of HR-ASD from high-risk infants without a subsequent diagnosis of ASD at 36 months (i.e. HR-Typical, HR-Atypical). Individual developmental trajectories were obtained from linear modeling between 8 and 24 months, and intercept and slope have been used as features for the classifiers. 95% confidence interval is also reported for each classifier.
Figure S2.5. Prediction of atypical development: AUC. In this figure the area under the curve (AUC, %) is reported for different classifiers based on temperament factors (surgency, negative affect, effortful control) and their combinations at different time points (developmental trajectory between 8 and 24 months). Performance is computed for classification of high-risk infants with atypical development at 36 months (i.e. HR-Atypical and HR-ASD) from their typically developing peers. Individual developmental trajectories were obtained from linear modeling between 8 and 24 months, and intercept and slope have been used as features for the classifiers. 95% confidence interval is also reported for each classifier.

Table S2.1. Best classifiers at each time point.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>p</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effortful control at 14 months</td>
<td>0.047</td>
<td>64.4</td>
<td>69.6</td>
<td>59.2</td>
<td>60.9</td>
<td>26.6</td>
<td>90.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[46.7, 80.3]</td>
<td>[37.4, 97.8]</td>
<td>[44.5, 73.4]</td>
<td>[47.9, 73.6]</td>
<td>[9.9, 45.2]</td>
<td>[78.3, 99.4]</td>
</tr>
<tr>
<td>All factors at 24 months</td>
<td>0.020</td>
<td>71.5</td>
<td>84.8</td>
<td>58.2</td>
<td>62.8</td>
<td>29.9</td>
<td>94.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[57.1, 83.5]</td>
<td>[60.5, 98.9]</td>
<td>[43.4, 72.6]</td>
<td>[49.4, 75.5]</td>
<td>[13.0, 48.5]</td>
<td>[85.7, 99.7]</td>
</tr>
<tr>
<td>Effortful control + Negative affect between 8 and 24 months</td>
<td>0.013</td>
<td>67.6</td>
<td>77.3</td>
<td>57.8</td>
<td>61.1</td>
<td>28.0</td>
<td>92.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[50.5, 81.7]</td>
<td>[47.3, 99.1]</td>
<td>[42.8, 72.1]</td>
<td>[47.9, 74.0]</td>
<td>[11.5, 46.2]</td>
<td>[81.0, 99.7]</td>
</tr>
</tbody>
</table>

Notes. Performance metrics for classifiers using as features temperament factors measured at different time points for classifying HR siblings who later develop ASD (HR-ASD vs. HR-Typical + HR-Atypical). Classifiers for HR atypically developing siblings (HR-ASD + HR-Atypical vs. HR-Typical) were not included due to performance not significantly different from chance level. All classifiers reported in this table significantly differed from prediction at chance level (shuffle test p<0.05). Decision on the best classifier was based on having the highest AUC within an observation time point. All metrics are reported as mean [95% confidence interval]. 95% confidence interval was computed using bootstrap. Abbreviations: AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value.

* Comparing performance of the best classifiers at different time points via Wilcoxon tests, we found that the integrated classifier at 24 months was marginally different from effortful control at 14 months after Bonferroni correction (z=-2.14, p=0.032 with \( \alpha_{\text{Bonferroni}} = 0.05/3 = 0.017 \)), but not from classifier built on the developmental trajectory of effortful control plus negative affect between 8 and 24 months (z=1.86, p=0.06 with \( \alpha_{\text{Bonferroni}} = 0.05/3 = 0.017 \)). Classifiers at 14 months and on the developmental trajectory between 8 and 24 months were not significantly different (z=-1.17, p=0.24 with \( \alpha_{\text{Bonferroni}} = 0.05/3 = 0.017 \)).
Table S2.2. Difference between classifier performance: HR-ASD vs HR-Atypical + HR-Typical.

<table>
<thead>
<tr>
<th>Paired classifiers at 14 months (effortful control vs.)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgency</td>
<td>-2.70</td>
<td>0.007</td>
</tr>
<tr>
<td>Negative affect</td>
<td>-2.80</td>
<td>0.005</td>
</tr>
<tr>
<td>Surgency + negative affect</td>
<td>-2.80</td>
<td>0.005</td>
</tr>
<tr>
<td>Surgency + effortful control</td>
<td>-0.70</td>
<td>0.484</td>
</tr>
<tr>
<td>Effortful control + negative affect</td>
<td>-1.13</td>
<td>0.260</td>
</tr>
<tr>
<td>All factors</td>
<td>-0.46</td>
<td>0.646</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paired classifiers at 24 months (all factors vs.)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgency</td>
<td>-2.70</td>
<td>0.007</td>
</tr>
<tr>
<td>Negative affect</td>
<td>-2.70</td>
<td>0.007</td>
</tr>
<tr>
<td>Effortful control</td>
<td>-0.51</td>
<td>0.610</td>
</tr>
<tr>
<td>Surgency + negative affect</td>
<td>-2.50</td>
<td>0.013</td>
</tr>
<tr>
<td>Surgency + effortful control</td>
<td>-1.58</td>
<td>0.114</td>
</tr>
<tr>
<td>Effortful control + negative affect</td>
<td>-0.98</td>
<td>0.327</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paired classifiers on longitudinal trajectories (effortful control + negative affect vs.)</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgency</td>
<td>-2.81</td>
<td>0.005</td>
</tr>
<tr>
<td>Negative affect</td>
<td>-2.81</td>
<td>0.005</td>
</tr>
<tr>
<td>Effortful control</td>
<td>-0.42</td>
<td>0.678</td>
</tr>
<tr>
<td>Surgency + negative affect</td>
<td>-2.81</td>
<td>0.005</td>
</tr>
<tr>
<td>Surgency + effortful control</td>
<td>-1.28</td>
<td>0.202</td>
</tr>
<tr>
<td>All factors</td>
<td>-1.12</td>
<td>0.262</td>
</tr>
</tbody>
</table>

Notes. Significant differences in performance between the best classifier at a specific time point and the other classifier within the same time point were tested by a two-sided Wilcoxon rank sum test when Friedman test on all classifiers performance at each time point was significant. Bonferroni correction was used to correct post-hoc Wilcoxon tests for multiple comparison (pairs=6) and results were considered significant for \( p_{\text{Bonferroni}} = 0.008 \). Results from Friedman tests are \( \chi^2(6)=40.6 \) and \( p<10^{-3} \) at 14 months; \( \chi^2(6)=34.8 \) and \( p<10^{-3} \) at 24 months; and \( \chi^2(6)=40.4 \) and \( p<10^{-3} \) using longitudinal trajectories between 8 and 24 months.
Parent-child interaction during the first year of life in infants at high and low risk of autism spectrum disorder

Pijl1,2, M.K.J., Bontinck3, C., Rommelse2,4, N.N.J., Begum Ali5, J., Cauvet6,7, E., Niedzwiecka8, A., Falck-Ytter6,7,9, T., Jones8, E.J.H., Van den Boomen10,11, C., Bölte6,7, S., Johnson5,12, M.H., Charman13,14, T., Warreyn1, P., Roeyers3, H., Buitelaar1,2, J.K., Oosterling2, I.J. & the EuroSibs Team*

Under review

1 Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, the Netherlands
2 Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, the Netherlands
3 Ghent University, Department of Experimental-Clinical and Health Psychology, Belgium
4 Radboud University Medical Centre, Department of Psychiatry, Nijmegen, the Netherlands
5 Centre for Brain and Cognitive Development, Birkbeck College, University of London, United Kingdom
6 Center of Neurodevelopmental Disorders at Karolinska Institutet (KIND), Neuropsychiatry Unit, Department of Women’s and Children’s Health, Karolinska Institutet, Stockholm, Sweden
7 Child and Adolescent Psychiatry, Center for Psychiatry Research, Stockholm County Council, Stockholm, Sweden
8 University of Warsaw, Faculty of Psychology, Department of Health and Rehabilitation Psychology, Poland
9 Uppsala University, Department of Psychology, Sweden
10 Helmholtz Institute, Department of Experimental Psychology, Utrecht, the Netherlands
11 Utrecht University, Department of Developmental Psychology, Utrecht, the Netherlands
12 Psychology Department, University of Cambridge, United Kingdom
13 King’s College London, Psychology Department, Institute of Psychiatry, Psychology & Neuroscience, London, United Kingdom
14 South London and Maudsley NHS Foundation Trust, London, United Kingdom

Abstract

The interaction with parents forms a key aspect of an infant’s social environment, but few prospective studies have examined parent-child interactions in the first year of life in infants at high (HR) and low (LR) risk of autism spectrum disorder (ASD). As part of a European multisite network, parent-child dyads were videotaped and observed at 5 months (62 HR siblings, 47 LR siblings) and 10 months (101 HR siblings, 77 LR siblings), using the Parent-Infant/Toddler Coding of Interaction (PIntCI) scheme. Linear mixed model analyses showed that at 10 months HR siblings had fewer initiations towards their parents than LR siblings ($p<0.05$, $d=0.44$). Group differences remained after correcting for the influence of infant’s sex and age, but were no longer significant after controlling for parental educational level ($p>0.05$, $d=0.36$). No group differences were found in infant behaviors at 5 months or in parent behaviors at 5 or 10 months of age (all $p>0.05$, $d<0.21$). This study provides preliminary evidence that 10-month-old HR infants differ from their LR peers in their interaction with their parents and may help to strengthen parent-child interaction as a promising avenue for early interventions.
During the first year of life, the interaction with parents forms a key aspect of an infant’s social environment. In the context of emerging autism spectrum disorder (ASD), investigation of early parent-infant interactions may provide insight into the developmental course of early social and communicative alterations, before ASD symptoms become clinically manifest (Wallace & Rogers, 2010). It is important to study the bidirectional nature of interactions and dissect the contribution of the infant and the parent for determining targets for parent-mediated interventions for ASD (Kasari et al., 2014; Rogers et al., 2014; Wetherby et al., 2014). In addition, parent-infant interaction research may also help to determine targets for pre-emptive interventions for infants or toddlers at risk (i.e. ‘prevention’ to redirect developmental trajectories before the full-blown disorder becomes manifest) (Green et al., 2017; Jones, Dawson, Kelly, Estes, & Jane Webb, 2017; Watson et al., 2017). As research into this domain is relatively new, improved knowledge about how perturbations in early parent-infant interaction arise and how they change over time may help to further improve these early parent-mediated (pre-emptive) interventions. The aim of the current study was therefore to investigate differences in parent-infant interactions in the first year between infants at high familial risk of ASD (HR) and infants at low risk (LR). Previous longitudinal studies using a HR design mainly showed that ASD-related precursors and/or early symptoms start to emerge toward the end of the first year (Jones, Gliga, Bedford, Charman, & Johnson, 2014). However, these HR studies have not yet performed a fine-grained observation of naturalistic parent-infant interactions across the first 12 months of life and may have missed more subtle alterations of the early parent-child interaction. In addition, given that parents of HR siblings may apply their learned interaction style based on the interaction with the older sibling with ASD, differences may also already be present in parent behaviors during the first year of life.

Previous retrospective and prospective research indicates that the interaction between parents and infants later diagnosed with (or at risk of) ASD in the first year of life differs from the interaction between parents and typically developing infants (Campbell, Leezenbaum, Mahoney, Day, & Schmidt, 2015; Harker, Ibanez, Nguyen, Messinger, & Stone, 2016; Northrup & Iverson, 2015; Rozga et al., 2011; Saint-Georges et al., 2011; Steiner, Gengoux, Smith, & Chawarska, 2018; Trevarthen & Daniel, 2005; Wan et al., 2012, 2013; Yirmiya et al., 2006). These studies analyzed videos of parent-infant interactions and used either a macro-level coding approach (i.e. global rating of qualitative and quantitative characteristics) or a micro-level coding approach (i.e. detailed observation of frequency, duration, and timing of behaviors). Only a few retrospective home video studies have focused specifically on differences in parent-infant interaction between infants subsequently diagnosed with ASD and controls (Saint-Georges et al., 2011; Trevarthen & Daniel, 2005).
Results revealed that infants who were later diagnosed with ASD (n=15) differed from typically developing infants (n=15) in their levels of social orientation, receptiveness to the parent, and seeking behaviors (i.e., spontaneous and intentional communication) across the first year (Saint-Georges et al., 2011). Their parents tended to use more stimulation while initiating interaction and fewer gestures in response to their child. Another study focusing on 11-month-old monozygotic twin girls, one of whom was later diagnosed with ASD, showed that the interaction between the father and the daughter diagnosed with ASD was characterized by an asynchronous pattern of behaviors and a lack of shared experience and enjoyment (Trevarthen & Daniel, 2005). Although retrospective home video research can provide insight into the developmental course of ASD, it also has its shortcomings, including a bias resulting from what parents make available for study and variability in the content of home videos (e.g., different settings). The key to overcome these issues is to prospectively study the interaction between parents and infants who are at high risk of developing ASD because they have an older sibling diagnosed with ASD (Szatmari et al., 2016).

To date, only one prospective study investigated parent-infant dyads in HR and LR siblings during the first half year of life (0 to 6 months) (Yirmiya et al., 2006). A microanalysis of parent and infant affective states at 4 months showed a weaker synchrony in infant-led interactions (i.e., baby leads, parent follows) in HR (n=21) as compared to LR dyads (n=21), suggesting that the parents of HR infants experience difficulties in adapting their affective behaviors to the affect initiated by the infant. With regard to infant behaviors during the second half-year of life (6 to 12 months), HR siblings (n=45) tended to be less lively than LR siblings (n=47), as shown by lower global ratings of physical activity (Wan et al., 2012, 2013). In addition, Campbell et al. (2015) concluded that HR siblings subsequently diagnosed with ASD (n=10) were less socially engaged when interacting with their parents than LR siblings (n=27), as indicated by a lower global rating of infant reciprocity with the parent and lower frequency of giving and showing toys. However, the same study reported similar levels of social engagement based on the frequency of directed vocalizations and shared positive affect. The latter findings were confirmed by other studies, showing no differences in the first year between HR and LR siblings in attentiveness toward a parent (Steiner et al., 2018; Wan et al., 2012, 2013), positive affect (Wan et al., 2012, 2013), directed gaze (Rozga et al., 2011), smiles (Harker et al., 2016; Rozga et al., 2011), gestures (Talbott, Nelson, & Tager-Flusberg, 2015), vocalizations (Northrup & Iverson, 2015; Rozga et al., 2011; Talbott, Nelson, & Tager-Flusberg, 2016), or integration of communicative behaviors (Parlade & Iverson, 2015). With regard to parental behavior, research has shown that the parents of HR siblings aged 6 to 12 months are more directive when interacting with their infant than the parents of LR siblings of the same age (Harker et al., 2016; Steiner et al., 2018; Wan et al., 2012,
2013), involving a more parent-directed course of the interaction including intrusive or demanding behaviors (e.g. redirecting child’s attention). Sensitivity tends to be similar among the parents of HR and LR siblings (Campbell et al., 2015; Harker et al., 2016), with a trend for lower sensitivity among the parents of HR siblings (Wan et al., 2012, 2013). Taken together, at least some differences between HR and LR dyads start to arise in the first year of life, which underlines the importance of studying early parent-infant interactions. However, the evidence is on a preliminary level, and far from conclusive.

Inconsistency of results so far is probably in part due to small sample sizes, differences across studies in age ranges and group comparisons (e.g. HR vs. LR, HR-ASD vs. HR no ASD), and the use of coding schemes that do not distinguish between behaviors that are spontaneously initiated by the infant and those made in a response to parental behaviors. In addition, only one prospective study focused on the first 6 months of life and did not report findings on parents and infants separately. None of the abovementioned studies investigated parent-infant interaction pathways across the first year of life. Thus, research into parent-child dyads focusing on both parent and infant behaviors at different time points in infancy are desirable to complement previous studies.

The current study aimed to advance this area of research by focusing on infant, parent, and dyadic behaviors during interactions between parents and infants at HR and LR of ASD at 5 and 10 months of age. To this end, the Parent-Infant/Toddler Coding of Interaction (PInTCI) was developed, which rates infant/toddler (including both initiations and responses), parent, and dyadic behaviors considered possible predictors for ASD or general cognitive or language development. We examined 195 parent-infant dyads (113 HR, 82 LR) during free-play as part of an ongoing longitudinal study involving a large international sample. On the basis of previous research, we hypothesized that HR dyads at 10 months would show lower ratings of infant social communicative behaviors, higher levels of parental negative control, and lower dyadic reciprocity. No specific hypotheses were formulated regarding parent-infant dyads at 5 months, as no previous prospective research exists. Owing to the bidirectional nature of parent-child interactions, we expected effects to be accumulative across time, showing larger differences between groups at 10 months than at 5 months.
Methods

Participants
As part of the ongoing EuroSibs Autism Research Network (www.eurosibs.eu’), 195 infants (113 HR and 82 LR) were assessed at 5 and/or 10 months of age. Ethical approval was given by local ethics committees in participating countries, and parents gave informed consent. The HR infants had at least one older sibling with a community clinical diagnosis of ASD (hereafter: ‘proband’). The LR siblings had at least one older sibling with typical development and no ASD within first-degree family members (as confirmed through a parent interview regarding family medical history). Exclusion criteria for both HR and LR infants included diagnosis of epilepsy, preterm birth (i.e. ≥36 weeks, N=1 was born at 35 weeks), and genetic syndromes clearly related to ASD in infant or proband (e.g. fragile X syndrome, tuberous sclerosis).

Table 3.1. Sample characteristics by risk group at 5 and 10 months.

<table>
<thead>
<tr>
<th></th>
<th>5 months</th>
<th>10 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR siblings (n=62)</td>
<td>LR siblings (n=47)</td>
</tr>
<tr>
<td>Sex infant (% male)</td>
<td>51.6</td>
<td>57.4</td>
</tr>
<tr>
<td>Sex parent (% male)</td>
<td>1.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Chronological age (months)</td>
<td>5.3 (0.7)</td>
<td>5.5 (0.6)</td>
</tr>
<tr>
<td>MSEL Non-verbal IQ</td>
<td>45.1 (6.6)</td>
<td>47.5 (6.7)</td>
</tr>
<tr>
<td>Educational level parent (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>40.7</td>
<td>6.8</td>
</tr>
<tr>
<td>Medium</td>
<td>37.3</td>
<td>34.1</td>
</tr>
<tr>
<td>High</td>
<td>22.0</td>
<td>59.1</td>
</tr>
</tbody>
</table>

* p < 0.05 ** p < 0.001
Notes. HR = high-risk; LR = low-risk; MSEL = Mullen Scales of Early Learning (Mullen, 1995).

Videotaped parent-child dyads were collected from sites in different countries – there were 45 clips from Belgium, 58 from the United Kingdom, 47 from the Netherlands, and 45 from Sweden. Of the 195 infants, 62 HR infants (32 male, 51.6%) and 47 LR infants (27 male, 57.4%) attended the 5-month visit, and 101 HR infants (53 male, 52.5%) and 77 LR infants (40 male, 51.9%) the 10-month visit. Complete 5- and 10-month data were available for 50 HR infants (25 male, 50.0%) and 42 LR infants (25 male, 59.5%). See Table 3.1 for detailed sample characteristics.

1 Participating sites in the current study: Ghent University (Ghent, Belgium), Birkbeck College (London, UK), Radboud University Medical Centre (Nijmegen, the Netherlands), University Medical Centre Utrecht (Utrecht, the Netherlands), Uppsala University (Uppsala, Sweden).
Procedure
Free play interactions between parents and infants were videotaped. At both time points, parents were instructed to play as usual on a play mat on the floor, without making any additional demands on their child. Instructions and toy categories (i.e. pretend play materials, construction toys, spinning toy, exploratory toy, book) were identical across sites. To keep the context as similar as possible across time, toy categories were kept constant (see Figure 3.1) and the same parent was involved in the play sessions at both time points. A play session typically lasted about 10 minutes, of which the first 5 minutes were coded. The observation of the interaction started at the moment that the researcher had left the room, parent and child started the interaction, and the video cameras sufficiently captured both parent and child.

To ensure the quality of video clips, standard procedures were followed and regular quality checks of video clips from each site were arranged by a data monitoring panel. Six coders (one English, one Swedish, two Belgian, two Dutch) were trained to use the coding scheme reliably before they independently coded clips. To prevent drift and to ensure reliable scoring throughout, coders regularly scored booster clips on which they received feedback from the trainers (MP, CB). Booster clip scores were included in the inter-rater reliability (IRR) and core analyses.

Measures
Parent-Infant/Toddler Coding of Interaction (PInTCI). The PInTCI (Pijl, Bontinck, Oosterling, & Warreyn, 2016) was used to evaluate parent-child interactions. This global coding scheme was developed after an extensive literature review on

Figure 3.1. Toys at the 5- and 10-month time point, including similar categories (i.e. pretend play materials, construction toys, spinning toy, exploratory toy, book).
characteristics of parent-child interaction that predicted ASD or general cognitive or language development in previous research. Many coding schemes are available, but, to our knowledge, no existing coding scheme includes all the constructs found to predict subsequent child development, and which can be used in different age groups across early development (5-36 months). To be applicable across early development, scales from existing micro and macro measures were adapted and combined: Coding Interactive Behavior (Feldman, 1998), coding scheme for the Communication Play Protocol (Adamson & Bakeman, 1999; Adamson, Bakeman, Deckner, & Nelson, 2012), Dyadic Communication Measure for Autism (Aldred, Green, & Adams, 2004), Erickson coding scales (Erickson, Sroufe, & Egeland, 1985), Manchester Assessment of Caregiver-Infant Interaction (Wan et al., 2012, 2013), Maternal Behavior Rating Scale (Mahoney & Perales, 2003; Mahoney, Powell, & Finger, 1986), Siller’s and Sigman’s coding scheme (Siller & Sigman, 2002), Social Interaction Rating Scale (Ruble, McDuffie, King, & Lorenz, 2008), infant coding scales (Clifford & Dissanayake, 2009), scaffolding scales (Baker, Fenning, Crnic, Baker, & Blacher, 2007; Dieterich, Assel, Swank, Smith, & Landry, 2006; Hoffman, Crnic, & Baker, 2006), and coding maternal response behaviours (Flynn & Masur, 2007; Landry, Smith, & Swank, 2006; Lloyd & Masur, 2014). Before the actual application of the coding scheme, there was a period of extensive pilot work during which two of the developers coded ten video clips of HR and LR infants interacting with their parents, including infants in the age range between 0 and 36 months. Based on these codings the developers iteratively revised and improved the rating scales.

The PInTCI consists of five child constructs (attentiveness – initiations – sharing of affect – positive affect – negative affect), five parent constructs (sensitive responsiveness – negative control – scaffolding – positive affect – negative affect), and one dyadic construct (dyadic reciprocity), rated on a 1-7 scale. A score of 1 consistently reflects maladjusted/negative behavior while a score of 7 reflects more adaptive/optimal behavior. A brief description of the PInTCI scales can be found in the Supplemental Material.

Inter-rater reliability was calculated for 24 clips (12 5-month and 12 10-month dyads) to investigate the relative agreement between the coders, with values classified as poor (0.00–0.40), fair (0.41–0.59), good (0.60–0.75), and excellent (>0.75) (Fleiss, 1986). To overcome language barriers, all clips contained English-speaking dyads. The IRR clips included both HR and LR dyads at both 5 and 10 months of age, and were included in the core analyses. Coders were blind for risk status and did not have access to the scores coded by the other coders. Constructs with intraclass correlations (ICC) below 0.60 were removed from further analyses (McHugh, 2012). ICC values showed excellent inter-rater reliability for all PCI constructs.
at 10 months: ICCs ranged from 0.68 to 0.95 (p<0.01). At 5 months, reliability ranged from good to excellent: ICCs ranged from 0.67 to 0.92 (p<0.01), except for infant initiations (ICC=0.42, p>0.05) and parental negative affect (ICC=0.31, p>0.05). The fair reliability for infant initiations at 5 months may imply that early infant behaviors are too subtle to enable coders to reliably observe initiative behavior. In typically developing infants, clear initiative behaviors start to develop at around 8 to 10 months, with gestures such as pointing and showing or alternating gaze (i.e. joint attention) (Bates & Dick, 2002). The poor reliability of parental negative affect was likely caused by the limited range of codes (i.e. mainly 5 to 7) in the IRR clips (Hallgren, 2012). See Supplemental Material for an overview of ICC values. More information about this measure can be obtained from the first author.

**Statistical analyses**

To analyze group differences in the parent-child dyads, linear mixed models were used with risk status (HR, LR) as fixed factor, time (5 months, 10 months) as repeated measure, and site (Belgium, the Netherlands, the United Kingdom, Sweden) as a random effect to account for within site correlation. A group by time interaction effect was included to analyze group differences over time. This approach was applied because it allowed modeling the statistical dependency among observations by including site as a random effect and enabled the use of information from participants with missing data across time (only one time point available for n=63 HR siblings and n=40 LR siblings). Age was variable at the 5- and 10-month time points (between 4 and 7 months, and between 9 and 12 months, respectively), introducing potential noise in results. Therefore, infant age was included as a covariate. We also adjusted for parental educational level, as a measure of social economical status, and infant sex. Cognitive functioning was not included as a covariate, because measurement of the cognitive development in infants is likely to include components of neurodevelopmental disorders (such as ASD) that make this measure impossible to disentangle from the disorder itself (Dennis et al., 2009). Including the MSEL as a covariate in our analyses may therefore lead to overcorrection. A correction for multiple comparisons was applied, using the false discovery rate controlling procedure with a q-value of 0.05 (Benjamini & Hochberg, 1995). The dependent variables were not normally distributed, which is a prerequisite for variance analyses. Therefore, on all PInTCI variables a Van der Waerden transformation was applied, which transforms raw scores into z-scores corresponding to the estimated cumulative proportion of the distribution analogous to a particular rank (using Statistical Package for the Social Sciences [SPSS] version 22). Data across both time points together were transformed, otherwise data would be standardized to a mean level of zero at both time points preventing the analyses of a main effect of time (and group by time interaction). Analyses were carried out on the transformed values,
Table 3.2. Spearman correlations between PInTCI scales at the 5- and 10-month time points.

<table>
<thead>
<tr>
<th>Dyad</th>
<th>Infant 1</th>
<th>Infant 2</th>
<th>Infant 3</th>
<th>Infant 4</th>
<th>Infant 5</th>
<th>Parent 1</th>
<th>Parent 2</th>
<th>Parent 3</th>
<th>Parent 4</th>
<th>Parent 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initiations (Infant 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.52*</td>
<td>0.44*</td>
<td>0.35*</td>
<td>0.07</td>
<td>0.14</td>
<td>0.27*</td>
</tr>
<tr>
<td>Attentiveness (Infant 2)</td>
<td>0.40*</td>
<td></td>
<td></td>
<td></td>
<td>0.55*</td>
<td>0.55*</td>
<td>0.12</td>
<td>0.25*</td>
<td>0.22</td>
<td>0.41*</td>
</tr>
<tr>
<td>Sharing of Affect (Infant 3)</td>
<td>0.48*</td>
<td>0.54*</td>
<td></td>
<td></td>
<td>0.65*</td>
<td>-0.03</td>
<td>0.16</td>
<td>0.10</td>
<td>0.16</td>
<td>0.38*</td>
</tr>
<tr>
<td>Positive Affect (Infant 4)</td>
<td>0.34*</td>
<td>0.54*</td>
<td>0.69*</td>
<td></td>
<td>0.13</td>
<td>0.23</td>
<td>0.21</td>
<td>0.22</td>
<td>0.42*</td>
<td>0.05</td>
</tr>
<tr>
<td>Negative Affect (Infant 5)</td>
<td>-0.16</td>
<td>0.15</td>
<td>-0.03</td>
<td>0.27</td>
<td>0.03</td>
<td>0.07</td>
<td>0.11</td>
<td>-0.05</td>
<td>0.10</td>
<td>0.07</td>
</tr>
<tr>
<td>Sensitive Responsiveness (Parent 1)</td>
<td>0.23</td>
<td>0.42*</td>
<td>0.30</td>
<td>0.24</td>
<td>0.20</td>
<td></td>
<td>0.34*</td>
<td>0.56*</td>
<td>0.54*</td>
<td>0.23</td>
</tr>
<tr>
<td>Negative Control (Parent 2)</td>
<td>-0.07</td>
<td>0.03</td>
<td>0.14</td>
<td>0.14</td>
<td>0.19</td>
<td>0.17</td>
<td></td>
<td>0.26</td>
<td>0.22</td>
<td>0.18</td>
</tr>
<tr>
<td>Scaffolding (Parent 3)</td>
<td>0.18</td>
<td>0.53*</td>
<td>0.27</td>
<td>0.28</td>
<td>0.08</td>
<td>0.67*</td>
<td></td>
<td>0.07</td>
<td>0.41*</td>
<td>-0.01</td>
</tr>
<tr>
<td>Positive Affect (Parent 4)</td>
<td>0.22</td>
<td>0.43*</td>
<td>0.50*</td>
<td>0.56*</td>
<td>0.23</td>
<td>0.41</td>
<td>0.01</td>
<td></td>
<td>0.44*</td>
<td>0.18</td>
</tr>
<tr>
<td>Negative Affect (Parent 5)</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.10</td>
<td>-0.13</td>
<td>0.08</td>
<td>0.05</td>
<td>0.03</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Dyadic Reciprocity (Dyad)</td>
<td>0.45*</td>
<td>0.86*</td>
<td>0.49*</td>
<td>0.52*</td>
<td>0.23</td>
<td>0.50*</td>
<td>0.08</td>
<td>0.55*</td>
<td>0.43</td>
<td>-0.05</td>
</tr>
</tbody>
</table>

* *p < 0.001

Notes. Numbers below diagonal represent 5-month time point, numbers above diagonal represent 10-month time point. **Bold** values indicate correlations > 0.80. **Italic** values indicate that data were removed from core analyses due to insufficient inter-rater reliability. PInTCI = Parent-Infant/Toddler Coding of Interaction.
but to facilitate interpretation of the findings, raw mean scores are reported in the Results and Supplemental Material. Results were similar for raw and transformed data.

Results

Descriptive statistics and correlations
No significant differences were found between the HR and LR groups in infant’s sex, parent’s sex, or infant’s chronological age at either time point (see Table 3.1). However, at the 10-month time point infant’s non-verbal IQ differed between the HR and LR groups ($t(176)=2.71, p=0.007$), showing that the HR siblings had lower levels of cognitive development than the LR siblings. In addition, an association between parental educational level and risk status was observed for both the 5-month and 10-month time point ($\chi^2(2)=20.24, p<0.001; \chi^2(2)=16.79, p<0.001$, respectively).

Table 3.2 reports the correlations between the PInTCI scales at the separate time points, and between 5- and 10-month codings. At both time points, the correlation between attentiveness toward the parent and dyadic reciprocity exceeded $r_s > 0.80$. This indicates the presence of multicollinearity and therefore the dyadic scale was removed from further analyses. Correlations between scales at the 5- and 10-month time point are reported in the Supplemental Material.

Group differences in parent-child dyads
Linear mixed models (LMM) were applied to analyze group differences in the parent-infant dyads (see Table 3.3 for an example). LMM analysis revealed a significant main group effect for infant initiations ($F(1, 282.88)=8.57, p<0.05, d=0.44$), indicating that HR siblings initiated less toward their parents than did LR siblings. Given the low ICC for infant initiations at 5 months, LMM analysis was performed for the 10-month time point only (i.e. age and the interaction effect age x group were excluded from the model). Further, main time effects were found for infant positive affect ($F(1, 131.20)=16.54, p<0.001, d=0.59$), infant negative affect ($F(1, 140.56)=19.04, p<0.001, d=0.63$), parental sensitive responsiveness ($F(1, 127.38)=9.65, p<0.01, d=0.45$), and parental scaffolding ($F(1, 130.45)=12.20, p<0.01, d=0.51$). These time effects showed that at 10 months there were higher levels of infant positive affect, parental sensitive responsiveness, and parental scaffolding, and lower levels of infant negative affect than at 5 months. Importantly, no significant group by time interactions were present (all $p>0.87$), indicating that the changes in parent-infant interactions over time were similar in HR and LR siblings. The random effect of site was not significant (all $p>0.33$), suggesting that the site at which parent-infant dyads were collected did not influence outcomes. See Figure 3.2 and 3.3 for the PInTCI
ratings at the 5- and 10-month time points by risk group and Supplemental Material for detailed information.

Remarkably, the significant group and time effects remained after correcting for the influence of infant’s sex and age. However, the adjustment for parental educational level resulted in a non-significant group effect for infant initiations ($F(1, 232.66)=5.71, p>0.05, d=0.36$), suggesting that the group difference between HR and LR infants was at least partly attributable to different levels of parental education. Significant time effects remained after adjusting for parental educational level.

**Table 3.3.** Example of a Linear Mixed Model (LMM) analysis.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>(Sex)</td>
<td>0.16</td>
<td>0.84</td>
</tr>
<tr>
<td>(Age T1)</td>
<td>0.27</td>
<td>0.89</td>
</tr>
<tr>
<td>(Age T2)</td>
<td>0.27</td>
<td>0.95</td>
</tr>
<tr>
<td>(Parental educational level)</td>
<td>2.26</td>
<td>0.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Repeated effects</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>9.65</td>
<td>0.004</td>
</tr>
<tr>
<td>Time by Group</td>
<td>0.10</td>
<td>0.87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effect</th>
<th>Estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site</td>
<td>1.12</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Notes. This table presents data for the scale Parental Sensitive Responsiveness. Text between brackets indicate the application of covariates. Estimates for the Group, Time, and Site effects are derived from the model without covariates. Reported p-values are FDR corrected.
Figure 3.2. Mean ratings of the Parent-Infant/Toddler Coding of Interaction (PiNTCI) at the 5-month time point by risk group with error bars representing standard errors (1=maladjusted/negative behaviors; 7=more adaptive behaviors).

Figure 3.3. Mean ratings of the Parent-Infant/Toddler Coding of Interaction (PiNTCI) at the 10-month time point by risk group with error bars representing standard errors (1=maladjusted/negative behaviors; 7=more adaptive behaviors).
Discussion

The purpose of this study was to obtain a detailed picture of the interactions between parents and their 5- and 10-month-old infants at HR or LR of ASD. Whereas previous studies investigated parent-infant interactions mainly starting in the second half of the first year, the current study complements research by prospectively examining parent-infant free play at 5 and 10 months, and by including infant’s initiating and responding behaviors. Results showed that the newly developed coding scheme had adequate inter-rater reliability for all scales, except for infant initiations and parent negative affect at 5 months (which were removed from further analyses). Preliminary findings revealed that group differences started to emerge in the second half year, with 10-month-old HR infants initiating fewer social interactions than LR infants. However, although group differences remained after correcting for the influence of infant’s sex and age, they were no longer significant after controlling for parental educational level. No group differences were found in infant behaviors at 5 months or in parent behaviors at 5 or 10 months. Examination across time showed that in both HR and LR dyads there were higher levels of infant positive affect, parental sensitive responsiveness, and parental scaffolding, and lower levels of infant negative affect at 10 months than at 5 months, but that these changes were similar in both groups.

Our preliminary finding that 10-month-old HR siblings initiated fewer interactions with their parents than did LR siblings is consistent with previous retrospective research (Saint-Georges et al., 2011), showing a much smaller increase in spontaneous and intentional communication during the first year of life in infants later diagnosed with ASD than in typically developing infants. Only one prospective study focused specifically on infant-initiated communication and found slower growth in the coordination of initiative behaviors of HR siblings who were later diagnosed with ASD as compared to their HR and LR peers (Parlade & Iverson, 2015). However, differences were only significant from 12 months onward, and not before their first birthday. This may be due to differences in the applied coding measures, in that frequency coding (as was used in the study by Parlade and Iverson (2015)) may fail to capture more subtle behaviors as compared to global coding that allows to rate the quality of initiations and integrate contextual information (i.e. the meaning and appropriateness of behaviors). Given that parental behaviors did not differ between HR and LR dyads at either 5 or 10 months in our sample, differences in infant behaviors are likely to reflect early emerging atypicalities within the broader autism phenotype (BAP) or later ASD instead of reflecting the impact of parental behaviors. However, the group difference in infant initiations was non-significant after correcting for parental educational level, suggesting that the group difference was at least partly attributable to different levels of parental education. Although
previous research suggests that socioeconomic status (SES) predicts parenting, little is known about the causal pathways that link components of SES (e.g. education) to infant behaviors, potentially mediated by parental behaviors (Bornstein, 2002). Therefore, in order to fully understand parent-infant dyads, future research should further investigate the influence of parental education on parent-infant interaction. Future research should also include other parent and infant characteristics that may influence parent-infant dyads. Given this gap in our understanding of parent-child interactions and the snapshots from which parent-child dyads are analyzed, results have to be interpreted carefully. Although firm conclusions cannot be drawn, this observation may support the use of intervention strategies at an early prediagnostic stage, with a focus on the initiation of social interaction as a core competence (Green et al., 2017; Watson et al., 2017).

The patterns of other infant behaviors (i.e. attentiveness to parent, shared affect, positive affect, and negative affect) were similar at 10 months, with HR infants being more likely to show limited social communication than LR infants, but this difference was not statistically significant. This supports retrospective work by Saint-Georges et al. (2011) who proposed low initiation of social interaction as being a primary indicator in infants subsequently diagnosed with ASD. This may reflect an initial disruption in social motivation mechanisms, starting with a low drive for social initiative that deprives the child of social experiences, which, ultimately, leads to more pronounced impairments in social communication. This would also explain our null findings at 5 months, substantiating previous findings indicating no clear behavioral markers of ASD during the first 6 months of life (Jones et al., 2014; Landa, Gross, Stuart, & Faherty, 2013; Zwaigenbaum et al., 2005). The idea of a growing deviant developmental trajectory in infants subsequently diagnosed with ASD is supported by findings of initial similar developmental levels at 6 months of age in ASD and non-ASD HR and LR infants (Landa et al., 2013), or only some lower developmental levels in the most severely affected ASD infants (Estes et al., 2015). Thereafter children later diagnosed with ASD show an increasingly atypical development between the first and second birthday (Estes et al., 2015; Landa et al., 2013).

Parental behavior did not differ between HR and LR dyads at 5 or 10 months, implying that the HR parents in our sample appear to be similarly tuned in to their infants’ communicative signals as compared to the LR parents. In line with our results, Campbell et al. (2015) found no differences in macro- and micro-analytically observed parental behaviors between HR and LR dyads. In contrast, other studies found that the parents of HR infants tended to be more directive or demanding than the parents of LR infants (Harker et al., 2016; Steiner et al., 2018; Wan et al., 2012, 2013). This may be partly because of differences in the content of coding
scales. For example, the non-directiveness scale used by Wan et al. (2012, 2013) may also include aspects of scaffolding (“behavior that accepts and encourages the infant to lead interaction and positive comments that reflect the infant’s experience”), whereas we investigated these constructs separately (i.e. negative control, scaffolding). Another explanation for this inconsistency is that our multi-site research project may have generated more variation in the sample, which may have reduced the power to detect differences. Additionally, inconsistent findings may be caused by differences between research samples based on specific parental characteristics that might affect the parent-child interaction. These include the presence of the BAP or clinical ASD symptoms in parents that may impact on the way parents interact with their infant. Also, parental stress, which is often elevated in the parents of children diagnosed with ASD, may influence the emotional availability of parents (Estes et al., 2009; Hayes & Watson, 2013; Kasari & Sigman, 1997). Further, given that the parents of HR siblings might already have received parent-mediated interventions for their child with ASD, learned strategies may have changed their parenting style when they interact with their other children. A future step will be to include measures focusing on clinical symptomatology, stress, and levels of received parent-mediated intervention to reveal explanatory mechanisms in the parent’s interactive behaviors.

Independently of risk group, infant affect, parental sensitivity, and parental scaffolding changed with age. The increased levels of infant positive affect and decreased levels of infant negative affect imply that emotional regulation improves from 5 to 10 months of age, as infants start to use more self-soothing behaviors (e.g., ability to shift attention away from a distressing stimulus) and less crying and fussing as their primary emotion regulation strategies (Calkins & Hill, 2007; Mangelsdorf, Shapiro, & Marzolf, 1995). Furthermore, infants become more interested in playing with objects and this may elicit clearer sensitive and supportive behaviors in the parent. Our findings underline the importance of investigating child and parent behaviors across infancy and raises questions about how these behaviors develop further into toddlerhood. The low correlations between parent-infant interactions at 5 and 10 months, indicating a low stability during infancy, also necessitates more frequent measures of parent-child interaction.

Limitations

Although the current study provides new insights into parent-infant interactions, some study limitations should be acknowledged. First, further follow-up of the current sample of HR and LR siblings will allow us to look more closely at patterns of parent-infant interactions in HR siblings with different outcomes (HR with ASD versus HR no ASD). This provides information about whether differences in social initiations reflect early emerging atypicalities within the BAP or later ASD, and allows the
investigation of predictive relationships, which is important given the bidirectional nature of interaction. Although our sample was significantly larger than in previous studies, future work should include even larger samples given the low proportion of children who later develop ASD. Second, as mentioned above, the multi-site character of this study may have generated more variation in the sample, which may have reduced the power to detect differences. However, a multi-site approach also provides benefits, including a larger sample size and generalizability across countries. Third, parental behaviors may not be representative of their behavior in daily life because of their awareness of being videotaped for the study. Parents may therefore show more socially acceptable behaviors and/or higher levels of stress, and therefore daily life parental behaviors that also affect the observed infant behaviors may have been missed. Finally, although the PInTCI seems to be a promising measure to code parent-infant interactions, our findings should be interpreted with caution given the relatively low number of clips on which inter-rater reliability was calculated. More research into the PInTCI is therefore required and should include a larger percentage of IRR clips. In addition, the low to medium correlations between the PInTCI scales across time also ask for more research into the psychometric properties of the coding measure. Although infant behaviors are likely to vary across the first year due to large developmental changes (i.e. transitioning from a lying, dependent baby to a more independent crawling, babbling (almost) toddler), parent behaviors are expected to be more stable across time. However, given the bidirectional nature of interactions, changes in the child’s development may also explain the lower correlations for parent sensitivity and parent scaffolding. Furthermore, while the coding scheme allowed us to evaluate both qualitative and quantitative aspects of parent-infant interactions while incorporating contextual information, we did not apply a fine-grained micro-coding measure (e.g., Dyadic Communication Measure for Autism (Aldred et al., 2004)). Given that micro- and macro-level coding are complementary approaches (Mesman, 2010), the inclusion of both should be considered in future studies. Also, the additional value of new approaches that efficiently code parent-infant interactions should be investigated instead of manual coding, for example by performing musical micro-analysis (Suvini, Apicella, & Muratori, 2017) or automated movement analysis (Lopez Perez et al., 2017).

Conclusions and future recommendations
In conclusion, this study found preliminary evidence for differences in the initiation of interactions with their parents between 10-month-old HR and LR siblings. Although our findings ask for additional research into the role of parental education and replication within larger cohorts (including longitudinal designs capturing parent-infant and parent-toddler interaction and subsequent child developmental
outcomes), they contribute to a growing body of research designed to provide information to support the development of early interventions.
Supplemental Material

Table S3.1. Definition of the global rating scales of the PInTCI.

<table>
<thead>
<tr>
<th>Category</th>
<th>Scale</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infant</strong></td>
<td>Initiations</td>
<td>Amount and quality of social initiations directed to the parent, either verbal (e.g. vocalizing, babbling) and/or nonverbal (e.g. sharing affect, showing, giving).</td>
</tr>
<tr>
<td></td>
<td>Attentiveness</td>
<td>Amount and quality of 1) infant’s spontaneous orientation to the parent, not elicited by parental behaviour; and 2) infant’s responsiveness to parental behaviour, either positive or negative.</td>
</tr>
<tr>
<td></td>
<td>Shared affect</td>
<td>Amount and quality of the infant’s sharing and directing affective states with/to the parent. Affect can be either positive or negative, but must be shared with the parent.</td>
</tr>
<tr>
<td></td>
<td>Positive affect</td>
<td>Amount and quality of infant’s positive affect, e.g. relaxed body language, smiles, laughs, giggles, happiness, enthusiasm, excitement, positive vocalizations, positive facial expressions.</td>
</tr>
<tr>
<td></td>
<td>(Absence of) Negative affect</td>
<td>Amount and quality of infant’s negative affect, e.g. body language (i.e. tension, discomfort, restlessness), showing anger, dislike, or hostility, negative facial expressions, negative vocalizations, negative bodily gestures (e.g. distress, rejection).</td>
</tr>
<tr>
<td><strong>Parent</strong></td>
<td>Sensitive responsiveness</td>
<td>1) The accuracy of identification and interpretation of the infant’s cues or needs, and 2) the timing and appropriateness of the parent’s responses to these cues.</td>
</tr>
<tr>
<td></td>
<td>(Absence of) Negative control</td>
<td>The extent to which the interaction is determined by the infant’s preferences and the infant’s focus of attention, or whether the parent mainly determines the course of the interaction in a directive, controlling, and/or intrusive way.</td>
</tr>
<tr>
<td></td>
<td>Scaffolding</td>
<td>The level of adequately facilitating the infant’s development and guiding the infant’s actions so that the child can do and say things that he/she would likely not achieve without guidance and encouragement.</td>
</tr>
<tr>
<td></td>
<td>Positive affect</td>
<td>Amount and quality of parent’s positive affect, e.g. positive tone of voice, enthusiasm, smiles/laughter, happy facial expressions, relaxed body posture, and physical affection toward the infant.</td>
</tr>
<tr>
<td></td>
<td>(Absence of) Negative affect</td>
<td>Amount and quality of parent’s negative affect, e.g. negative tone of voice, tightened or angry facial expressions, tense body posture and angry or hostile acts.</td>
</tr>
<tr>
<td><strong>Dyad</strong></td>
<td>Dyadic reciprocity</td>
<td>The amount and quality of engagement, mutuality, cooperation, reciprocity, and sharedness between parent and infant.</td>
</tr>
</tbody>
</table>

Note. PInTCI = Parent-Infant/Toddler Coding of Interaction.
Table S3.2. Intra-class correlation coefficients (ICC) between coders for PInTCI constructs.

<table>
<thead>
<tr>
<th></th>
<th>Inter-rater reliability ICC at 5 months (N=12)</th>
<th>Inter-rater reliability ICC at 10 months (N=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infant</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiations</td>
<td>0.42 (-0.33 - 0.81)</td>
<td><strong>0.68</strong> (0.24 - 0.90)</td>
</tr>
<tr>
<td>Attentiveness</td>
<td><strong>0.84</strong> (0.62 - 0.95)</td>
<td><strong>0.95</strong> (0.87 - 0.98)</td>
</tr>
<tr>
<td>Sharing of Affect</td>
<td><strong>0.67</strong> (0.25 - 0.89)</td>
<td><strong>0.81</strong> (0.54 - 0.94)</td>
</tr>
<tr>
<td>Positive Affect</td>
<td><strong>0.92</strong> (0.82 - 0.97)</td>
<td><strong>0.90</strong> (0.76 - 0.97)</td>
</tr>
<tr>
<td>(Absence of) Negative Affect</td>
<td><strong>0.86</strong> (0.68 - 0.95)</td>
<td><strong>0.83</strong> (0.59 - 0.95)</td>
</tr>
<tr>
<td><strong>Parent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitive Responsiveness</td>
<td><strong>0.86</strong> (0.68 - 0.96)</td>
<td><strong>0.80</strong> (0.52 - 0.94)</td>
</tr>
<tr>
<td>(Absence of) Negative Control</td>
<td><strong>0.88</strong> (0.72 - 0.96)</td>
<td><strong>0.85</strong> (0.64 - 0.95)</td>
</tr>
<tr>
<td>Scaffolding</td>
<td><strong>0.75</strong> (0.43 - 0.92)</td>
<td><strong>0.84</strong> (0.61 - 0.95)</td>
</tr>
<tr>
<td>Positive Affect</td>
<td><strong>0.90</strong> (0.76 - 0.97)</td>
<td><strong>0.91</strong> (0.78 - 0.97)</td>
</tr>
<tr>
<td>(Absence of) Negative Affect</td>
<td>0.31 (-0.58 - 0.77)</td>
<td><strong>0.67</strong> (0.21 - 0.89)</td>
</tr>
<tr>
<td><strong>Dyad</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyadic Reciprocity</td>
<td><strong>0.91</strong> (0.80 - 0.97)</td>
<td><strong>0.89</strong> (0.71 - 0.96)</td>
</tr>
</tbody>
</table>

**Notes.** Bold values indicate good to excellent ICC (≥ 0.60). PInTCI = Parent-Infant/Toddler Coding of Interaction.

Table S3.3. Spearman correlation coefficients for the PInTCI scales between the 5- and 10-month time point by risk status.

<table>
<thead>
<tr>
<th></th>
<th>HR siblings</th>
<th>LR siblings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infant</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiations</td>
<td>-0.24</td>
<td>-0.04</td>
</tr>
<tr>
<td>Attentiveness</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>Sharing of Affect</td>
<td>0.22</td>
<td>0.13</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>0.06</td>
<td>0.36*</td>
</tr>
<tr>
<td>(Absence of) Negative Affect</td>
<td>0.12</td>
<td>-0.07</td>
</tr>
<tr>
<td><strong>Parent</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitive Responsiveness</td>
<td>0.23</td>
<td>0.14</td>
</tr>
<tr>
<td>(Absence of) Negative Control</td>
<td>0.30*</td>
<td>0.31*</td>
</tr>
<tr>
<td>Scaffolding</td>
<td>0.33*</td>
<td>0.15</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>0.36*</td>
<td>0.50**</td>
</tr>
<tr>
<td>(Absence of) Negative Affect</td>
<td>-</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Dyad</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyadic Reciprocity</td>
<td>0.14</td>
<td>0.19</td>
</tr>
</tbody>
</table>

* p<0.05, ** p<0.01

**Notes.** Italic values indicate that data were removed from analyses for methodological reasons. The correlation coefficient of (Absence of) Negative Affect for the HR group could not be determined due to low variance. PInTCI = Parent-Infant/Toddler Coding of Interaction; HR = high-risk; LR = low-risk.
Table S3.4. Global ratings on the PInTCI by time point and risk status.

<table>
<thead>
<tr>
<th></th>
<th>5 months</th>
<th></th>
<th>10 months</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR siblings</td>
<td>LR siblings</td>
<td>HR siblings</td>
<td>LR siblings</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Range</td>
<td>Mean (SD)</td>
<td>Range</td>
</tr>
<tr>
<td>Infant</td>
<td>2.02 (0.80)</td>
<td>1-4</td>
<td>2.28 (0.90)</td>
<td>1-5</td>
</tr>
<tr>
<td>Attentiveness</td>
<td>3.90 (0.94)</td>
<td>2-6</td>
<td>4.00 (1.06)</td>
<td>2-6</td>
</tr>
<tr>
<td>Sharing of Affect</td>
<td>2.21 (1.40)</td>
<td>1-7</td>
<td>2.34 (1.43)</td>
<td>1-6</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>3.11 (1.32)</td>
<td>1-7</td>
<td>3.23 (1.39)</td>
<td>1-7</td>
</tr>
<tr>
<td>(Absence of)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Affect</td>
<td>5.84 (1.24)</td>
<td>2-7</td>
<td>5.85 (1.22)</td>
<td>2-7</td>
</tr>
<tr>
<td>Parent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitive Responsiveness</td>
<td>4.37 (1.22)</td>
<td>2-7</td>
<td>4.34 (1.11)</td>
<td>2-7</td>
</tr>
<tr>
<td>(Absence of)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Control</td>
<td>4.66 (1.07)</td>
<td>2-7</td>
<td>4.74 (1.51)</td>
<td>2-7</td>
</tr>
<tr>
<td>Scaffolding</td>
<td>4.13 (1.09)</td>
<td>1-6</td>
<td>4.13 (1.04)</td>
<td>2-6</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>4.89 (1.34)</td>
<td>1-7</td>
<td>5.11 (1.46)</td>
<td>2-7</td>
</tr>
<tr>
<td>(Absence of)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative Affect</td>
<td>6.94 (0.51)</td>
<td>3-7</td>
<td>6.91 (0.28)</td>
<td>6-7</td>
</tr>
<tr>
<td>Dyad</td>
<td>3.47 (1.13)</td>
<td>1-6</td>
<td>3.53 (1.06)</td>
<td>1-6</td>
</tr>
</tbody>
</table>

Notes. Analyses were carried out on the transformed z-values, but to facilitate interpretation the mean scores in this table were given in raw scores. Ratings from 1 to 7, with 1 reflecting maladjusted/negative behavior while a score of 7 reflects more adaptive/optimal behavior. *Italic* values indicate that data were removed from analyses for methodological reasons. PInTCI = Parent-Infant/Toddler Coding of Interaction; HR = high-risk; LR = low-risk.
PART II

Early detection and intervention of autism spectrum disorder
Sustainability of an Early Detection Program for Autism Spectrum Disorder over the Course of 8 Years

Pijl\textsuperscript{1,2}, M.K.J., Buitelaar\textsuperscript{1,2}, J.K., de Korte\textsuperscript{1,2}, M.W.P., Rommelse\textsuperscript{1,3}, N.N.J., and Oosterling\textsuperscript{1,2}, I.J.

*Autism, 2018; 22 (8), 1018-1024.*

\textsuperscript{1}Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, the Netherlands
\textsuperscript{2}Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, the Netherlands
\textsuperscript{3}Radboud University Medical Centre, Department of Psychiatry, Nijmegen, the Netherlands
Abstract

The importance of early detection of autism spectrum disorder (ASD) followed by early intervention is increasingly recognized. This quasi-experimental study evaluated the long-term effects of a program for the early detection of ASD (consisting of training of professionals and use of a referral protocol and screening instrument), to determine whether the positive effects on the age at referral were sustained after the program ended, while controlling for overall changes in the number of referrals. Before, during, and after the program, the proportion of children referred before 3 years (versus 3-6 years) of age was calculated for children subsequently diagnosed with ASD (N=513) or another, non-ASD, condition (N=722). The odds of being referred before 3 years of age was higher in children with ASD than in children with another condition during the program than before (3.1, 95% CI 1.2-7.6) or after (1.7, 95% CI 1.0-3.0) the program, but was not different before versus after the program. Thus, although the program led to earlier referral of children with ASD after correction for other referrals, the effect was not sustained after the program ended. This study highlights the importance of continued investment in the early detection of ASD.
Autism spectrum disorder (ASD) can be reliably diagnosed before 3 years of age and the diagnosis is remarkably stable in both clinical (Kleinman et al., 2008; Lord et al., 2006; van Daalen et al., 2009) and high-risk samples (Brian et al., 2015; Ozonoff et al., 2015). Over the last years, findings have emphasized the importance of early identification and treatment of ASD to children’s adaptive, social, and cognitive functioning (Zwaigenbaum, Bauman, Choueiri, et al., 2015). Because of the dynamic and plastic nature of the brain, very early interventions may alter the course of brain and behavioral development in children with ASD (Webb et al., 2014), and recent studies have provided evidence of the long-term effectiveness of early interventions (Estes, Munson, et al., 2015; Pickles et al., 2016). In addition, family and society costs for subsequent professional services may be reduced (Penner et al., 2015; Peters-Scheffer, Didden, Korzilius, & Matson, 2012).

Despite the benefits of early detection, the mean age at diagnosis in daily clinical practice still lies around the late preschool years or even later. A review of 42 studies published from January 1999 through March 2012 found that the mean age ranged from 38 to 120 months (Daniels & Mandell, 2014). In order to lower the age at diagnosis, strategies for the early detection of ASD have been developed and promising results have been obtained in some non-randomized studies, showing positive effects on age at diagnosis (Chakrabarti et al., 2005; Holzer et al., 2006; Koegel et al., 2006), percentage of early identified ASD cases (Swanson et al., 2014), and self-efficacy of primary care providers (Mazurek et al., 2017). In a study that used a control region, Oosterling, Wensing, et al. (2010) examined the effect of a screening approach for the early detection of ASD that was integrated in routine developmental surveillance in a specific region of the Netherlands. This early detection program involved (a) training of primary care providers to recognize early signs of autism, (b) use of a specially designed referral protocol that included the Early Screening of Autistic Traits questionnaire (ESAT - Dietz, Swinkels, van Daalen, van Engeland, & Buitelaar, 2006; Swinkels et al., 2006), and (c) formation of a multidisciplinary diagnostic team at the regional psychiatric academic center. After implementation of the screening program, the mean age at ASD diagnosis decreased significantly by 19.5 months to an average of 63.5 months. In the control region (without active investment in early detection of ASD), the mean age at diagnosis did not change significantly. Although this study provided valuable insight into strategies to improve the early detection of ASD, the question remained whether positive effects are sustained in the long term.

To our knowledge, only one study has investigated the long-term effects of an early detection program (Holzer et al., 2006). This non-randomized study examined the

---

2 Currently known as the CoSoS (Communication and Social development Signs).
effects of a 2-year program which contained the following aspects: (a) familiarizing primary care providers with early developmental problems, (b) informing pediatricians and general health practitioners about a screening tool (Checklist for Autism in Toddlers (CHAT) - Baron-Cohen et al., 1996), and (c) performing a diagnostic assessment by a child neurologist, child psychiatrist, and/or child psychologist. With the program, the mean age at diagnosis decreased by 1.5 years, but this effect was not sustained after the program ended. These findings emphasize the importance of a maintenance strategy. Additional understanding of long-term effects can be useful to help policy makers in developing effective programs that will have a permanent effect.

The aim of this study was to evaluate the sustainability of the integrated early detection program developed by Oosterling, Wensing, et al. (2010). The age at referral was recorded of children (aged 0-6 years) subsequently diagnosed with ASD (N=513) or a non-ASD condition (N=722) before, during, and after the early detection program. The program ended when financial and staffing support ended, which affected the training of primary care personnel in the recognition of the early signs of ASD and in use of the referral protocol. The multidisciplinary diagnostic team continued to function. We wanted to determine whether the positive effect on age at referral was sustained after the program ended.

**Methods**

**Study design and setting**

To facilitate the early detection of ASD in the Netherlands, our group used an integrated early detection program (Oosterling, Wensing, et al., 2010), which was approved by the Dutch ethical committee. In the current follow-up study, we performed a natural examination of the age at referral before (January – December 2003), during (January 2004 – December 2006), and 2 years after (January 2009 – December 2011) the early detection program. Even after the early detection program ended, the multidisciplinary diagnostic team at Karakter Child and Adolescent Psychiatry University Center continued to provide highly specialized mental health care for infants and toddlers in the region. However, the lack of funding and staff meant that no specific effort was put into training primary care providers to use the screening protocol, including the ESAT/CoSoS. The control region was not included in this follow-up study because we did not receive permission from one of the institutions to use their data.

This study must be viewed in the context of the healthcare setting in the Netherlands. In general, children of all ages can be referred to psychiatric assessment
centers by general practitioners, medical specialists (e.g. neurologist or pediatrician), professionals from other mental health services or institutions for language development, and primary care providers. Primary care providers can be the doctors of well-baby clinics or members of specific infant-toddler development teams. These independent infant-toddler development teams work in close collaboration with the doctors and nurses of well-baby clinics and provide parents who may have specific concerns about their child’s development with easily accessible first-line care. They also carry out case management, investigate children’s developmental problems in general, and when necessary refer children to secondary or tertiary health services for diagnostic assessment and treatment. To the best of our knowledge, no changes in the healthcare policies or healthcare system (e.g. referral strategies) occurred during the time frame of the study.

Participants
The sample included N=1235 infants, toddlers, and preschoolers (aged 0–6 years) who were referred for clinical psychiatric evaluation before (N=119), during (N=531), or after (N=585) the early detection program was implemented. Of these, 38%, 47%, and 37% were newly diagnosed with ASD as compared to non-ASD diagnoses, respectively. Children with a non-ASD diagnosis had other diagnoses (including absence of a psychiatric diagnosis). This group reflected the general population of referrals in Dutch child and adolescent psychiatry settings, including two thirds of the non-ASD referrals having externalizing problems and a minority having internalizing or other problems (see Table 4.1 for more demographics).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at referral (years)</td>
<td>4.1 (1.6)</td>
<td>3.7 (1.6)</td>
<td>4.4 (1.5)</td>
</tr>
<tr>
<td>Male</td>
<td>75.6</td>
<td>74.8</td>
<td>77.1</td>
</tr>
<tr>
<td>Diagnosis ASD</td>
<td>37.8</td>
<td>47.1</td>
<td>37.3</td>
</tr>
</tbody>
</table>

Notes. ASD = autism spectrum disorder; SD = standard deviation.

Early detection program
The early detection program comprised three components: (a) training of primary care providers to recognize early signs of autism, (b) use of a systematic screening protocol including the ESAT/CoSoS, and (c) formation of a multidisciplinary diagnostic team. The training sessions aimed to raise awareness and familiarize trainees with ASD (especially its early signs) and the use of the referral protocol. In total 39 sessions were delivered,
of which 22 were for primary care providers, because they are most likely to refer young children at risk of ASD. The remaining 17 sessions were attended by other health professionals (e.g. general practitioners, speech-and-language therapists). Attendance was compulsory for primary care workers, who were awarded CME (Continuing Medical Education) points. In the program, professionals were required to administer the ESAT/CoSoS (with the assistance of the parents) before they referred children younger than 36 months for assessment of ASD. This 14-item questionnaire focuses on early social communication skills and restricted and repetitive behaviors in children younger than 36 months (Dietz et al., 2006; Swinkels et al., 2006). Children who failed three or more items were considered to be at risk and underwent further assessment. For those children that screened negative with the ESAT/CoSoS (failing <3 items), the referring professional had to provide additional information showing that the child was at risk. Within 2 weeks of referral, families were invited to the Karakter Child and Adolescent Psychiatry University Center. A multidisciplinary team for infant psychiatry, specialized in the early diagnosis of ASD, carried out the diagnostic assessment of all referrals. More information about the early detection program can be obtained from the last author.

Measures
Age at referral. To identify the long-term effects of the early detection program, the age at referral was compared before, during, and after the program. Data were retrieved from electronic medical records. Since the early detection program focused on children younger than 36 months of age, a differentiation was made between referrals before 3 years of age (0–35 months) and referrals 3 to 6 years of age (36–83 months).

Cognitive functioning. IQ scores were assessed with age-appropriate measures, most frequently the Mullen Scales of Early Learning (Mullen, 1995), the Psycho-Educational Profile-Revised (Schopler, Reichler, Bashford, Lansing, & Marcus, 1990), the Snijders-Oomen Non-verbal Intelligence test (Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998), and the Wechsler tests (Wechsler, 1997, 2002). The level of cognitive functioning was categorized in three groups: <70, 70–89, ≥90.

Data analysis
Binary logistic regression modeling was used to evaluate the long-term effects of the early detection program after it ended, while controlling for overall changes in the number of referrals. The effect of diagnosis (ASD, non-ASD), time point (before, during, after the program), and the interaction diagnosis x time point on the likelihood that children younger than 3 years (versus children aged 3–6 years) were referred was investigated.

Binary logistic regression modeling was also used to investigate the effect of other potential predictors (i.e. cognitive functioning, and sex) on the age at referral of children diagnosed with ASD, using IQ (<70, 70–89, ≥90), sex (male, female), and time point as
potential predictors. The interaction effects IQ x time point and sex x time point were also included in the model. If the interaction or main effects were non-significant, predictors were dropped from the final model.

Information about the level of cognitive functioning was missing in 8.8% (n=45) of the ASD cases (cognitive data was unavailable for children with non-ASD diagnoses). Multiple imputation with the expectation maximization algorithm was used to account for the missing data in the group of children with ASD.

**Results**

The use of multiple imputation only marginally changed outcomes and did not alter conclusions. Below, we present the results with multiple imputation.

**Change in proportion of children referred before 3 years**

Figure 4.1 shows the change in proportion of children referred before 3 years of age by diagnosis and time point. Binary logistic regression analyses revealed that there was a significant interaction effect for diagnosis x time point on the age at referral ($\chi^2(2) = 7.90, p = 0.019$). The odds of being referred before 3 years for children diagnosed with ASD versus a non-ASD condition was significantly higher during the program than before (3.1, 95% CI 1.2–7.6, $p < 0.05$) or after (1.7, 95% CI 1.0–3.0, $p < 0.05$) the program, but not before versus after the program (1.8, 95% CI 0.7–4.5, $p > 0.05$).

![Figure 4.1. Percentage of children referred before 3 years of age by diagnosis and time point with error bars representing 95% confidence intervals.](image)
Predictors of age at referral for children with an ASD diagnosis

Regarding ASD participants, there was no significant overall interaction effect IQ x time point ($\chi (4) = 5.43, p = 0.25$), which implied that the effect of cognitive functioning on age at referral was not related to the effect of the program. Nonetheless, the main effect of cognitive functioning on the age at referral was significant ($\chi (2) = 72.50, p < 0.001$). The odds of being referred before 3 years was 2.6 times (95% CI 1.6–4.2, $p < 0.001$) and 10.0 times (95% CI 5.9–16.9, $p < 0.001$) higher if a child with ASD had an IQ < 70 as compared to an IQ between 70 and 89 and an IQ > 90, respectively. The odds ratio for being referred was 3.9 times (95% CI 2.3–6.5, $p < 0.001$) higher for children with ASD with an IQ between 70 and 89 than for children with ASD with an IQ > 90. No significant effects of sex were found. See Figure 4.2 for the percentage of children with ASD referred before 3 years versus between 3 and 6 years by level of cognitive functioning at all three time points.

![Figure 4.2. Percentage of children with autism spectrum disorder at all time points by age at referral and level of cognitive functioning with error bars representing 95% confidence intervals.](image)

Discussion

This is the first quasi-experimental study to investigate the sustainability of the effect of an integrated screening program for the early detection of ASD after the program ended. Results clearly underline the need for continued investment: the odds of being referred before 3 years of age for children with ASD versus non-ASD was higher when the early detection program was implemented than before or after
its implementation, but not before versus after the implementation of the program. Thus, although the early detection program led to earlier referral of children with ASD when corrected for other referrals, the effect was not sustained after the program ended. At all time points, children with intellectual disabilities were more likely to be referred before 3 years of age than were children with a higher level of cognitive functioning.

**Lack of sustainable effects and how to deal with it**

The findings suggest that the early detection program fulfilled its objective to improve the early detection of ASD over the years of active investment, but the effect disappeared after active investment had faded out. This is in line with the study of Holzer et al. (2006), who showed that a decrease in mean age at diagnosis was not sustained 2 years after implementation. In the current study, the lack of financial support was the major barrier, as was also noted by Zwaigenbaum, Bauman, Fein, et al. (2015) as one of the key barriers to the early detection of ASD. The lack of financial support ended the intensive collaboration between primary and specialist care (e.g. no prompts were given to the primary care providers), which may have resulted in a lack of awareness among primary care providers. Additionally, no training sessions were provided after the program ended, which meant that experience and knowledge about early detection were lost if staff left. Continuation of the program, with training of personnel, would be expensive. Given that the financing of mental health care is generally under pressure, the question is how to develop effective strategies that require minimal financial support but which have a permanent impact.

In order to improve the early detection of ASD, collaboration within and between countries, with the sharing of knowledge, is essential and will help to develop a standardized approach to the detection of ASD. As such, the European network COST Action ‘Enhancing the Scientific Study of Early Autism’ [ESSEA] was initiated with a view to increasing knowledge about the early signs of autism, by combining techniques from cognitive neuroscience with those from clinical sciences, and by reviewing the state of art of early identification (Garcia-Primo et al., 2014) and intervention (McConachie, Fletcher-Watson, & Working Group, 2015; Salomone et al., 2016) practices in Europe. The COST-ESSEA network inspired Dutch researchers and clinical experts in the field of early autism to form a national interdisciplinary network in 2013. This network promotes collaboration in scientific studies, the exchange of knowledge, and the translation of knowledge into practical tools (e.g. for use in primary care practice). The network operates in close collaboration with primary care providers and the parents of children with ASD as experience experts. As previously suggested by Miller et al. (2011), we strongly believe that intensifying
the collaboration between autism experts, primary care providers, and parents will facilitate the timely referral of children at risk of ASD.

Another key element in the early detection of ASD is the availability and accessibility of knowledge for both parents and healthcare professionals. In the Netherlands, a state-of-the-art online platform is currently being developed which will provide parents and professionals with red flag early signs of ASD and information about the diagnosis and relevant services. Also, e-learning and live-online-learning (LoL) modules are being developed. The e-learning module teaches professionals about the early signs of ASD whereas the LoL module involves the discussion of specific case studies of the professionals themselves in a virtual online classroom and with the support of an experienced clinician (Oosterling et al., 2016; Van ‘t Hof, Bailley, Hoek, & Ester, 2017). In the USA, similar collections of Web-based tools and courses are available (http://www.autismnavigator.com and http://www.cdc.gov/ncbddd/autism/actearly/), including a specific course about early signs of ASD. These trends imply that e-health technology can have a pivotal role in raising awareness and maintaining the effects of early detection programs without incurring high costs. Research into the effect of these methods is required to evaluate their effectiveness.

The role of cognitive functioning in early detection
Regardless of the time point, children with intellectual disabilities were more likely to be referred for assessment before 3 years of age than were children with a higher level of cognitive functioning. Previous research has shown that children with cognitive impairment (IQ<70) are younger at the time of ASD diagnosis than higher functioning children (Mazurek et al., 2014; Shattuck et al., 2009). Children with concurrent ASD and cognitive impairment may have more obvious behavioral and developmental challenges, which increase the likelihood of early detection. Vice versa, high-functioning children may be better able to compensate for their difficulties and/or may show more subtle impairments at a preschool age. This is now also acknowledged in the DSM-5, which states that “deficits may not become fully manifest until social demands exceed limited capacities” (American Psychiatric Association, 2013). On the one hand, given their more subtle symptoms, an early diagnosis is likely to be missed in these children and consequently they will not benefit from early intervention. On the other hand, it may simply not be possible to detect ASD in this group at an early stage, and attempts to do so might result in a high number of false positives. This argues for repeat evaluation in early childhood around 4–6 years of age, when potential problems in social relationships with peers start to emerge.
Strengths and limitations

This study had several strengths, including the embedding in a naturalistic context, a large sample size, and the use of a non-ASD group. However, it also had some limitations. To the best of our knowledge, there were no changes in the healthcare system or policy during the period covered by the study, although there may have been unknown changes in clinical practice (e.g. availability of healthcare providers) that influenced referral practices and hence the age at referral. Ideally, long-term effects should be examined in a controlled study including the same regions as in the original study, but we did not receive permission from one of the institutions in the control region to use their data. However, by comparing referrals later diagnosed with ASD with referrals diagnosed with other (non-ASD) conditions, the effect of the early detection program could be examined while controlling for the most important potential confounders. Although our findings provide evidence for the effectiveness of the program, the question remains whether children that were referred before 3 years of age had better outcomes than children referred later. For an ultimate justification of early detection, the long-term effects on child development and quality of life should be investigated. Therefore, future studies should longitudinally follow up children from the general population who are randomly assigned to undergo (or not) a screening program, regardless of whether they screen positive or negative with regard to the likelihood of ASD. In addition, the effectiveness of early interventions for young children identified by means of an early detection program should be further investigated.

Conclusion and clinical implications

In conclusion, although the program improved the early detection of ASD, its effect was not sustained when the financial support needed to train health professionals ended. These findings highlight the importance of maintaining early detection through continuous investment in active screening and ongoing training of primary care providers. Given the evidence that screening programs can detect autism in an early stage, long-term investment merits a high place on the political agenda. Policy makers and healthcare managers should consider specific strategies to overcome barriers when implementing (and maintaining the effects of) early detection programs. Although labels are needed to refer children to appropriate services, the ultimate goal of early detection is to identify early signs, discuss concerns with families, and guide parents in how to best support their child. Instead of a ‘wait-and-see’ approach, we strongly argue that follow-up services and appropriate interventions be made available for at-risk children.
Does the Brief Observation of Social Communication Change (BOSCC) help moving forward in measuring change in early autism intervention studies?

Pijl¹,2, M.K.J., Rommelse¹,3, N.N.J., Hendriks¹,2, M., de Korte¹,2, M.W.P., Buitelaar¹,2, J.K. and Oosterling¹,2, I.J.


¹Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, the Netherlands
²Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Department of Cognitive Neuroscience, Nijmegen, the Netherlands
³Radboud University Medical Centre, Department of Psychiatry, Nijmegen, the Netherlands
Abstract

The field of early autism research is in dire need of outcome measures that adequately reflect subtle changes in core autistic behaviors. This article compares the ability of a newly developed measure, the Brief Observation of Social Communication Change (BOSCC), and the Autism Diagnostic Observation Schedule (ADOS) to detect changes in core symptoms of autism in 44 toddlers. Results provide encouraging evidence for the BOSCC as a candidate outcome measure, as reflected in sufficient inter- and intra-rater reliability, independency from other child characteristics, and sensitivity to capture change. Although the BOSCC did not evidently outperform the ADOS on any of these quality criteria, the instrument may be better able to capture subtle, individual changes in core autistic symptoms. The promising findings warrant further study of this new instrument.
The choice of outcome measures in early autism research is highly varied and there is little consensus about which measures adequately monitor change over time (Anagnostou et al., 2014; Bolte & Diehl, 2013; Cunningham, 2012; McConachie, Parr, et al., 2015). In their review, McConachie, Parr, et al. (2015) identified 131 different tools to measure change in young children with autism. This enormous number of outcome measures has prevented autism research from successfully comparing the outcomes of intervention studies and has compromised conclusions regarding the most effective type and intensity of intervention for children diagnosed with autism spectrum disorder (ASD). There is a critical need for outcome measures to be standardized, which will help clinicians to identify which infant will benefit the most from a particular intervention. This study evaluates the utility of a new tool designed to measure change in core autistic behavior: the Brief Observation of Social Communication Change (BOSCC; Grzadzinski et al., 2016).

The highly varied use of outcome measures reflects the dearth of standardized tools that capture change in core autistic symptoms. Given the abstract nature of ASD (American Psychiatric Association, 2013), it is challenging to design an instrument that is able to translate the conceptual descriptions of the disorder into concrete behaviors. This requires a thorough understanding of the differences between normal versus abnormal communication and interpretation of behavior. For this reason, it would seem ‘logical’ to evaluate the effectiveness of interventions based on a standardized, direct observation of core ASD symptoms by a trained coder rather than based on a parent or teacher report. This approach was adopted in several early treatment studies that used the Autism Diagnostic Observation Schedule (ADOS - Lord, Rutter, DiLavore, & Risi, 1999; Lord et al., 2012) to measure outcomes (e.g. Estes, Munson, et al., 2015; Green et al., 2010; Oosterling, Visser, et al., 2010). However, the ADOS was not originally designed to measure change over time and has drawbacks when used to do so (Lord et al., 1999). A limitation of earlier versions of the ADOS was the use of different modules dependent on an individual’s developmental level (Lord et al., 1999), which made it difficult to compare ADOS scores across modules and over time. Therefore, revised algorithms and a severity metric were created (Gotham et al., 2009; Gotham et al., 2008; Gotham, Risi, Pickles, & Lord, 2007). The validity of the ADOS as an instrument to measure autism severity was improved (de Bildt et al., 2011; Shumway et al., 2012), which enabled researchers and clinicians to disentangle ASD-specific factors from non-ASD-specific factors, such as language or cognitive development. In order to evaluate the specific effect of an intervention, it is essential that an instrument can distinguish changes in core symptoms from changes in general abilities. However, a remaining limitation of the ADOS is the narrow range of scores used for each item, which limits the ability to dimensionalize severity of symptoms and the ability to detect...
subtle changes. Although the ADOS can identify changes in ASD symptoms over several years (Estes, Munson, et al., 2015; Gotham et al., 2012), it is less effective in detecting subtle changes over shorter periods of time (Dawson et al., 2010; Estes, Munson, et al., 2015).

Lord, Grzadzinski and colleagues developed the BOSCC to circumvent the limitations of commonly used measures (Grzadzinski et al., 2016). While this instrument contains items comparable to those of the ADOS (Lord et al., 2012), in that both instruments focus on core symptoms, there are two main differences between the BOSCC and the ADOS. First, the setting in which behavior is observed can differ. BOSCC scores are based on the observation of social communicative behavior during naturalistic interactions between a child and an adult (e.g. parent and child or professional and child), whereas ADOS scores are based on a standardized, semi-structured professional–child interaction. Second, the scoring scale of the BOSCC is broader in order to capture more nuanced variations in behavior. Instead of a score range of 0–2 used in the ADOS, the score range of the BOSCC is 0–5, with higher scores being indicative of more atypical behavior. Thus, the ADOS can be used to assess whether an individual meets a diagnostic threshold and helps to identify that individual’s current clinical status, whereas the BOSCC has been designed to measure subtle changes in autistic behaviors over time.

The BOSCC would seem a promising instrument to study change in early autism – it focuses on core ASD symptoms, can be scored blind, and is based on direct observation rather than on parent or teacher report, thereby minimizing subjectivity. However, before the BOSCC can be used in intervention studies, its measurement properties, that is, reliability, validity, and responsiveness to change, need to be established. There is preliminary evidence of the instrument’s usefulness in early autism research (Grzadzinski et al., 2016), with inter-rater and test-retest reliability being adequate for assessing a child’s behavior during parent–child interactions. As compared to a no change alternative, the instrument was able to capture changes in the social communicative behaviors of young children with ASD (N=56) over time, which the ADOS severity metric did not. In addition, Kitzerow, Teufel, Wilker, and Freitag (2015), who used the BOSCC to score behavior during an ADOS assessment (N=21), also showed that the BOSCC was sensitive to change, but not more so than the ADOS. These findings highlight the need for more research into the utility and efficacy of the BOSCC, applied to the heterogeneous group of young children with ASD, in different interactional contexts and in relation to other ASD measures.

The aim of this study was to investigate the utility of a preliminary version of the BOSCC as instrument to measure outcomes in early autism intervention studies. Given the limitations of the ADOS, we investigated whether the BOSCC has greater potential for measuring change in core autistic behavior than the ADOS. The
main objective of this study was to compare the BOSCC and the ADOS, aiming at:
1) investigating inter- and intra-rater reliability; 2) providing preliminary evidence
for construct validity (based on correlational analyses with theoretical similar and
dissimilar measures); and 3) examining sensitivity to change. Our a priori hypothesis
was that the BOSCC and the ADOS would show a comparable reliability and validity,
but that the BOSCC would be better able to detect change because its broader
range of scores.

Methods

Participants
This study re-used data collected by Oosterling, Visser, et al. (2010) who investigated
the efficacy of a parent training intervention in a randomized controlled trial involving
toddlers diagnosed with ASD. All children received some form of intervention, either
the treatment condition in which parental skills were promoted or care as usual (see
Oosterling, Visser, et al., 2010 for details regarding intervention type). Participants
were recruited during the period 2004–2007 at Karakter Child and Adolescent
Psychiatry University Center Nijmegen, the Netherlands. The study was approved
by the local ethics board and informed consent was obtained for each individual
participant included in the study. All participants had screened positive on the Early
Screening of Autistic Traits Questionnaire3 (ESAT - Swinkels et al., 2006) and met
criteria for autism according to a consensus diagnosis made by a child psychiatrist and
psychologist. This consensus diagnosis was based on an extensive assessment process,
including direct observations of child behavior, cognitive and language testing, and
parental information.

Of the final sample included in the study of Oosterling, Visser, et al. (2010) on an
intention-to-treat basis (n=67), the data of 44 participants were re-analyzed in this
study. For the remaining 23 participants (34.3%) no video data of parent-child play
sessions, on which the BOSCC coding was based (see measures), and/or no ADOS
data were available at two time points. Given that the effect of the intervention itself
was not the focus of the current study, data from the intervention and control groups
were pooled. See Table 5.1 for detailed demographics of the included participants.

Measures and procedures
BOSCC. A preliminary research version of the BOSCC (February 2014 version) that
intends to measure subtle change in core autistic behaviors was used (Grzadzinski et

3 Currently known as the CoSoS (Communication and Social development Signs).
al., 2016). This version contains 16 items associated with key features of ASD, such as making eye contact or the frequency and function of social overtures. Items were scored on a scale ranging from 0 to 5, with higher scores reflecting more atypical behaviors. The BOSCC provides an ASD total score for items 1 to 13, which reflect key autistic behaviors. In addition, a social communication (SC) domain score and a restrictive, repetitive behaviors (RRB) domain score can be computed. The present study computed the SC domain score based on items 1 to 8, as was also done in previous studies using a similar version of the BOSCC (Fletcher-Watson et al., 2015; Nordahl-Hansen, Fletcher-Watson, McConachie, & Kaale, 2016). The RRB domain score was based on items 10 to 13 (play, sensory interests, hand/finger mannerisms, and repetitive/stereotyped interests/behaviors). Items 14 to 16 (activity level, disruptive behavior, and anxiety) can be used for determining the validity of the BOSCC assessment, by evaluating whether the observation was representative of a child’s typical behavior. In the present study, Cronbach’s alpha for internal consistency was 0.91 for the BOSCC total score (items 1-13), 0.90 for the SC domain score (items 1-8), and 0.75 for the RRB domain score (items 10-13).

### Table 5.1. Participant characteristics (n=44) at baseline.

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (months)</td>
<td>31.2 (5.3)</td>
<td>77.3</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnosis (DSM-IV)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>90.9</td>
<td></td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>9.1</td>
<td></td>
</tr>
<tr>
<td>Non-verbal IQ</td>
<td>59.8 (21.4)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Native Dutch</td>
<td>90.9</td>
<td>9.1</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Educational level mother</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>38.6</td>
<td>38.6</td>
</tr>
<tr>
<td>Middle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>22.7</td>
<td></td>
</tr>
<tr>
<td>Educational level father</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>44.2</td>
<td>23.3</td>
</tr>
<tr>
<td>Middle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>32.6</td>
<td></td>
</tr>
<tr>
<td>T2-T1 duration (months)</td>
<td>15.7 (2.1)</td>
<td></td>
</tr>
</tbody>
</table>

Notes. SD = standard deviation; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders (4th ed.); PDD-NOS = pervasive developmental disorder—not otherwise specified. * No value for 1 participant.

Videotaped parent-child interactions at baseline and endpoint were evaluated using the BOSCC. In the original study, parents and toddlers had been filmed during a free play interaction, at the clinic (baseline) and at home (endpoint). At both time
points, parents had been instructed to play as usual, without making any additional demands on their child. In order to keep the context as similar as possible across assessments, toy categories (i.e. pretend play materials, construction toys) had been kept constant and the same parent had been involved in the play sessions at both time points. In total, the interaction between parent and child took approximately 15 minutes, of which the first 10 minutes were scored. All clips were scored on the basis of two 5-minute segments. Scores of the two segments were summed and averaged to obtain a total change score.

**BOSCC training.** In this study, the training procedure was as follows. Step 1: The first and last author (MP; IO), both experienced in the assessment and scoring of the ADOS, were informed by Catherine Lord (primary author of the BOSCC and the ADOS) about the use of the BOSCC (COST Enhancing the Scientific Study of Early Autism (ESSEA) training school 2013, Manchester UK). Their scoring was considered reliable based on two criteria formulated by the authors of the instrument: 1) within 1 point for ≥80% of items on a segment; and 2) total change scores within 3 points, with each coder meeting both criteria for three consecutive clips of 10 minutes (5 minutes per scoring session, six segments in total). Step 2: MP and IO then trained two master students (MH; MdK) to score parent–child interactions using the BOSCC. To establish the reliability of these coders (according to the abovementioned criteria), their scores were compared with the consensus scores of the trainers. The scores of the students on the training clips were not included in the core analyses. Step 3: Once the student coders achieved reliable scores with the BOSCC, regular reliability checks with the trainers and group discussions were arranged to prevent drift and to ensure reliable scoring throughout.

**ADOS-Generic.** This ‘gold standard’ diagnostic instrument was used to assess social communicative and repetitive behaviors during an examiner–child interaction (Lord et al., 1999). Both at baseline and endpoint the most appropriate ADOS module was assessed, depending on the child’s language level. ADOS module 1 was used for all children at baseline. At endpoint, seven children (15.9%) switched to ADOS module 2. To examine autism symptoms over time, ADOS calibrated severity scores (CSS) were applied, including one total and two domain severity scores: algorithm total (CSS Total), Social Affect (CSS SA) and Restrictive, Repetitive Behaviors (CSS RRB - Gotham et al., 2009; Hus, Gotham, & Lord, 2014). The ADOS CSS range from 1 to 10, with higher scores indicating greater autism severity.

**Additional outcome measures.** BOSCC and ADOS scores were correlated with scores on the following measures: (1) The MacArthur-Bates Communicative Development Inventory (MCDI - Fenson et al., 1993; Zink & Lejaegere, 2002), which focuses on word production and comprehension; (2) Three age-appropriate psychometric tests, which focus on non-verbal IQ, namely the Mullen Scales of
Early Learning (MSEL - Mullen, 1995), Psycho Educational Profile-Revised (PEP-R - Schopler et al., 1990), and Snijders-Oomen Nonverbal Intelligence Test (SON-R - Tellegen et al., 1998); and (3) the Child Behavior Checklist (CBCL - Achenbach & Rescorla, 2000), which focuses on other behavioral problems. All measures were assessed at baseline and endpoint. More information about data collection and measures can be found in Oosterling, Visser, et al. (2010).

Statistical analyses
First, the reliability of the BOSCC was investigated by calculating inter- and intra-rater reliability for total scores, domain scores, and scores for individual items, based on 20 video segments (11.4%). Inter-rater reliability was established by comparing the scores of the two student coders with the consensus score. In addition, the intra-rater reliability of each coder was calculated, based on repeat scoring after an interval of 2–8 weeks. Intraclass correlation coefficients (ICC) were used to investigate the absolute agreement between and within coders, with values classified as poor (0.00–0.40), fair to good (0.41–0.75), and excellent (>0.75) (Fleiss, 1986). The current study compared the reliability of the BOSCC with the reliability of the ADOS-G, as indicated by Lord et al. (1999; 2000).

Second, as was also done by Grzadzinski et al. (2016), construct validity was assessed by investigating whether cross-sectional scores (i.e. BOSCC domain scores related to ADOS domain scores within one time point) and change scores (i.e. change in BOSCC or ADOS scores related to change in scores for other measures) for the two instruments were correlated. Change scores were determined by calculating the difference between baseline and endpoint. High correlations between theoretically similar measures (e.g. BOSCC and ADOS) and low correlations between theoretically dissimilar measures (e.g. BOSCC and CBCL) were expected. In order to disentangle changes in core symptoms from changes in general abilities, the correlations between the ADOS or BOSCC scores with scores for developmentally sensitive measures (i.e. MCDI, non-verbal IQ, age) should be low. However, some (i.e. low to moderate) association with the MCDI is to be expected, because of the conceptual overlap in language and social communicative behaviors. The Pearson or Spearman correlations were interpreted according to Cohen’s guidelines (Cohen, 1988): 0.10–0.29 (weak), 0.30–0.49 (moderate), and 0.50–1.0 (strong).

Finally, the sensitivity to measure change of the BOSCC and the ADOS CSS was investigated. First, change scores at a group level were assessed using the paired sample t-test or Wilcoxon signed-ranks test (p<0.05), depending on the distribution of the variables. Second, to determine at an individual level whether the change was clinically meaningful, the reliable change index (RCI) was calculated (Jacobson & Truax, 1991; Wise, 2004). RCI values indicate whether a participant’s change score
exceeds that which could reasonably be expected on the basis of measurement error. In order to calculate the RCI, test–retest reliability estimates of the BOSCC (February 2014 version) were applied: 0.84, 0.82, and 0.78 for the BOSCC total, SC and RRB scores, respectively (Grzadzinski et al., personal communication 2015). With regard to the ADOS, the test–retest reliability estimates reported by Lord et al. (2000) were used: 0.71 (calculated mean of the SA and RRB scores), 0.82, and 0.59, for the ADOS total, SA and RRB scores, respectively. In addition, correlational analyses were performed to assess the relationship between reliable change detected with the BOSCC and with the ADOS CSS and to examine their relationship with child characteristics at baseline.

A correction for multiple comparisons was applied for all correlational analyses, using the False Discovery Rate (FDR) controlling procedure with a q-value of 0.05 (Benjamini & Hochberg, 1995). Furthermore, multiple imputation with the expectation maximization algorithm was used to account for missing data. The maximum percentage of missing data was 9.1% (n=4), which was the case for the CBCL subscales at baseline. The MCDI and the CBCL at endpoint had missing data for 2.3% of the sample (n=1). There were no missing data for the BOSCC and ADOS CSS total and domain scores.

Results

The use of multiple imputation only marginally changed outcomes and did not alter conclusions. Below, we present results with multiple imputation.

Reliability
The inter-rater reliability for separate items, domain scores, and total score of the BOSCC was excellent: all ICC values exceeded 0.82. Specifically, ICC values for the total, SC, and RRB scores were 0.99, 0.99, and 0.97, respectively. The intra-rater reliability was fair to excellent for both coders: ICC values ranged from 0.68 to 0.99 (see Supplemental Material).

The inter-rater reliability of the BOSCC was comparable to that of the ADOS-G reported by Lord et al. (2000). In that study, the inter-rater reliability for the ADOS social-communication total was excellent: ICC values ranged from 0.84 to 0.98 across modules. The inter-rater reliability for the ADOS RRB score was also relatively high, ranging from 0.75 to 0.90. Intra-rater reliability was not investigated.
Construct validity

Correlations between BOSCC and ADOS. Correlational analyses were used to investigate the association between BOSCC and ADOS CSS scores at baseline and endpoint. At baseline, the total and RRB scores of the BOSCC and the ADOS were moderately correlated ($r_1 = 0.37$, $p < 0.05$; $r_2 = .36$, $p < 0.05$, respectively), whereas the BOSCC SC score was only weakly correlated with the ADOS CSS SA score ($r_3 = 0.25$, $p > 0.05$). At endpoint, the total, SA/SC, and RRB scores of the BOSCC and the ADOS were moderately correlated ($r = 0.48$, $p < 0.05$; $r = 0.45$, $p < 0.05$; $r = 0.32$, $p < 0.05$, respectively). Post-hoc analyses revealed that the correlations between the BOSCC and the ADOS CSS did not significantly differ between baseline and endpoint (total: $z = -0.69$, $p > 0.05$; SA/SC: $z = -1.10$, $p > 0.05$; RRB: $z = 0.23$, $p > 0.05$). Correlations between changes in BOSCC and ADOS CSS scores are reported in the section ‘Sensitivity to detect change in autistic behaviors’.

Correlations between BOSCC and ADOS change scores and other measures. None of the BOSCC and ADOS CSS total and domain change scores was significantly correlated with the change scores of other measures (details are given in Table 5.2). This implies that BOSCC and ADOS CSS change scores are independent of changes in other measures.

Table 5.2. Pearson or Spearman correlation coefficients between change on BOSCC scores, ADOS CSS, and additional measures.

<table>
<thead>
<tr>
<th></th>
<th>IQ Non-verbal</th>
<th>Age</th>
<th>MCDI Comprehension</th>
<th>MCDI Production</th>
<th>CBCL Internalizing</th>
<th>CBCL Externalizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOSCC total</td>
<td>$r_1 = 0.01$</td>
<td>$r_2 = -0.18$</td>
<td>$r = -0.40$</td>
<td>$r = -0.24$</td>
<td>$r_1 = 0.07$</td>
<td>$r = -0.16$</td>
</tr>
<tr>
<td>BOSCC SC</td>
<td>$r_1 = 0.02$</td>
<td>$r_2 = -0.09$</td>
<td>$r = -0.35$</td>
<td>$r = -0.25$</td>
<td>$r_1 = -0.06$</td>
<td>$r = -0.07$</td>
</tr>
<tr>
<td>BOSCC RRB</td>
<td>$r_1 = -0.04$</td>
<td>$r_2 = -0.18$</td>
<td>$r = -0.09$</td>
<td>$r = 0.06$</td>
<td>$r_1 = 0.35$</td>
<td>$r = -0.07$</td>
</tr>
<tr>
<td>ADOS CSS total</td>
<td>$r_1 = 0.04$</td>
<td>$r_2 = -0.26$</td>
<td>$r = -0.14$</td>
<td>$r = -0.03$</td>
<td>$r_1 = 0.02$</td>
<td>$r = 0.07$</td>
</tr>
<tr>
<td>ADOS CSS SA</td>
<td>$r_1 = 0.04$</td>
<td>$r_2 = -0.30$</td>
<td>$r = -0.18$</td>
<td>$r = -0.05$</td>
<td>$r_1 = 0.04$</td>
<td>$r = 0.11$</td>
</tr>
<tr>
<td>ADOS CSS RRB</td>
<td>$r_1 = 0.18$</td>
<td>$r_2 = -0.09$</td>
<td>$r = 0.13$</td>
<td>$r = 0.16$</td>
<td>$r_1 = -0.07$</td>
<td>$r = -0.23$</td>
</tr>
</tbody>
</table>

Notes. MCDI = MacArthur Communicative Development Inventory; CBCL = Child Behavior Checklist; BOSCC = Brief Observation of Social Communication Change; SC = social communication; RRB = restrictive and repetitive behavior; ADOS = Autism Diagnostic Observation Schedule; CSS = calibrated severity score; FDR = false discovery rate.

All correlations were non-significant after applying the FDR controlling procedure.
Table 5.3. Paired-sample t-test and Wilcoxon signed ranks analyses for BOSCC and ADOS CSS total and domain scores.

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Endpoint</th>
<th>t-Test / Wilcoxon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min.</td>
<td>Max.</td>
<td>Min.</td>
</tr>
<tr>
<td>BOSCC total</td>
<td>1</td>
<td>49</td>
<td>4.5</td>
</tr>
<tr>
<td>BOSCC SC</td>
<td>16</td>
<td>37</td>
<td>27.6 (6.2)</td>
</tr>
<tr>
<td>BOSCC RRB</td>
<td>1.5</td>
<td>14</td>
<td>5.6 (3.0)</td>
</tr>
<tr>
<td>ADOS CSS total</td>
<td>2</td>
<td>10</td>
<td>6.8 (2.2)</td>
</tr>
<tr>
<td>ADOS CSS SA</td>
<td>2</td>
<td>10</td>
<td>7.2 (2.2)</td>
</tr>
<tr>
<td>ADOS CSS RRB</td>
<td>1</td>
<td>10</td>
<td>6.4 (2.5)</td>
</tr>
</tbody>
</table>

* p < 0.05; ** p < 0.01.

Notes. SD: = standard deviation; BOSCC: = Brief Observation of Social Communication Change; SC: = social communication; RRB: = restrictive and repetitive behavior; ADOS: = Autism Diagnostic Observation Schedule; CSS: = calibrated severity score; SA: = social affect.

Sensitivity to detect change in autistic behaviors

The sensitivity of both instruments to detect behavioral change with time was first investigated at a group level. A paired-sample t-test indicated a statistically significant decrease in the total score for core autistic behaviors assessed with the BOSCC (t (43) = 2.53, p < 0.05, d = 0.38; Table 5.3), reflecting improvement over time. This was clearly driven by improvements in the RRB score (z = -3.19, p < 0.01, r = -0.34), since the SC score did not change significantly between baseline and endpoint (z = -1.61, p > 0.05, r = -0.17). A similar pattern emerged with the ADOS: the ADOS CSS total and RRB scores were significantly lower at endpoint than at baseline (z = -2.99, p < 0.01, r = -0.32; z = -2.33, p < 0.05, r = -0.25, respectively), whereas the ADOS CSS SA score did not significantly change over time (z = -1.80, p > 0.05, r = -0.19).
Figure 5.1. Scatter Plot of Reliable Change Index and Endpoint Score for BOSCC Total.

Figure 5.2. Scatter Plot of Reliable Change Index and Endpoint Score for ADOS CSS Total.
Figure 5.3. Scatter Plot of Reliable Change Index and Endpoint Score for BOSCC SA.

Figure 5.4. Scatter Plot of Reliable Change Index and Endpoint Score for ADOS CSS SA.
Chapter 5 | Measuring change in early autism interventions

**Figure 5.5.** Scatter Plot of Reliable Change Index and Endpoint Score for BOSCC RRB.

**Figure 5.6.** Scatter Plot of Reliable Change Index and Endpoint Score for ADOS CSS RRB.
The sensitivity to detect change at an individual level was explored by using RCIs. The RCIs for the total, SC/SA, and RRB scores of the BOSCC and ADOS are shown in Figures 5.1 to 5.6. Jacobson and Truax (1991) recommended using an RCI ± 1.96 \( (p < 0.05 \text{ two-tailed}) \) to indicate a significant change, as is shown by the dotted vertical lines in the graphs. A less conservative criterion \( \text{RCI} \pm 1.28; p < 0.10 \text{ two-tailed} \) is also included in the graphs in order to specify the group that showed no reliable change (as is shown by the ‘mildly improved’ and ‘mildly deteriorated’ groups; Wise, 2004). The RCI values for the BOSCC total score revealed that 20.5% of the children showed a reliable improvement and 9.1% deterioration, whereas based on the ADOS CSS total score 11.4% exhibited significant improvement and 0% deterioration. The RCI values for the domain scores of the BOSCC indicated that 15.9% of the children showed a reliable improvement in SC and 20.5% a reliable improvement in RRB. Calculations based on the ADOS CSS domain scores indicated that 22.7% showed a reliable improvement in SA and 6.8% a reliable improvement in RRB.

Correlational analyses between the RCIs of the BOSCC and the RCIs of the ADOS CSS revealed non-significant correlations for the total, SC/SA, and RRB scores: \( r = 0.03; r = 0.10; r_s = 0.14 \), respectively. In addition, post-hoc chi-square tests revealed that there was no association between improvement (RCI ≤ -1.96) versus no improvement (RCI ≥ -1.96) on both measures \( \chi^2 (1, N=44) = 0.001, p = 0.98, \chi^2 (1, N=44) = 1.92, \)
Correlational analyses between the RCIs and baseline characteristics showed that the RCI of the BOSCC total score was significantly correlated with the BOSCC total score at baseline \((r = -0.66, p > 0.05)\). This implies that children with high BOSCC scores at baseline were more likely to show improvement than were children with low scores at baseline. Likewise, the RCI of the ADOS CSS total score was significantly correlated with the ADOS CSS total score at baseline \((r = -0.43, p > 0.05)\). The RCI values of both the BOSCC total score and the ADOS CSS total score were not significantly correlated with non-verbal IQ \((\text{BOSCC}: r = 0.29; \text{ADOS}: r = -0.11)\) or age \((\text{BOSCC}: r = 0.09; \text{ADOS}: r = 0.24)\) at baseline.

**Discussion**

One of the challenges for researchers and clinicians when evaluating the progress of children with ASD is the choice of outcome measures, because there is no agreed ‘gold standard’ instrument to measure outcomes. The aim of this study was to investigate the usefulness of a new and promising tool, the BOSCC, for detecting changes in core symptoms of autism. Given that the ADOS has often been used for this purpose, it is important to determine whether the BOSCC is a more appropriate instrument for use in intervention studies involving young children. Our findings showed that although the BOSCC may indeed be a promising instrument, it did not evidently outperform the ADOS in terms of reliability, relation to other measures, or sensitivity to change.

The BOSCC had a satisfactory inter- and intra-rater reliability, as also reported by others (Fletcher-Watson et al., 2015; Grzadzinski et al., 2016; Kitzerow et al., 2015; Nordahl-Hansen et al., 2016). Research has shown comparable reliability for the ADOS, both for the ADOS-G (Lord et al., 2000) and for the ADOS-2 (Lord et al., 2012; Zander et al., 2016). However, in contrast to the ADOS, the developers of the BOSCC have proposed that the instrument can be reliably used by less experienced coders without formal training. We partly confirmed this, in that no training in the administration of the BOSCC was needed, but the coders did need intensive training and ongoing supervision in how to score the BOSCC. Given that items of the BOSCC are closely related to the items of the ADOS, coders who have received ADOS training are likely to become reliable coders quicker than those without previous experience with the ADOS, but this remains to be investigated.
Although both the BOSCC and ADOS had high inter-rater reliability, the BOSCC could be economically advantageous, because training is not required in how to administer the instrument and specific materials are not needed.

The changes in scores on both the BOSCC and the ADOS were generally independent of the change in age, non-verbal IQ, language, and other behavioral problems. Both the BOSCC and the ADOS were demonstrated to measure changes in core ASD symptoms instead of a more general level of impairment.

The correlations between BOSCC and ADOS scores at both time points suggest that both instruments measure fairly similar constructs. However, changes in BOSCC and ADOS scores were not correlated, suggesting that the instruments measure different aspects of change in autistic behaviors. This may be because the scoring range of the ADOS is narrow, which hampers its ability to measure subtle changes over time. In addition, dissimilarities in interactional context between the BOSCC (free play parent–child interaction) and the ADOS (standardized professional–child interaction) might have influenced findings. For example, given that parents and children influence each other’s behavior, changes in parent behaviors might have affected changes in child social skills (Haven et al., 2014). This study did not explore the effect of parental behaviors on child social communicative behaviors. The bi-directional nature of interactions may argue for applying the BOSCC to a semi-structured interaction, in which the adult is provided with specific instructions that are consistent across observations. The subject of contextual influence deserves additional attention in future research.

Although the BOSCC and ADOS had a comparable sensitivity to measure change at a group level, the BOSCC exceeded the ADOS in detecting subtle changes at an individual level. While evidence-based treatments for children with ASD are established on the basis of group differences, it is important to also consider individual variability in this heterogeneous group. Individual trajectories of growth can then be identified, which would help to predict which individual would benefit the most from a particular intervention, which in turn would facilitate individual treatment planning (Kasari, 2002; Zwaigenbaum, Bauman, Choueiri, et al., 2015). This study showed that the BOSCC seems to be a more effective instrument to measure these subtle, individual changes (both in terms of improvement and deterioration), possibly because it uses a broader range of scores. However, it should be noted that the individual change was highly dependent on test-retest reliability, which is used to calculate the RCI. Given that the BOSCC is a relatively new instrument, little is known about its reliability. More research in this area is needed to examine whether the BOSCC measures ‘true’ change in ASD symptoms.

A detailed examination of the reliable individual change revealed that children with more severe autistic behaviors at baseline were more likely to show improvement.
than children with less marked impairments at baseline. This may indicate that there was little room for improvement in the children that performed well at baseline. However, no children performed on ceiling levels on the BOSCC at baseline, as indicated by baseline scores higher than 1 standard deviation (SD) of the lowest (best) score. According to the ADOS, 2 children (4.5%) had baseline scores within 1 SD of the lowest score. Change in autistic behaviors did not depend on the child’s developmental level or age at baseline. In general, these findings suggest that the children with the best chance of showing improvement with time are those children with more severe autistic symptoms at baseline, independent of other developmental impairments.

**Limitations and future directions**

This study should be viewed as a first step toward the use of the BOSCC as an outcome measure, although several study limitations should be borne in mind. We used a preliminary research version of the BOSCC that has already undergone some adaptations by its authors. In addition, the interpretation of results was hampered by the relatively small sample size. For this reason, the definitive version of the BOSCC should be investigated in larger samples. Furthermore, the calibrated severity scores for the ADOS domains (not total score) were investigated in a single study (Hus et al., 2014). Although this study sheds light on the usefulness of the ADOS severity metric as outcome measure, the ‘domain’ findings should be interpreted with caution. Another limitation was the difference in the setting in which behavior was scored with the BOSCC at baseline (clinic) and endpoint (home). The expression of core autistic behaviors may be more intense in an unfamiliar setting than in the child’s home. However, given that there were no significant differences in correlations between the BOSCC and the ADOS (assessed in the clinic at baseline and endpoint) across time points, it is not likely that the change in setting influenced the presence of core autistic behaviors. In order to distinguish between the effect of an intervention and the effect of the setting, it is recommended that intervention studies use the same setting for all assessments. Finally, we were not able to investigate the sensitivity of the BOSCC to measure the response to treatment because the original study did not detect a treatment response (Oosterling, Visser, et al., 2010). Future research should investigate whether the BOSCC is able to identify the level of responsiveness to different evidence-based treatments in relation to other commonly used outcome measures. Moreover, the clinical relevance of the changes detected with the BOSCC should be investigated, because statistical significance or effect size does not provide information about whether a relationship between variables is clinically meaningful (Kraemer et al., 2003). Jacobson and Truax (1991) recommended that, in addition to reliable change indexes (as applied in this study), a clinical cut-off point for clinically
meaningful change should be used. More research is needed to determine how large the change in BOSCC scores (e.g. 1 or 2 SDs decrease) should be to be clinically meaningful.

Conclusion
In conclusion, the findings of this study provide encouraging evidence that the BOSCC is a promising instrument to measure observation-based outcomes. In comparison to the ADOS, the BOSCC seems an adequate instrument to guide personalized care and its use entails less cost. However, the current version of the BOSCC has no greater potential with regard to all quality criteria of an outcome measure. Investigating the final version of the BOSCC, possibly including different modules, may reveal more advantages of the BOSCC.

Although we examined the usefulness of the BOSCC as compared with other outcome measures, it was not our aim to recommend the BOSCC as a sole instrument. The evaluation of a child’s progress requires the use of multiple instruments focusing on both global (e.g. autistic symptoms) and specific (e.g. behaviors directly related to the intervention’s target) characteristics of the child (Lord et al., 2005; Nordahl-Hansen et al., 2016). This combination provides a better understanding of the underlying mechanisms by which an intervention has an effect. Nonetheless, in order to be able to compare the effectiveness of different interventions, it would help if a key outcome, and hence instrument, could be used across intervention studies. Future studies should explore further whether the BOSCC can serve this purpose.
## Supplemental Material

**Table S5.1.** Intra-class correlation coefficients (ICC) for BOSCC items, domain scores and total score (n=20 segments).

<table>
<thead>
<tr>
<th>Items</th>
<th>Inter-rater reliability ICC</th>
<th>Intra-rater reliability ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coder 1</td>
<td>Coder 2</td>
</tr>
<tr>
<td>1. Eye contact</td>
<td>.96</td>
<td>.82</td>
</tr>
<tr>
<td>2. Facial expressions</td>
<td>.82</td>
<td>.95</td>
</tr>
<tr>
<td>3. Gestures</td>
<td>.96</td>
<td>.80</td>
</tr>
<tr>
<td>4. Vocalizations</td>
<td>.96</td>
<td>.96</td>
</tr>
<tr>
<td>5. Integration vocal/non-vocal</td>
<td>.98</td>
<td>.92</td>
</tr>
<tr>
<td>6. Social overtures</td>
<td>.87</td>
<td>.84</td>
</tr>
<tr>
<td>7. Responses</td>
<td>.89</td>
<td>.68</td>
</tr>
<tr>
<td>8. Enjoyment</td>
<td>.97</td>
<td>.95</td>
</tr>
<tr>
<td>9. Engagement with materials</td>
<td>.98</td>
<td>.73</td>
</tr>
<tr>
<td>10. Play with objects</td>
<td>.94</td>
<td>.96</td>
</tr>
<tr>
<td>11. Sensory interests</td>
<td>.89</td>
<td>.75</td>
</tr>
<tr>
<td>12. Hand/finger/body mannerisms</td>
<td>.96</td>
<td>.99</td>
</tr>
<tr>
<td>13. Repetitive interests or stereotype behaviors</td>
<td>.99</td>
<td>.72</td>
</tr>
<tr>
<td>14. Activity level</td>
<td>.98</td>
<td>.72</td>
</tr>
<tr>
<td>15. Disruptive behavior</td>
<td>.95</td>
<td>.89</td>
</tr>
<tr>
<td>16. Anxious behavior</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>BOSCC SC</td>
<td>.99</td>
<td>.96</td>
</tr>
<tr>
<td>BOSCC RRB</td>
<td>.97</td>
<td>.86</td>
</tr>
<tr>
<td>BOSCC Total</td>
<td>.99</td>
<td>.98</td>
</tr>
</tbody>
</table>

*Not applicable, no variance was found*

Notes. BOSCC = Brief Observation of Social Communication Change; SC = Social Communication (item 1-8); RRB = Restricted and Repetitive Behaviors (item 10-13)
Summary and general discussion
Over the last decades there have been rapid advances in ASD research, informing us about early risk markers, the role of both genetic and environmental etiological factors, and the effectiveness of early detection and intervention programs. The overall objective of this thesis was to further enhance our understanding of ASD in infancy and toddlerhood. The first aim was to give additional insight into the earliest signs of ASD by investigating potential early risk markers in infants and toddlers at familial high and low risk of ASD. Specifically, infant temperament and characteristics of parent-infant interaction were studied. Second, this thesis examined clinical referrals from community samples, investigating the long-term effects of a program for early detection of ASD, and evaluating the usefulness of an instrument in measuring change during early intervention of ASD. Taken together, this research was undertaken to contribute to earlier diagnosis and, consequently, intervention. This general discussion provides an overview of the studies described in the thesis, summarizes and discusses key findings, points out limitations, suggests recommendations for improvement of clinical practice as well as directions for future research, and closes with a general conclusion.

Part I - Early risk markers of autism spectrum disorder in infancy and toddlerhood

Summary

Prospective studies of infants at familial risk can provide information about early markers that may reliably differentiate siblings who eventually develop ASD from those who do not. In Chapters 2 and 3, potential risk markers of ASD were investigated in two relatively large samples of young children at risk of ASD (HR) and their low-risk control peers (LR). In the first study, reported in Chapter 2, we examined parent-reported temperament as a potential risk marker. We performed group-based ANCOVAs and individual-based machine learning analyses to investigate whether temperament traits (i.e. surgency, negative affect, effortful control) can help to predict ASD on a group and an individual level. Group-based analyses showed linear risk gradients, with more atypical temperament for HR-ASD, followed by HR-Atypical, HR-Typical, and LR siblings. Early temperament traits were significantly different between outcome groups, with the most striking vulnerabilities in negative affect and effortful control (i.e. regulation of attention, emotion, and behavior). However, the effect sizes were generally medium ($0.03 \leq \eta_p^2 \leq 0.34$), especially regarding differences within the HR group. Machine learning analyses showed that, at an individual level, HR-ASD siblings could not be identified accurately, whereas HR infants without ASD could, as based on the temperament trait effortful control (and its combinations with other traits). Overall, these results emphasize the discrepancy between group-based and individual-based predictions and suggest that while temperament seems not to facilitate the early identification of ASD at an individual level, it does help identify HR infants who do not develop ASD.
In the study described in Chapter 3, we focused on characteristics of early parent-child interaction. A newly developed global coding scheme was applied that includes constructs that were found to predict subsequent child development. Our main finding was that 10-month-old HR siblings had fewer initiations toward their parents as compared to LR siblings ($p<0.05$, $d=0.43$). Group differences remained after correcting for the influence of infant’s sex and age, but were not significant after controlling for parental educational level ($p>0.05$, $d=0.34$). No group differences were found in infant behaviors at 5 months or in parent behaviors at 5 or 10 months of age (all $p>0.05$, $d<0.19$). We preliminary concluded that the observation of early parent-child interaction may help to inform about early emerging atypicalities within the broader autism phenotype (BAP) or later ASD.

**Discussion**

The findings in Chapter 2 question the usefulness of temperament as an early risk marker of ASD. Although our group-based analyses support previous research showing that temperament can differentiate between outcome groups, our analyses at an individual level provided a different picture. The positive predictive value and specificity were not clinically useful, indicating that based on temperament a substantial number of HR siblings would be falsely classified as having ASD at 36 months (i.e. false positives). Given that temperament did not accurately predict which HR infants move on to develop ASD, it may not be useful as an early risk marker of ASD. Our results highlight the difficulties of translating findings from a group to an individual level. In fact, there is often substantial overlap between groups in individual variation, making it more difficult to make predictions for individual infants, and this signals that the temperamental profile within ASD is heterogeneous. However, the high negative predictive values imply that temperament, specifically effortful control and its combinations with other traits, can accurately predict which HR infants are not going to develop ASD in all likelihood. As an alternative to a risk marker of ASD, temperament may be seen as a *stratification marker* that allows to classify individuals with ASD into biologically more homogeneous subtypes (Loth et al., 2017). Temperament can then be conceived as an individual difference amongst the autistic population, rather than as a core feature of ASD itself (Johnson, 2012). These individual differences may modify the course or severity of ASD symptoms or the response to treatment, as is suggested by pathoplasticity or modifier models of temperament (Mundy et al., 2007; Tackett, 2006). For example, those individuals with ASD and low levels of effortful control may be the most likely to exhibit a severe ASD profile. In this way temperament may help to unravel the heterogeneous character of ASD, and may be used to individualize interventions. In addition, although the information about a child’s temperament may lead to false positives,
it may still be useful for prevention in a familial HR group. Based on our findings, the level of effortful control in combination with a child’s risk status can be used to reliably determine who do not develop ASD. For the group of HR siblings with low levels of effortful control it is more difficult to predict their outcomes. Given that familial HR families are often not without concerns regarding the development of the younger sibling of their child with ASD, being aware of the genetic risk factor, the strategy ‘no harm, no foul’ may be followed. By applying early monitoring and short pre-emptive interventions, the developmental trajectory of some of these ‘double at risk’ siblings may change their atypical trajectory into a typical one.

When considering the findings in Chapter 3, it is important to bear in mind that this study was only based on differences between risk groups (i.e. HR vs. LR), because information about the child’s outcome was not (yet) available. Additionally, the group difference in infant initiations was non-significant after correcting for parental educational level, suggesting that the group difference was attributable to different levels of parental education. Therefore, parental effects on infant behaviors that may not have been captured by the applied coding scheme (e.g. number of parental social initiations) and may be influenced by parental educational level should be considered. Also, due to the research setting in which the parent-infant dyads were observed, parent behaviors may have been socially acceptable, and therefore daily life parental behaviors that also affect the observed infant behaviors may have been missed. Although firm conclusions cannot be drawn based on these preliminary findings, the observation of parent-infant dyads may provide valuable information about early markers for ASD already in the first 12 months of life. Differences in infant behaviors may reflect early emerging atypicalities within the BAP or later ASD instead of reflecting the impact of parental behaviors. This supports the view that there is an ASD prodrome in the first year(s) of life that is characterized by delays and impairments in early social interaction and communication (Yirmiya & Charman, 2010). These ‘prodromal’ characteristics may perturb early parent-child interactions, which in turn may increase the risk of the infant developing the full ASD phenotype, as is pictured in Figure 6.1 (Mandy & Lai, 2016). The key question remains whether the atypicalities in infant initiating behaviors during interaction with their parents (1) should only be seen as potential early markers for ASD, or (2) may indeed further alter the interaction between parent and infant (pathway b), and whether a perturbed parent-infant interaction actually has an impact on subsequent child development and ASD diagnosis (pathways c and d).

Another important aspect to consider is the generalizability of the findings reported in Chapter 2 and 3, as they are both embedded in a HR sibling design. It is unclear whether and to which extent findings from familial HR studies generalize to infants identified as HR from community samples, who may not have an older sibling
diagnosed with ASD (Johnson, Gliga, Jones, & Charman, 2015; Szatmari et al., 2016). For example, parents who participate in HR studies may do so because of concerns about their infant’s development, and they may over-report symptoms in the infant sibling because they are primed and sensitized to the presentation of ASD through the older sibling with ASD. Additionally, there may be different genetic pathways to ASD in the HR sample (on the basis of familial risk) as compared to rare cases (i.e. single gene or de novo mutations), complicating generalizability. Furthermore, the implications of early screening are different in a familial HR sample (i.e. parents who are informed and may already have concerns) as compared to a HR community sample (i.e. parents who may be naive and unconcerned). We will continue this discussion later in this chapter.

Figure 6.1. A proposed gene-environment interplay in the emergence of autism spectrum disorder (based on Kong et al., 2018; Mandy & Lai, 2016).

**Future directions**

*Unraveling the role of early temperament*

Our findings prompt further research into early risk markers of ASD. The innovative analysis approach that was applied in Chapter 2 provides important information about temperament as an individual predictor for subsequent child development. Future work should investigate whether the integration of clinical (e.g. MSEL, VABS) and biological (e.g. eye tracking, function imaging) measures can improve the positive predictive value for the clinical diagnosis of ASD at an individual level (Bussu et al., 2018), and to investigate the additional value of temperament. Also, it should be examined to which extent an atypical temperament reflects brain alterations that
predispose to ASD and/or are shared between atypical development and ASD. For example, mapping the longitudinal relations between early brainstem-related physiological regulation and frontal cortex development (which both have been linked to later ASD and self-regulation) and effortful control will provide insight into the roots of effortful control atypicalities in young children with ASD (Johnson et al., 2015).

Exercising mechanisms of parent-infant interaction

Another important future goal is to investigate longitudinal trajectories of parent-child interaction across early childhood. This allows the investigation of causal relationships, which is important given the bidirectional nature of interaction, and provides information about the child's outcome status at 36 months. Future research should specifically focus on whether characteristics of parent-infant interaction can have a protective influence on the risk of an individual developing psychopathology, as is reflected in pathway d in Figure 6.1 (Mandy & Lai, 2016). This can be examined by conducting randomized controlled trials (RCTs) of interventions, as was for example done by Green et al. (2017). Findings of this pre-emptive intervention study are consistent with the idea that parent-infant interaction may influence the ASD phenotype in children who already have an elevated familial risk of developing ASD. This is in line with findings from ADHD research, showing that higher levels of motor activity at 4 months signal risk for later ADHD, but in the presence of higher quality parenting the effect of this risk on the ADHD phenotype is attenuated for boys, and even reversed for girls (Miller, Degnan, Hane, Fox, & Chronis-Tuscano, 2018). In other words, characteristics of parent-infant interaction may be viewed as a modifier of subsequent child development, and thus are a very relevant target for early intervention in children at early risk. This strategy will be followed in a new research project initiated by our team, aiming to improve direct care in early childhood by offering a short home-based early intervention to parents of children (12 to 30 months) who screened positive (and are therefore at risk of ASD) during regular well-baby clinic visits. This will also allow us to examine whether and to which extent findings from familial HR studies generalize to infants identified as HR from community samples.

In addition, other factors that may influence parent-infant interaction characteristics should be further investigated, such as parental BAP, experienced stress, parental emotion regulation, infant temperament, or pre- and perinatal factors. For example, recently the role of parental emotion regulation have been emphasized in ADHD research (Gershy & Gray, 2018). Parents of children with ADHD who had more difficulties regulating their emotions showed more coercive parenting. In the light of the differential susceptibility theory, children at risk for
psychopathology may be differentially susceptible to their early social environment, such that both negative and positive interactions with their parents have a high impact. In other words, higher levels of coercive behaviors during early interaction will likely have a negative impact on the child’s development, and providing an optimal primary environment is especially important for children at risk. Moreover, parental genotypes may not only directly affect the genotypic architecture of the child (e.g., familial risk for ASD), but non-transmitted alleles can also impact the parent-child interaction via the phenotype of the parent, influencing the child’s development (i.e., genetic nurture) (Kong et al., 2018).

Overall, one of the challenges in accomplishing these future goals will be the selection of a coding measure for parent-child interaction. The current thesis applied a newly developed global measure, which yielded promising results, but more research is required into its psychometric properties, its use at older ages, and its potential utility in clinical practice. The use of a macro-level coding approach (applying a global rating of qualitative and quantitative behaviors) in combination with a micro-level coding approach (focusing on frequency, durations and timing of behaviors) should also be considered in future work (Mesman, 2010; Yoder & Symons, 2010). Macro-level coding can help to inform about the quality of the interaction, integrates contextual information, and enables to code a range of behaviors in a relatively time-efficient manner (as was done in this thesis). Micro-level coding is more time-consuming and often requires specific coding software, but this coding tend to be more objective and can help to address questions of quantity. By using these approaches simultaneously, a coherent integrative view of the parent-child interaction can be collected and macro-level and micro-level coding can be compared. In addition, the use of novel methods to examine parent-child interaction can be considered in future research, such as head cams, eye tracking, or EEG measures. For example, the use of delicate head cams for both parent and child allows to record parent-child interactions in the natural home setting instead of a lab setting in which ceiling cameras are used, and may also be valuable for evaluating parental and child behaviors during interventions. Also, the use of dual eye tracking methods facilitates the interpretation of the use of eye gaze, and allows to analyze gaze behavior in ongoing social interactions (Pfeiffer, Vogeley, & Schilbach, 2013).

**Studying early deviant development**

In general, in order to help the field of early detection forward, three broad aspects should be considered in future studies. First, investigation of very early brain development may provide information about pre-behavioral markers of risk and offers an opportunity to investigate causal pathways to ASD (Varcin & Jeste, 2017). Techniques like EEG, fMRI, fNIRS, and eye tracking have started to demonstrate
that atypical brain structure and function can be identified in the first year of life (Emerson et al., 2017; Hazlett et al., 2017; Varcin & Jeste, 2017), and the hope is that future studies may eventually translate these findings into clinical practice. This may open doors to the use of mobile devices in the future, using digital phenotyping of ASD in addition to clinical rating in practice. However, more research into the early markers of ASD is needed before such markers can be considered to be used as clinically relevant measures in the early detection of ASD. Future studies should include large sample sizes that enable sophisticated analyses that document changes with age, and investigate correlations between brain and behavior.

Second, to help improve the clinical identification of ASD, future studies should include both group-level and individual-level analyses. Thus far, the majority of studies have focused only on group-level differences, but this thesis has indicated that there may be discrepancies between group-based and individual-based findings. There is often substantial overlap between groups in individual variation, making it more difficult to make predictions for individual infants. By investigating whether early characteristics also represent markers at the individual level, we can learn more about the specificity and sensitivity of these markers to later clinical diagnosis (Johnson et al., 2015).

The current thesis did not include other outcome groups (e.g. ADHD, anxiety), so it does not allow to draw any conclusions about unique and shared underlying causal mechanisms. However, the necessity to make definite conclusions regarding a categorical diagnosis very early in child development is questionable. This applies, for example, to the group of young HR children following an atypical development without meeting the criteria for a clinical diagnosis at 36 months of age (in Chapter 2: ‘HR-Atypical siblings’). Some children in this group may develop clinical psychopathology at a later stage, and early parent-mediated pre-emptive intervention may be helpful to change their deviant developmental trajectory. In addition, the coexistence of disorders and the overlap between them (e.g. ASD and ADHD) argues for multidisciplinary evaluation without the need of a discrete clinical label (Gillberg, 2010). Parent-mediated (pre-emptive) interventions may always be valuable in case of a (sub)deviant development.

Part II - Early detection and intervention of autism spectrum disorder

Summary
The severity of ASD, its rising prevalence (and its economic and social costs), and the growing evidence for early intervention, all underline the importance of early detection of ASD. In Chapter 4, we presented one of the few studies available in the literature that reports on the sustainability of the effect of an early detection
program for ASD. Our team developed and evaluated an integrated early detection program that comprised of a) a training of primary care providers to recognize early signs, b) the use of a specially designed referral protocol that included a screening questionnaire, and c) the formation of a multidisciplinary diagnostic team (Oosterling, Wensing, et al., 2010). Results revealed that the odds of being referred before the age of 3 years was higher in children with ASD than in children with another condition during the program than before (3.1, 95% CI: 1.2–7.6) or after (1.7, 95% CI: 1.0–3.0) the program, but was not different before versus after the program. In other words, the program proved to be effective over the years of active investment, but the effect disappeared when active investment had faded out. Furthermore, children with intellectual disabilities were more likely to be referred before 3 years of age than were children with a higher level of cognitive functioning at all time points. The odds of being referred before 3 years was 2.6 times (95% CI 1.6–4.2) and 10.0 times (95% CI 5.9–16.9) higher if a child with ASD had an IQ <70 as compared to an IQ between 70 and 89 and an IQ >90, respectively. The odds ratio for being referred was 3.9 times (95% CI 2.3–6.5) higher for children with ASD with an IQ between 70 and 89 than for children with ASD with an IQ >90. Our study highlights the importance of maintaining early detection through continuous investment in active screening and ongoing training of primary care providers. Additionally, given that children with higher levels of cognitive functioning may be easily missed during early screening, there is a need for investigating strategies how to detect this group as early as possible.

As there has been an increasing focus on the early detection of ASD, as highlighted in part I, the availability of evidence-based early intervention of ASD is also highly important. To draw conclusions about the effectiveness of the intervention type and intensity for children with ASD, an enormous variety of outcome measures have been used in intervention studies to monitor change over time. However, there is a need for suitable valid and reliable outcome measures that capture meaningful change and minimize risk of bias, for example by using blinding outcome assessments (French & Kennedy, 2017). To compare the effectiveness of different interventions, it would help if one outcome measure is systematically applied across intervention studies. The first steps toward a key outcome measure may have been taken by developing the Brief Observation of Social Communication Change (BOSCC) (Grzadzinski et al., 2016). In Chapter 5, we compared the ability of the BOSCC with a commonly used outcome measure, the Autism Diagnostic Observation Schedule (ADOS | Lord et al., 1999; Lord et al., 2012), to detect changes in core symptoms of ASD in toddlers. Whereas the ADOS was specifically designed to assess whether an individual meets a diagnostic threshold and helps to identify that individual’s current clinical status, the BOSCC was designed to measure subtle changes in autistic behaviors over time.
We concluded that although the BOSCC did not evidently outperform the ADOS on its inter- or intrarater reliability, independency from other child characteristics, and sensitivity to capture change, the BOSCC may be better able to capture subtle individual changes.

**Discussion**

Although our findings described in Chapter 4 provide evidence of the short-term effectiveness of an early detection program, the question remains whether children who are early screened do have better outcomes than children who are referred later. No information was available about the post-diagnostic process, so we do not know whether the children who were screened before the age of 3 years actually had earlier access to treatment and had better long-term outcomes than children who were screened later. In this light there is expert debate about whether screening for ASD in young children is beneficial. While the American Academy of Pediatrics recommends systematic screening at different intervals during infancy and early childhood for all children Committee on Practice and Ambulatory Medicine (2016), the US Preventative Services Taskforce (USPSTF) has published recommendations stating that 'current evidence is insufficient to assess the balance of benefits and harms of screening for ASD in young children for whom no concerns of ASD have been raised by their parents or a clinician' (Siu et al., 2016). According to the USPSTF there is a lack of research into the efficacy of treatment and association between early screening and long-term effects on child health outcome. There may be harms in terms of misdiagnosis and the anxiety related to further testing after a positive screening result. Also, interventions can be time-absorbing and financially consuming for families. In contrast, proponents of early systematic screening in the community have emphasized that there is evidence for the ability of universal screening tools to detect ASD in toddlers at an earlier stage, and for the effectiveness of early interventions on child and family outcomes (Dawson, 2016; Veenstra-VanderWeele & McGuire, 2016). Also, early screening may be less potential harmful in families who already have concerns about their child’s development, such as families with an older child that is diagnosed with ASD. Screening and monitoring of these families may therefore be different than screening of families who are not aware of risk factors and do not (yet) have any concerns about their child. The early detection program as was applied in our study was based on a two-stage screening approach in which a complete screening instrument is only applied to those children who are found to have a deviant developmental path during routine development surveillance (Filipek et al., 1999). Although this approach only focuses on young children for whom concerns exist without immediately adding the burden of a complete diagnostic assessment, we still do not know whether screening of these ‘at risk’ children has positive long-term effects. Another issue that should be discussed is that we do not have any information about missed cases during
the three phases of our study (i.e. before, during, and after the early detection program). Nonetheless, our study showed that children with intellectual disabilities were more likely to be referred before the age of 3 as compared to higher functioning children, suggesting that the latter group may be missed in early detection programs. As was also suggested in Chapter 4, these children may show more subtle behavioral symptoms as compared to the lower functioning children. Possibly, higher functioning children may possess equal ASD impairments as compared to children with intellectual disabilities, but they may be better able to compensate for their difficulties due to their greater cognitive capacity (Livingston & Happe, 2017).

Our findings in Chapter 5 suggest the additional value of the BOSCC as compared to the ADOS in its ability to capture subtle individual changes, but we do not recommend the BOSCC as a sole instrument to measure change in intervention studies. The evaluation of a child’s progress requires the use of multiple instruments focusing on both global and specific characteristics of the child (Lord et al., 2005; Nordahl-Hansen et al., 2016). While specific instruments focus on behaviors directly related to the intervention’s target (e.g. parent interaction style), global instruments can measure change in behaviors related to broader ASD symptoms. In addition, a differentiation can be made based on the context in which change in behaviors is measured (Yoder, Bottema-Beutel, Woynaroski, Chandrasekhar, & Sandbank, 2013). Context-bound measures are based on a context that is highly similar to the intervention setting whereas the context of generalized measures differs from the intervention setting. In many intervention studies outcome measures are highly proximal to the intervention target (i.e. specific) and determined within or in very similar contexts to the treatment setting (i.e. context-bounded). How does this translate into more distal behaviors in generalized contexts? To show ‘true’ effectiveness of a parent-mediated intervention it is essential to prove effectiveness based on behaviors during parent-infant interactions directly related to the intervention’s target, and generalization into wider social contexts (e.g. interaction with an unfamiliar adult instead of parent-child interaction in parent-mediated interventions) and global ASD symptoms (Green & Garg, 2018). In this view, the BOSCC may be applied as an outcome measure on both parent-infant interactions and interactions with an unfamiliar adult.

Future directions
Early screening, better outcomes?
For an ultimate justification of early detection programs, two important issues remain to be further investigated. The first question is whether early screening actually leads to better outcomes on the long term. Currently it is not clear whether early screening does in fact allow earlier access to treatment, and the long-term effects
on child development and quality of life remain to be unknown. Future studies should longitudinally follow up children from the general population who are randomly assigned to undergo (or not) a screening program, regardless of whether they screen positive or negative with regard to the likelihood of ASD (or another developmental disorder). The children who are not screened can be offered the same treatments on request of potential professionals involved. In addition, a cost-effectiveness analysis should be conducted, including potential cost savings, to examine whether the cost of screening is worth the benefit.

**Investigating the effectiveness of (pre-emptive) interventions**

Inevitably, the second step is to grant access to support services following early concerns (or diagnosis), and to answer the urgent need for high-quality evidence-based early interventions. Parental experiences about the post-diagnostic process are often negative, including a dissatisfaction with post-diagnostic support (if any) and a feeling of ‘being left in the dark’ (Crane, Chester, Goddard, Henry, & Hill, 2016). Although progress has been made in intervention research, most studies lack in the overall quality of the trial design (French & Kennedy, 2017; Green & Garg, 2018). Problematically, interventions that are applied in common health care practice are based on these studies, resulting in applied interventions with little or no rigorous evidence base. We think now the time is right for more ambitious trial designs (including large samples, allocation concealment, and blinding of participants and outcome measures) that investigate long-term change in children and their families.

Some studies have already started to show positive effects on the long term (Estes, Munson, et al., 2015; Green et al., 2017; Pickles et al., 2016). Ideally, the outcome measures used in these study designs should incorporate both global vs. specific and context-bound vs. generalized outcome measures. To investigate underlying mechanisms of the intervention, it is important to examine mediating effects between these different outcomes. For example, Pickles et al. (2014) revealed that positive changes in specific behaviors (i.e. child initiations during parent-child interaction) mediated the treatment effect on global symptom change across context (i.e. ADOS scores). These encouraging findings need replication by using a combination of different outcome measures that provide a better understanding of the underlying mechanisms, or ‘active ingredients’, by which an intervention has an effect. In other words, how, why and for whom do interventions work? Individual trajectories of growth can then be identified for the heterogeneous group of children diagnosed with ASD, which would help to predict which individual would benefit the most from a particular intervention, which in turn would facilitate individual treatment planning (for instance by using sequential multiple assignment randomized ‘SMART’ trials |
Kasari, Kaiser, et al., 2014). A final note in selecting outcome measures is to consider the use of ‘real-world’ targets (e.g. quality of life | Jonsson et al., 2017; McConachie, Parr, et al., 2015), and, instead of purely psychosocial measures, neural outcome measures (e.g. Jones et al., 2017). Neural outcome measures can involve the use of EEG or eye tracking to measure social attention, or wearables to detect emotion recognition or regulation during naturalistic social interactions.

Nonetheless, in order to be able to compare the effectiveness of different interventions, it would help if a key outcome, and hence instrument, could be used across intervention studies. Our study in Chapter 5 showed that the BOSCC exceeded the ADOS in detecting subtle changes at an individual level, suggesting that the BOSCC may be able to guide personalized care as a global measure. Future studies should further investigate the usefulness of the BOSCC, focusing on its use in different settings (i.e. parent-child vs. examiner-child) and on its response to treatment. Importantly, to investigate which intervention works for whom, both group-based and individual-based analyses should be performed.

**General clinical implications**

*The role of temperament in early detection and intervention*

The primary motivation for identifying markers of ASD during the first year(s) of life is the hope to change the course of an early emerging atypical development by applying early interventions and even emptive interventions with a preventative character. The study reported in Chapter 2 showed that temperament did not accurately predict which HR infants move on to develop ASD, but can accurately predict which HR infants do not go on to develop ASD. Temperament may therefore still be an informative marker to provide early monitoring (and possibly pre-emptive intervention) for those children that are familial at risk and show low levels of effortful control. Additionally, child’s temperament may have a potential clinical role in interventions. On the one hand, if a particular temperament profile may lead to more positive outcomes, pre-emptive interventions can be developed that are designed to target this (Mundy et al., 2007). For example, by integrating emotion regulation strategies in parent-mediated interventions. Importantly, child temperamental challenges likely have an impact on the child’s environment. For example, children showing high levels of negative affect and low levels of emotion regulation (resulting in daily tantrums) may influence the interaction with their parents, which in turn may lead to more inconsistent parenting and parental stress. This may have a cascade of effects ultimately effecting child’s development. On the other hand, information about a child’s temperament may be used to determine the type and intensity of intervention, in this way individualizing therapies and providing the most effective conditions for improving child and family outcomes. Based on
the research conducted in this thesis, conclusions about the role of temperament in community settings cannot be drawn.

**The role of parent-infant interaction in early detection and intervention**

Although firm conclusions cannot be drawn based on the findings in this thesis, preliminary findings suggest that parent-infant interactions may help to inform about early emerging atypicalities within BAP or later ASD. The observation of parent-infant interactions may therefore be an important part of diagnostic procedures and the additional value of standardized coding schemes should be further investigated in clinical settings. Although the findings in Chapter 3 require replication within larger cohorts including information about the child’s outcome, they support the use of intervention strategies at an early pre-diagnostic stage (<12 months of age) and help to strengthen parent-infant interaction as a promising avenue for early intervention. The early family environment is essential in supporting a healthy child development and parent-child interactions are likely involved in shaping etiological pathways for atypical development. Given the dynamic and plastic nature of the brain during infancy and toddlerhood, this is a critical period to possibly influence interactions between the brain and its social environment (Webb et al., 2014). One of the advantages of parent-mediated interventions is that parents have the ability to practice skills with their child throughout the day and across settings. Therefore, we argue for the importance of future research into the critical role of parents at an early stage for helping shape their children’s future, specifically in at risk children who might not have a diagnosis yet.

**Implementation of early detection**

Knowledge about risk profiles and the interaction with the environment may help to signal early challenges, and to provide adequate care to children about whom serious concerns exist (not necessarily after a diagnosis is given). Growing scientific evidence indicates that early and appropriate access to interventions improves long-term outcome and may prevent escalation of problems (e.g. negative spirals in parent-child interaction), or development of comorbid behavioral or emotional problems (Estes, Munson, et al., 2015; Green et al., 2010; Pickles et al., 2014; Pickles et al., 2016). Until there is new evidence for the contrary, we assume that early signaling for risk groups (instead of universal screening), has positive effects on the long-term effects of children and families. Yet, how to develop effective strategies that require minimal financial support but have permanent positive impact?

The main explanation for the disappearance of positive effects in Chapter 4 was a lack of financial support, which resulted in the end of an intensive collaboration between primary and specialist care (resulting in a lack of awareness among primary
care providers). Also, the training sessions for professionals in recognizing early signs of ASD were not continued after the end of the program, which meant that experience and knowledge about early detection were lost if staff left. Continuation of the program, with training of personnel, would be expensive. Therefore, specific strategies should be considered to overcome barriers when implementing (and maintaining the effects of) early detection programs. First, collaboration within and between countries, with the sharing of knowledge, is essential, and will help to develop a standardized approach to the detection of ASD and other developmental challenges. (Inter)national guidelines for early detection and diagnosis should be made widely available. Second, the training of healthcare professionals in the early signs of deviant developmental trajectories is one of the key aspects in early detection programs, so the availability and accessibility of knowledge is highly important. To make this affordable at the long-term, e-health technology can play a pivotal role. We recommend e-learning for professionals as a promising tool to teach them about early signs. One important contribution to this, which has been initiated in the Netherlands, is a Live Online Learning (LOL) module. This involves the discussion of specific case studies of the professionals themselves in a virtual online classroom and with the support of an experienced clinician (Van ‘t Hof et al., 2017). In addition, state-of-the-art online platforms should be made available including information about red flag early signs of ASD (and other developmental challenges) and information about diagnoses and relevant services. Fortunately, over the past years a number of tools have been developed that inform about the red flags for ASD with the ultimate goal to lower the age at diagnosis (e.g. http://www.autismejongekind.nl; http://www.autismnavigator.com; http://www.cdc.gov/ncbddd/autism/actearly). Given that concerns of parents about their child’s development in the first year of life can help to predict their child’s development (including a potential ASD diagnosis) (Ozonoff et al., 2009; Sacrey et al., 2015), they play an important role in early detection as well. Therefore, these platforms should also be easily available for all families with concerns. In addition, they could be informed via magazines for young parents and promotional material provided in waiting rooms of general practitioners and well-baby clinics. The time is right for the standard use of such methods which can help to raise awareness and maintain the effects of early detection programs without incurring high costs. Future studies should evaluate programs that apply these strategies to assess their effectiveness.

Furthermore, early detection methods should be further improved by examining which children are yet to be missed and, importantly, why they are missed. The study reported in Chapter 4 has provided information about the impact of cognitive functioning on the age at diagnosis, but this thesis did not provide information about other child, family and community factors that may contribute to inequality in the
identification of children with ASD, including socio-economic, cultural, and sex disparities (Daniels & Mandell, 2014; Dworzynski, Ronald, Bolton, & Happe, 2012; Mandell, Novak, & Zubritsky, 2005). The question arises how to avoid false negatives (and positives) during early screening? First, screening can be repeated at different ages as a follow-up, which allows screening programs to act as developmental surveillance tools (Magan-Maganto et al., 2017). In some children it may simply not be possible to detect ASD at an early stage, and attempts to do so might result in a high number of false positives. Repeat evaluation in early childhood when potential problems with peers start to emerge (e.g. around 4 to 6 years) may help to increase the efficacy of the screening process. Second, another strategy is the use of different tools in multiple contexts in the screening process (Magan-Maganto et al., 2017). For example, by using information obtained from both the parent and a day-care centre professional (Dereu et al., 2010). Third, current screening and diagnostic tools may be more adapted to certain groups of individuals and less sensitive for others (e.g. the female ASD phenotype) (Mandy & Lai, 2017). This issue should be further investigated, and if it holds true, the sensitivity of tools should be improved. Finally, brief pre-emptive interventions may be applied to those infants who are at risk (e.g. based on familial risk or screen-positives on a screening instrument), without the presence of a psychiatric diagnosis.

Notably, this thesis specifically focused on ASD, but the questionability of definite conclusions regarding a categorical diagnosis very early in development has been discussed previously in this chapter. Identifying risk markers for a broader defined risk group (not only for ASD specifically) enables early intervention, even before a diagnosis is given. This can optimize outcomes for those infants at risk for ASD or other developmental challenges. This is also in line with the holistic approach as suggested by Gillberg (2010), placing multiple conditions under the acronym ‘Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examinations’ (ESSENCE). The ESSENCE concept focuses on the interrelatedness and co-existence of child neurodevelopmental problems and includes categorical disorders such as ASD and ADHD. In this view preschool children should be multidisciplinary assessed and receive adequate intervention as early as possible. Instead of a wait-and-see stance, a pre-emptive intervention can then be used to enhance social communicative abilities, monitor the child’s development, and evaluate the individual needs of the child and his/her family.
Box 6.1 Key points

- Infants at familial risk who subsequently develop ASD exhibited temperamental atypicalities as compared to their at risk non-ASD peers and low-risk controls, with the most striking vulnerabilities in negative affect and effortful control.
- Individual-based analyses, using machine-learning algorithms, revealed that temperament seems not to be helpful for individual prediction of ASD, but can accurately predict which familial HR infants are not going to develop ASD.
- 10-month old infants at familial risk of ASD showed fewer social initiations during interaction with their parents as compared to low-risk controls, likely reflecting early emerging atypicalities within the BAP or later ASD. However, group differences were not significant after controlling for parental educational level.
- An integrated early detection program for ASD proved to be effective over the years of active investment, but the effect was not sustained on the long-term (after the program had ended), highlighting the importance of maintaining early detection through continuous investment in active screening and ongoing training of primary care providers.
- Children with intellectual disabilities were more likely to be referred before 3 years of age than were children with a higher level of cognitive functioning. Children with higher levels of cognitive functioning may be easily missed during early screening.
- A newly developed outcome measure, the BOSCC, did not evidently outperform the ADOS on its inter- or intrarater reliability, independency from other child characteristics, and sensitivity to capture change.
- The BOSCC may be better able to capture subtle individual changes and seems a promising measure to guide personalized care.
Box 6.2 Clinical implications

- Temperament may be an informative risk marker to provide early monitoring (and possibly pre-emptive intervention) for those children that are familial at risk and that show atypical temperament traits.

- Temperament may also function as a stratification marker, allowing to classify individuals with ASD into biologically more homogeneous subgroups.

- Infant’s temperament may be targeted in very early pre-emptive interventions, and may also facilitate individual treatment planning in children diagnosed with ASD.

- The observation of early parent-child interaction (<12 months) may help to inform about early emerging atypicalities within BAP or later ASD, and offer the opportunity to also evaluate parental behaviors.

- Parent-mediated intervention efforts to optimize child’s social communicative behaviors need to start early in infancy.

- Specific cost-effective strategies should be considered to overcome barriers when implementing and maintaining the effects of early detection programs, such as e-learning tools for primary care professionals.

- There is a need for suitable outcome measures that are sensitive to meaningful change and minimize risk of bias.
Box 6.3 Future directions

- Future work should investigate whether the integration of biological and clinical measures can improve the individual prediction of ASD, and examine the extent to which early temperament could be of additional value.
- Future studies should include both group-level and individual-level analyses to help improve the specificity and sensitivity of markers to later clinical diagnosis.
- More research is needed into very early brain development that may inform about pre-behavioral markers and causal pathways to ASD.
- Future work should investigate the early emergence of ASD symptoms in population-based longitudinal cohorts (instead of solely focusing on familial HR designs), and those that define risk status in other ways (such as prematurity).
- Future work should investigate the causal mechanisms of parent-child interaction on subsequent child functioning across early childhood, including information about subsequent child’s outcome and factors that may also influence parent-child interaction characteristics (e.g. infant temperament, parent emotion regulation).
- More research is required into the psychometric properties of the applied coding scheme (PInTCI), its use at older ages, and its potential clinical utility.
- Future work should consider the use of both micro-level and macro-level coding of parent-child dyads, and may also use novel methods, such as head cams or dual eye tracking.
- Future studies should longitudinally follow-up children from the general population who are randomly assigned to undergo (or not) a screening program, and analyze the cost-effectiveness of the program.
- There is a need for investigating strategies how to detect children with higher levels of cognitive functioning (and other groups that are likely to be missed) as early as possible.
- More research into the usefulness of the BOSCC as a (key) outcome measure is required, also examining its use in different settings (e.g. parent-child vs. examiner-child) and response to treatment.
- Future studies that evaluate the effectiveness of interventions should consider the use of both global and specific outcome measures, and both general and context-bound outcome measures, and investigate potential mediating effects that provide understanding about active ingredients of the intervention.
- Future work should also consider the use of real-world targets as outcome measures (e.g. quality of life), and neural outcome measures (e.g. EEG or eye tracking).
General conclusions
This thesis aimed to improve our knowledge of several aspects of early development of ASD. The different studies provided candidate risk markers (i.e. child temperament, infant initiations during parent-infant interaction) worthy of further investigation as well as some future directions for early detection programs and the use of outcome measures in early interventions for ASD. Nonetheless, the findings of our studies are small pieces of a big puzzle and results should be interpreted with caution, particularly with respect to generalization. For example, whether the findings from the HR siblings studies will generalize to infants with ASD who do not have a family history of ASD remains unclear, as does the extent to which the findings are specific to ASD as opposed to other neurodevelopmental disorders (e.g. ADHD). Therefore, findings from the HR sibling design need to be considered in the context of other designs that also address early emergence of ASD symptoms in both population-based longitudinal cohorts, and those that define risk status in other ways (such as prematurity), also including other neurodevelopmental disorders (Szatmari et al., 2016). Furthermore, investigating developmental trajectories of brain and behaviors (and the interaction between them) across early childhood (including the first 12 months of life), considering both group-based and individual-based predictions, testing dimensional and categorical outcomes, and examining the meaning of findings to early detection and intervention, are exciting possible directions for future research.

Although labels (i.e. DSM-classifications mostly) are needed to refer children to appropriate follow-up services or interventions, the ultimate goal of early detection is to identify early signs, discuss concerns with the family, and guide parents in how to best support their child. We hope that the findings in this thesis will help to shift from an approach of ‘diagnosis and treat’ to ‘predict and promote’. This enables early intervention, even before a diagnosis is given, to optimize outcomes for those infants at risk for ASD or other developmental challenges. Given the high co-existence of disorders and the overlap of symptoms across disorders, a multi-problem and multidisciplinary assessment in the population of young infants (0-6 years) may be more appropriate instead of the current trend in child and adolescent psychiatry toward more compartmentalizing disorders. Given that the majority of false positive cases in early detection programs for ASD tend to be children who have other developmental problems, early detection programs could also assist in detecting other neurodevelopmental problems.

Our understanding of the early emerging ASD phenotype has undergone a revolution in the past 30 years. A future challenge is to improve our understanding of the underlying biological and environmental influences that lead to ASD symptoms, and to identify prodromal signs that will help mark out infants at risk before the onset of the full-blown disorder (Yirmiya & Charman, 2010). An increased focus on the translation of findings into clinical practice (both early detection and intervention) may
be fruitful to ensure that all children with early developmental challenges and their families benefit from the full extent of research efforts.
APPENDICES

Nederlandse samenvatting
(Dutch summary)
References
Dankwoord
(Acknowledgements in Dutch)
Publication list
About the author
NEDERLANDSE SAMENVATTING

Het onderzoek in dit proefschrift
Dit proefschrift heeft ten doel de kennis te vergroten van autismespectrumstoornis (ASS) in de baby- en peutertijd, en bij te dragen aan het verbeteren van vroegtijdige screening, diagnostiek, en vandaar uit vroegtijdige interventie. **Ten eerste** wordt meer inzicht verkregen in de vroege signalen van ASS door potentiële risicofactoren te onderzoeken gedurende de vroege ontwikkeling. Dit is mogelijk door gebruik te maken van een groep kinderen uit een Europees onderzoeksproject met een verhoogd risico op het ontwikkelen van ASS, namelijk de broertjes en zusjes van kinderen met de diagnose ASS. Deze groep ‘hoog-risico’ kinderen wordt vergeleken met een groep kinderen met een broer of zus met een gezonde ontwikkeling, de zogenaamde ‘laag-risico’ kinderen. Door de vroege ontwikkeling (0 tot 36 maanden) van beide groepen te volgen kan er meer inzicht worden verkregen in de vroege kenmerken en onderliggende mechanismen in de ontwikkeling van ASS. Meer specifiek worden in dit proefschrift de vroege kenmerken van temperament en de ouder-kind interactie onderzocht. **Ten tweede** worden kinderen uit een Nederlandse klinische populatie onderzocht, dat wil zeggen kinderen die werden verwezen naar de kinder- en jeugdpsychiatrie en bij wie al dan niet een klinische diagnose ASS werd vastgesteld. Op basis van deze populatie beschrijft dit proefschrift 1) een wetenschappelijke studie waarin de langetermijneffecten van een vroegdetectie programma worden bepaald, en 2) een studie waarin de bruikbaarheid van een nieuwe uitkomstmaat voor interventiestudies wordt onderzocht.

Wat is een autismespectrumstoornis?
ASS is een ontwikkelingsstoornis die al in de vroege ontwikkeling aanwezig is. Binnen de (kinder- en jeugd)psychiatrie wordt gebruik gemaakt van een classificatiesysteem waarin de precieze gedragskenmerken van onder andere ASS staan omschreven: de DSM-5 (American Psychiatric Association, 2013). Aan de hand van dit systeem kan worden beoordeeld of er sprake is van een klinisch beeld dat samen te vatten is in een bepaalde classificatie. Daarbij zijn van belang: de hoeveelheid symptomen, de mate van ernst, de fase waarin de symptomen voor het eerst voorkwamen, en de belemmeringen in het functioneren. Op basis van deze criteria is bij ongeveer 1% van de bevolking (116 op de 10.000 personen) sprake van ASS (Baird et al., 2006). Individuen met ASS hebben beperkingen op het gebied van de sociale communicatie en interactie, en vaak vallen stereotiepe of repetitieve patronen in het gedrag op. Daarnaast kan er sprake zijn van sensorische gevoeligheden, zoals voor harde geluiden of bepaalde texturen van kleding. Ondanks dat alle individuen met ASS over het algemeen beperkingen laten zien in deze domeinen, kunnen bijvoorbeeld de ernst van
de symptomen en het verloop over tijd zeer verschillen tussen personen. De diagnose ASS wordt meestal gezien als een levenslange diagnose, maar op basis van onderzoek is bekend dat de inzet van vroege behandeling een positief effect heeft op de latere ontwikkeling (hoe vroeger, hoe beter). In de afbeelding hieronder staat een aantal cijfers over ASS (Dietz & Oosterling, 2017).

Zoals in de DSM-5 is beschreven, moeten symptomen van ASS reeds zichtbaar zijn in de vroegkinderlijke ontwikkeling. Kennis over de uiting van ASS gedurende de eerste levensjaren is hiermee van groot belang voor het betrouwbaar diagnosticeren van ASS. In het eerste levensjaar kunnen de problemen zich uiten in een beperkte sociale interesse (bijv. minder gericht zijn op de ouders, of niet reageren op aanspreken) (Campbell et al., 2015; Feldman et al., 2012a; Miller et al., 2017). In het tweede en derde levensjaar worden de beperkingen vaak duidelijker zichtbaar, bijvoorbeeld door een vertraagde spraaktaal-ontwikkeling, geen tot weinig gebruik van non-verbale communicatie (bijv. door beperkt gebruik maken van oogcontact, wijzen, of laten zien van voorwerpen om een interesse te delen), en een verminderde mate van delen van plezier (Jones et al., 2014). Voorbeelden van stereotiepe of repetitieve gedragsspatronen op jonge leeftijd zijn bijvoorbeeld: weinig gevarieerd spel, stereotiepe bewegingen met de handen (fladderen), in hoge mate moeite hebben met veranderingen in het dagelijks leven, of een over- of onderreactie op sensorische prikkels (bijv. herhaaldelijk aan voorwerpen willen ruiken).

De oorzaken van ASS
De diagnose ASS wordt gesteld op basis van gedragskenmerken, maar neurobiologische processen lijken ten grondslag te liggen aan het waarneembare gedrag. ASS wordt beschreven als een complexe stoornis van de informatieverwerking in de hersenen, betreffende afwijkingen in de structuur, connectiviteit, en functies van de hersenen, maar er bestaat nog veel onduidelijkheid over de etiologie. Genen lijken hierin een belangrijke rol te spelen (Abrahams & Geschwind, 2008; Constantino & Charman, 2016; Geschwind & State, 2015), en het belang van de interactie tussen genen en omgeving wordt steeds vaker onderstreept door onderzoekers (Modabbernia et al., 2017; Visser et al., 2013). Mogelijk weerspiegelt het heterogene karakter van
ASS de veelheid aan genetische factoren, de invloed van omgevingsfactoren, en het samenspel van genen en omgeving. Alhoewel het genetische onderzoeksveld zich snel ontwikkelt, zal de toekomstige rol van genetische factoren in vroegdetectie en behandeling van ASS nog bepaald moeten worden.

**Box A.1 Casusbeschrijving Thomas**
Thomas was 19 maanden toen zijn ouders zich voor het eerst zorgen gingen maken om zijn ontwikkeling. De zorgen hadden met name betrekking op de vertraging in zijn spraak en motorische ontwikkeling. Gedurende het eerste levensjaar van Thomas had zijn moeder al het gevoel dat er iets ‘anders’ aan hem was, maar uitte deze zorgen niet bij een zorgprofessional. Met een leeftijd van 24 maanden werd fysiotherapie ingezet op advies van het consultatiebureau. De jeugdarts had ouders gerustgesteld wat betreft zijn taalontwikkeling en gaf aan dat de taalontwikkeling bij sommige kinderen vertraagd kan verlopen, maar waarschijnlijk zou Thomas op den duur zich de taal wel eigen maken. Op het kinderdagverblijf, waar Thomas 1 dag per week naar toe ging, werd gezien dat hij niet veel met leeftijdsgenoten speelde, maar hij was goed te hanteren en hij genoot ervan alleen te spelen.

Net na zijn 4e verjaardag ging Thomas naar een reguliere basisschool. In de klas liep hij vaak wat rond of rommelde met speelgoed, maar hij was nog weinig taakgericht en kwam nauwelijks tot spel. Ondanks dat hij tevreden en vrolijk was, toonde hij een voorkeur om alleen te spelen. De zorgen die zijn ouders al lang geleden hadden benoemd op het consultatiebureau, werden nu herkend door de leerkracht. Zijn ouders besloten toen professionele hulp in te schakelen, en de diagnose ASS werd gesteld toen Thomas 4 jaar en 5 maanden oud was. De behandeling Pivotal Response Training (PRT) werd ingezet. Er was sprake van een grote vertraging in diagnostiek en behandeling, ondanks de vroege zorgen bij ouders. Deze vertraging veroorzaakte stress en onzekerheid bij ouders over de vroege ontwikkeling van Thomas, en leidde tevens tot gemiste kansen om Thomas (en ouders) al op een zeer jonge leeftijd interventieprogramma’s aan te bieden.

**Het belang van vroege(re) herkenning**
De casusbeschrijving van Thomas (zie Box A.1) schetst het beeld dat ouders van jonge kinderen met ASS vaak al vroeg het gevoel hebben dat hun baby of peuter zich ‘anders’ ontwikkelt. De eerste duidelijke zorgen zijn vaak gerelateerd aan de spraaktaalontwikkeling die moeizaam op gang komt (Kozlowski, Matson, Horovitz, Worley, & Neal, 2011). Ook bij Thomas werd de diagnose veel later gesteld, ondanks dat ouders al vroeg zorgen ervoeren. Dit gebeurt vaak, ondanks het feit dat ASS in
veel gevallen al voldoende betrouwbaar kan worden vastgesteld rond 24 maanden (Kleinman et al., 2008; van Daalen et al., 2009). Vertraging in de diagnosestelling is onwenselijk. Ten eerste leidt een vertraging tot een uitgestelde inzet van de beschikbare interventies, wat mogelijk van invloed is op de lange termijn uitkomsten van het kind. Gezien de flexibiliteit van het zeer jonge brein, kunnen gerichte interventies een positievere effect hebben op de ontwikkeling en het leren (Dawson, 2008). Ten tweede ervaren veel ouders van jonge kinderen met ASS stress gerelateerd aan de opvoeding (Estes et al., 2009), waarbij de zorgen en onzekerheid over de ontwikkeling van hun kind de spanningen nog verder op kunnen voeren. Vertraging in de diagnosestelling leidt tevens tot uitstel van professionele ondersteuning voor ouders (Keenan et al., 2010). Ten slotte toont onderzoek kostenefficiëntie aan; door in vroeg stadium behandeling en begeleiding te bieden worden problemen voorkomen vanwaar de behandeling op latere leeftijd beduidend meer zou kosten (Peters-Scheffer et al., 2012).

De diagnostiek van ASS wordt nu volledig gebaseerd op de aan- of afwezigheid van bepaalde gedragskenmerken die vanaf circa 12 maanden tot uiting lijken te komen. Meer onderzoek naar de vroege signalen van ASS, met name gedurende de eerste levensjaren, kan bijdragen aan het opsporen van cognitieve en neurobiologische mechanismen die aan ASS ten grondslag liggen, en mogelijk helpen tijdig kenmerken van ASS te signaleren. Met het ontdekken van de vroege signalen en onderliggende mechanismen wordt het wellicht mogelijk om ASS eerder en betrouwbaar vast te stellen. Hierdoor zouden passende behandelingstrajecets tijdig kunnen starten, welke kunnen bijdragen aan een verbeterde ontwikkeling van een kind met ASS en kwaliteit van leven voor het kind en zijn/haar omgeving.

De belangrijkste bevindingen

_Vroege risicofactoren van ASS in de baby- en peutertijd_

Aangaande vroege signalen van ASS richten de hoofdstukken 2 en 3 zich specifiek op risico (en protectieve) factoren van het temperament en de ouder-kind interactie gedurende de vroege ontwikkeling. Beide hoofdstukken beschrijven onderzoek naar twee groepen jonge kinderen, namelijk de broertjes en zusjes van kinderen met de diagnose ASS (de zogenaamde hoog-risico brusjes), en de broertjes en zusjes van kinderen met een gezonde ontwikkeling (de zogenaamde laag-risico brusjes). Uit onderzoek blijkt dat broertjes en zusjes van kinderen met de diagnose ASS een verhoogd risico hebben op dezelfde diagnose; 18.7% van deze kinderen heeft zelf ook de diagnose (Ozonoff et al., 2011). **Hoofdstuk 2** beschrijft een studie waarin het temperament van deze kinderen werd onderzocht als een risicofactor voor ASS. Temperament wordt hier gedefinieerd als relatief stabiele eigenschappen betreffende activiteit, affect, aandacht, en zelfregulatie van een persoon, welke worden
beïnvloed door genetische, biologische en omgevingsfactoren (Shiner et al., 2012). Temperament werd gemeten middels vragenlijsten ingevuld door de ouders van de hoog-risico (HR, n=170) en laag-risico (LR, n=77) kinderen op een leeftijd van 8, 14, en 24 maanden. Met 36 maanden werd de diagnostische uitkomst van elk kind bepaald op basis van verschillende klinische testen en observaties. Dit resulteerde in vier groepen, namelijk HR-ASS (hoog-risico brusjes met diagnose ASS), HR-Atypisch (hoog-risico brusjes zonder diagnose ASS, maar met een afwijkende ontwikkeling), HR-Typisch (hoog-risico brusjes met een gezonde ontwikkeling) en LR (laag-risico brusjes). De resultaten tonen een verschil aan tussen de groepsanalyses (gebaseerd op gemiddelden van groepen) en de individuele analyses. Temperament dimensies verschillen significant tussen de groepen, met een meer extreem temperament (m.a.w. meer negatief affect, minder regulatieve vaardigheden) voor HR-ASS, gevolgd door HR-Atypisch, HR-Typisch en LR brusjes. Echter, op individueel niveau kon de diagnose ASS op basis van temperament niet betrouwbaar worden voorspeld. Opmerkelijk was dat de HR kinderen zonder een latere diagnose ASS wel betrouwbaar konden worden geïdentificeerd op basis van temperament. Meer specifiek, zij bleken relatief goed in staat om hun aandacht, emoties en gedrag te reguleren. Dit betekent dat de aanwezigheid van een ‘makkelijk’ temperament de kans op latere ASS erg klein maakt, terwijl de aanwezigheid van een ‘moeilijk’ temperament met onvoldoende precisie latere ASS voorspelt.

In hoofdstuk 3 worden, evenals in hoofdstuk 2, kinderen uit familiair hoog- en laag-risico gezinnen bestudeerd. In deze studie werden interacties tussen jonge kinderen en hun ouders onderzocht op een leeftijd van 5 en 10 maanden (n=113 HR, n=82 LR). Kind- en oudergedrag gedurende deze interacties werd met behulp van een codeerinstrument geëvalueerd, de PInTCI (Parent-Infant/Toddler Coding of Interaction). De PInTCI beschrijft de codering van kind- en ouderkenmerken die op basis van literatuur voorspellend zijn gebleken voor de latere ontwikkeling van het kind. Uit de voorlopige resultaten komt naar voren dat de HR kinderen in het tweede deel van hun eerste levensjaar minder initiatief tot contact toonden richting hun ouders dan LR kinderen. Dit verschil bleef bestaan nadat gecorrigeerd werd voor de invloed van geslacht en exacte leeftijd van het kind, maar verdween nadat gecorrigeerd werd voor opleidingsniveau van de ouders. Het groepsverschil was daarom mogelijk toe te rekenen aan verschillen in het opleidingsniveau. Er werden geen groepsverschillen gevonden in kindgedrag op 5 maanden, of oudergedrag op 5 en 10 maanden. Uit deze resultaten kan worden geconcludeerd dat de observatie van vroege ouder-kind interacties mogelijk informatief is voor het constateren van vroege afwijkingen in de sociale communicatie die gerelateerd zijn aan ‘kenmerken van het bredere fenotype’ (het zogenaamde Broader Autism Phenotyp, BAP) of een latere diagnose ASS. Gezien de diagnostische uitkomst van de kinderen in deze studie
nog niet bepaald is en gezien de mogelijke invloed van ouderkenmerken, kunnen we alleen met voorzichtigheid conclusies trekken en is vervolgonderzoek nodig.

**Vroegdetectie en behandeling van ASS**

De ernst van ASS, de stijgende prevalentie, en de groeiende evidentie voor effectieve vroege interventies benadrukken allen het belang van de vroegdetectie van ASS. Om vroege herkenning te stimuleren zijn verschillende vroegdetectie programma's geïmplementeerd. **Hoofdstuk 4** beschrijft één van de weinige studies die zich richten op de duurzaamheid van een vroegdetectie programma voor ASS. Dit programma bestond uit a) een training van potentiële verwijzers (zoals consultatiebureau artsen en medewerkers van integrale vroeghulp) in het signaleren van vroege kenmerken van ASS, b) het gebruik van een screeningsprotocol met een screeningsinstrument (Communicatieve en Sociale ontwikkelings Signalen | CoSoS4), en c) het vormen van een multidisciplinair diagnostisch team (Oosterling, Wensing, et al., 2010). Het programma bleek effectief gedurende de inzet van het programma (Januari 2004 – December 2006), maar zodra de actieve investering in het trainen van personeel werd stopgezet, verdween het positieve effect (Januari 2009 – December 2011). Verder bleek dat kinderen met cognitieve beperkingen meer kans hadden om verwezen te worden voor het derde levensjaar dan kinderen met sterkere cognitieve capaciteiten, ongeacht de inzet van het programma. Deze studie onderstreept het belang van een doorlopende investering in actieve screening en permanente training van verwijzers. Daarnaast is er een behoefte aan strategieën die zich richten op het detecteren van de groep kinderen met een hoger cognitief niveau, gezien deze groep mogelijk gemist wordt in de vroegdetectie van ASS.

Een verhoogde focus op de vroegdetectie en diagnostiek van ASS vraagt om de beschikbaarheid van effectieve vroege interventies voor de groep jonge kinderen met een diagnose. Er bestaat steeds meer evidentie voor de effectiviteit van interventies voor jonge kinderen met ASS (French & Kennedy, 2017), maar gezien de grote hoeveelheid uitkomstmaten die tot op heden wordt gebruikt is het lastig om interventiestudies met elkaar te vergelijken. Om conclusies te kunnen trekken over de effectiviteit van het type en de intensiteit van interventies, zou één uitkomstmaat systematisch moeten worden toegepast in verschillende studies. De eerste stappen richting een dergelijke uitkomstmaat zijn gezet door de ontwikkeling van de Brief Observation of Social Communication Change (BOSCC) (Grzadzinski et al., 2016). In **hoofdstuk 5** werd de bruikbaarheid van dit instrument onderzocht en vergeleken met een veelgebruikt instrument, de Autism Diagnostic Observation.

---

4 Aanvankelijk bekend als de Early Screening of Autistic Traits questionnaire (ESAT | Dietz et al., 2006; Swinkels et al., 2006)
Schedule (ADOS). Waar de ADOS oorspronkelijk is ontwikkeld om het huidige diagnostische beeld van een individu te bepalen, is de BOSCC specifiek ontwikkeld om subtiele veranderingen in autistisch gedrag over tijd te meten. Alhoewel de BOSCC en de ADOS gelijkenissen vertonen op het gebied van de inter- en intra-beoordelaarsbetrouwbaarheid, onafhankelijkheid van andere kindkenmerken, en sensitiviteit om veranderingen te meten, lijkt de BOSCC beter in staat om subtiele individuele veranderingen te bepalen.

Welke richting geven deze bevindingen voor toekomstig onderzoek en de klinische praktijk?

Temperament als risicofactor
Alhoewel jonge kinderen met een verhoogd familair risico op ASS zich op groepsniveau kenmerken door opvallendheden in temperament, lijkt dit niet duidelijk bij te dragen aan de vroegherkenning van ASS op individueel niveau. Desalniettemin kan het volgen van de vroege ontwikkeling van deze kinderen en het eventueel toepassen van een kortdurende interventie, vooral in het geval van het ontbreken van regulatieve eigenschappen, hun ontwikkeling positief beïnvloeden. Daarnaast kan informatie betreffende temperament bruikbaar zijn binnen de groep kinderen met een diagnose ASS. Op basis van temperament kunnen meer homogene groepen worden gevormd ter bevordering van een behandeling op maat. Dit proefschrift toont echter ook aan dat temperament op zichzelf niet voldoende voorspellend is voor de diagnose ASS op een individueel niveau, daarom is het van belang dat toekomstig onderzoek zich richt op de integratie van biologische en klinische maten die mogelijk gezamenlijk de voorspelbaarheid van ASS kunnen verbeteren (Bussu et al., 2018). Groep- en individuele analyses zijn daarbij van belang voor het verbeteren van specificiteit en sensitiviteit van markers voor de latere diagnose ASS. Aangezien gedragkenmerken van ASS in het eerste levensjaar slechts subtiel aanwezig lijken te zijn, is hersenenonderzoek op jonge leeftijd zinvol (middels bijv. elektro-encefalografie (EEG), magnetic resonance imaging (MRI), of near-infrared spectroscopy (NIRS)). De eerste onderzoeken tonen aan dat afwijkingen in hersenstructuur en -functie gevonden kunnen worden in het eerste levensjaar (Emerson et al., 2017; Hazlett et al., 2017; Varcin & Jeste, 2017). De vertaalslag naar de klinische praktijk moet echter nog worden gemaakt. Verder geldt voor beide studies in hoofdstuk 2 en 3 dat nader onderzoek nodig is naar de generaliseerbaarheid van deze HR studies naar de algemene populatie of andere risicogroepen (bijv. prematuriteit).

Ouder-kind interactie als risico (en beschermd) factor
De observatie van vroege ouder-kind interacties zou kunnen helpen in het signaleren van vroege signalen van ASS, en biedt de mogelijkheid om een belangrijke
omgevingsfactor, namelijk het gedrag van de ouder(s), te evalueren en positief bij te sturen. De voorlopige bevinding in hoofdstuk 3 dat reeds in het eerste levensjaar (kleine) verschillen aanwezig lijken zijn in het gedrag van de HR kinderen richting hun ouders, ondersteunt het belang van de start van een vroege behandeling. Nader onderzoek naar de rol van het opleidingsniveau van de ouder is echter noodzakelijk. Daarnaast is het van belang dat ouder-kind interacties op de lange termijn worden onderzocht, waarbij de focus ligt op voorspellende mechanismen van de ouder-kind interactie op de latere ontwikkeling en uitkomsten van het kind. De mogelijke invloed van andere factoren op de interactie (bijv. ouderlijke stress, emotieregulatie van het kind en de ouders, pre- en perinatale factoren) moeten daarbij worden onderzocht. Binnen toekomstig onderzoek naar de ouder-kind interactie speelt de rol van de ouder tijdens de interactie een essentiële rol als een mogelijke protectieve factor. Met andere woorden, kunnen bepaalde kenmerken van de ouder tijdens de ouder-kind interactie de ontwikkeling van het kind zo ondersteunen dat het risico op de ontwikkeling van psychopathologie vermindert? Dubbelblinde, gerandomiseerde gecontroleerde studies (RCTs) kunnen hierbij een belangrijke rol spelen. De eerste interventiestudies met een dergelijk design tonen al aan dat het stimuleren van de kwaliteit van de vroege ouder-kind interactie van invloed is op de latere ontwikkeling van het kind (Green et al., 2017). Daarnaast zal toekomstig onderzoek zich moeten richten op de psychometrische eigenschappen van het toegepaste codeerinstrument (PInTCI), het gebruik van het instrument op oudere leeftijden, en tevens de mogelijkheden van het instrument in een klinische, diagnostische setting. Ten slotte lijkt het zinvol om naast het gebruik van een dergelijk macroanalytisch instrument (gericht op kwalitatieve en kwantitatieve kenmerken aan de hand van een globale ratingschaal), ook het gebruik van microanalytische instrumenten (gericht op frequentie, duur en volgorde van gedragingen) te overwegen. Tevens kunnen nieuwe methodes worden overwogen waarbij gebruik wordt gemaakt van hoofdcamera’s of waarbij de oogbewegingen van ouder en kind worden gevolgd (dual eye tracking).

**Vroegdetectie programma’s op de lange termijn**

De motivatie om vroege risico of beschermende factoren te onderzoeken is om beter in staat te zijn afwijkingen in de ontwikkeling te signaleren en positief te beïnvloeden middels (preventieve) interventies. Dit proefschrift toont aan dat vroegdetectie programma’s een positief effect kunnen hebben op de leeftijd van verwijzing, maar kosteneffectieve strategieën moeten worden ontwikkeld om deze positieve effecten te behouden. Dit kan bijvoorbeeld middels e-learning modules, waarbij zorgprofessionals online kunnen worden onderwezen in de vroege signalen van ASS. Toekomstig onderzoek zal zich ook moeten richten op de vraag of de verlaging in de gemiddelde leeftijd van verwijzing daadwerkelijk resulteert in een snellere start
van de interventie, en wat de langetermijneffecten hiervan zijn op de ontwikkeling van het kind en zijn/haar kwaliteit van leven. Daarbij is specifieke aandacht voor de groep kinderen die nu mogelijk nog gemist wordt in vroegdetectie programma’s (zoals kinderen met hogere cognitieve capaciteiten of meisjes) van belang.

**Uitkomstmaten in interventiestudies**

Vroege signalering is een voorwaarde voor vroege interventies, en voor verschillende interventies zijn reeds positieve effecten op de lange termijn aangetoond (Estes, Munson, et al., 2015; Green et al., 2017; Pickles et al., 2016). Dit proefschrift toont aan dat de BOSCC een veelbelovende uitkomstmaat is voor gebruik binnen interventiestudies, maar nader onderzoek naar dit instrument is gewenst. Alhoewel de BOSCC een belangrijke rol zou kunnen spelen als een ‘key’ uitkomstmaat om interventies met elkaar te vergelijken, lijkt het belangrijk om in aanvulling op dit instrument ook andere uitkomstmaten toe te voegen die zich bijvoorbeeld richten op de specifiek aangeleerde gedragingen van een interventie. Op deze manier worden zowel globale als specifieke uitkomstmaten ingezet, en kunnen ook de onderliggende mechanismen van de interventie (wat werkt, voor wie?) worden onderzocht. Daarbij kan ook worden overwogen gebruik te maken van zowel context-gebonden als gegeneraliseerde instrumenten. Waar de context-gebonden meetinstrumenten worden afgenomen in contexten die vergelijkbaar zijn met de setting waarin de interventie plaatsvond (bijvoorbeeld de ouder-kind interactie), worden gegeneraliseerde instrumenten toegepast in contexten die verschillen van de bekende context (bijvoorbeeld de interactie met een leerkracht). Verder kan gebruik worden gemaakt van uitkomstmaten die zich meer richten op de ‘werkelijkheid’ (bijv. kwaliteit van leven), en van uitkomstmaten die een beeld geven van de hersenontwikkeling (bijv. EEG, MRI, of NIRS).

**Conclusie**

De studies gepresenteerd in dit proefschrift zijn het resultaat van verschillende nationale en internationale samenwerkingen, en hebben inzicht verschafte in diverse aspecten van vroege detectie, diagnostiek en behandeling van ASS. Ondanks dat een deel van de resultaten nog met voorzichtigheid dient te worden geïnterpreteerd en nader onderzoek nodig is, zijn de conclusies bruikbaar als input voor toekomstig onderzoek of reeds voor toepassing in de huidige klinische praktijk. Allereerst blijken de vroege kenmerken van het temperament evenals die van de ouder-kind interactie enigszins waardevol te kunnen zijn voor de herkenning van ASS. Hopelijk geven de wetenschappelijke studies in dit proefschrift aanzet tot nader onderzoek naar beide aspecten. De afgelopen 30 jaar is onze kennis over de vroege ontwikkeling van ASS zeer vergroot, maar toekomstig onderzoek zal zich verder moeten richten
op biologische en omgevingsinvloeden die uiteindelijk leiden tot symptomen van ASS, en het identificeren van signalen die voorafgaan aan de uiteindelijke klinische diagnose (Yirmiya & Charman, 2010).

Verder toont dit proefschrift aan dat een specifiek programma gericht op de vroege signalering van ASS effectief is gedurende de implementatie, met name voor kinderen met een ontwikkelingsachterstand. Echter, actieve investering is nodig om het positieve effect van een dergelijk programma te behouden. Deze resultaten hebben er onder andere toe geleid dat er in Nederland een website beschikbaar is gekomen voor zowel ouders als professionals met informatie over de vroege signalen van ASS. Daarnaast is een e-learning module ontwikkeld voor professionals in de jeugdhulp of in de (jeugd)gezondheidszorg die werken met jonge kinderen (0-4 jaar) (Dietz & Oosterling, 2017).

Ten slotte lijkt een nieuw ontwikkelde uitkomstmaat, de BOSCC, beter in staat subtiele veranderingen over tijd te meten dan de uitkomstmaat die tot op heden in interventiestudies wordt toegepast. Momenteel wordt internationaal nader onderzoek gedaan naar de bruikbaarheid van dit instrument en zijn onderzoekers het er over eens dat dit een veelbelovend instrument lijkt te zijn.

Een belangrijke vraag rest in hoeverre we op een zeer jonge leeftijd al categorische labels moeten en kunnen toewijzen. Het doel van vroegdetectie zou zich vooral moeten richten op het identificeren van vroege signalen, bespreekbaar maken van zorgen met het gezin, en ouders begeleiden in hoe de ontwikkeling van hun kind zo optimaal mogelijk te ondersteunen. Dit is in feite een pleidooi voor een verandering in visie van ‘diagnostiek en behandeling’ naar ‘voorspelling en stimulering’. Op deze manier kan interventie eventueel al worden ingezet voordat een diagnose wordt gegeven. Gezien het feit dat stoornissen vaak tegelijkertijd voorkomen en de kennis over de aanwezigheid van symptomen, zou een breed multidisciplinair onderzoek moeten worden uitgevoerd bij jonge kinderen (0-6 jaar) in plaats van de huidige trend in de kinder- en jeugdpsychiatrie waarin steeds meer onderscheid tussen stoornissen en disciplines ontstaat. Er zou dan gesproken kunnen worden van een bredere risicogroep op basis van vroege afwijkingen in (a) de algemene ontwikkeling, (b) communicatie en spraak, (c) sociale wederkerigheid, (d) motorische vaardigheden, (e) aandacht, (f) activiteitsniveau, (g) gedrag, (h) stemming, en/of (i) slaap. In de wetenschap wordt dit ook wel het ‘ESSENCE’ concept genoemd, wat een afkorting is voor ‘Early Symptomatic Syndromes Eliciting Neurodevelopmental Clinical Examination’ (Gillberg, 2010). (Preventieve) Interventie zou dan altijd beschikbaar kunnen zijn in het geval van een afwijkende ontwikkeling met aandacht voor een bredere risicogroep.
REFERENCES

Images by Peter Brouwers commissioned by the Dutch national network Autisme Jonge
Kind (AJK, Dutch for ‘Autism Young Child’).


Appendices


Appendices


Appendices


Schopler, E., Reichler, R. J., Bashford, A., Lansing, M. D., & Marcus, L. M. (1990). *The psychoeducational profile revised (PEP-R).* Austin, TX: Pro-Ed.


Appendices


DANKWOORD

Il Cammino, come la vita, non è una competizione.
Mai farsi prendere dalla voglia di fare troppo:
Il tuo organismo te ne chiederà presto conto.
Guardarsi intorno, osservare, fermarsi, assaporare.
Questo è ciò che ti insegnerà il Cammino.1

Abbazia di Sant’Antimo, Montalcino, Italia
Associazione Ecclesiale Italiana della Via Francigena e delle Antiche Vie di Pellegrinaggio

È finita! Met dit proefschrift als de eindbestemming, maar de reis ernaar toe heeft met name veel betekend. Ik heb enorm veel geleerd van de wereld van de wetenschap en het belang van een overdracht naar de klinische praktijk, en ik bleek onverwachts (relatief 😊) bestendig voor de bijbehorende stress en werkdruk tijdens het traject. Mijn interesse in de wetenschap is zelfs zodanig aangewakkerd dat ik na mijn promotie gedeeltelijk als onderzoeker aan het werk ben gegaan met als missie de kwaliteit van de zorg binnen de kinder- en jeugdpsychiatrie te verbeteren. Ik ben heel dankbaar dat ik gedurende het traject door zeer gewaardeerde experts werd omringd die mij inspireerden en motiveerden om dit proefschrift te realiseren. Ook de personen die binnen mijn sociale leven met mij ‘meereisden’ waren hierbij onmisbaar. Enkelens van hen wil ik hieronder expliciet bij naam noemen.

Allereerst wil ik mijn dank uitspreken naar alle gezinnen met jonge kinderen in binnen- en buitenland die deelnamen aan de onderzoeksprojecten. Dankzij hen zijn we meer te weten gekomen over de vroege herkenning van ASS. Ondanks dat deze gezinnen vaak al belast waren met allerlei uitdagingen in het dagelijks leven, was het opvallend hoe gemotiveerd ze waren om te participeren in wetenschappelijk onderzoek. Ik heb hier veel bewondering voor en ik ben hen zeer erkentelijk voor hun deelname.

Iris, dr. Servatius-Oosterling, mijn eerste co-promotor. Wat ben ik dankbaar voor jouw vraag in 2012 om aan het werk te gaan als junior onderzoeker binnen het Zebra-project. Eerder had je me al begeleid binnen het DIANE project (jouw

1 The Route, like life, is not a competition.
Never get seized by the wish to do too much:
Your body will soon demand an explanation.
Look around, observe, stop and taste.
That’s what the Route will teach you.
promotietraject), waarin jij mij als student-assistent leerde om op kritische wijze ouder-kind interacties te observeren. Grappig hoe dit vervolgens een belangrijk thema is geworden binnen mijn eigen promotietraject, ‘iets met autisme en OKI’? Gedurende het traject ben je mij blijven inspireren op het gebied van de wetenschap en de praktijk, en leerde je me een balans te vinden tussen gedrevenheid (of perfectionisme?!?) en tijd voor ontspanning. Ik heb grote bewondering voor jouw positivisme, enthousiasme, en doorzettingsvermogen. Zelfs met drie stralende kleintjes thuis, een wetenschappelijke en klinische baan, en een KP-opleiding, blijft jouw oprechte betrokkenheid en enthousiasme onverminderd. Ik ben dankbaar dat we zo'n vriendschappelijk contact hebben kunnen ontwikkelen. Ik hoop dit contact en onze gedeelde passie zowel in de wetenschap als in de praktijk samen voort te kunnen zetten. Dankjewel!

Jan, prof. dr. Buitelaar, mijn promotor. Al tijdens ons eerste gesprek vond ik jouw unieke gevoel voor humor opvallend, wat gedurende de afgelopen jaren verder werd voortgezet tijdens onze afspraken waarin altijd ruimte was voor een vleugje humor. Ik heb grote bewondering voor jouw kennis, de grote hoeveelheid onderzoeksprojecten die door jou zijn geïnitieerd, en ik was na onze afspraken altijd geïnspireerd om weer verder te gaan. Jij hield de rode lijn van het verhaal scherp en wist de kern altijd naar boven te halen. Veel dank voor het delen van jouw kennis en het vertrouwen dat je in me hebt gesteld.

Nanda, dr. Lambregts-Rommelse, mijn tweede co-promotor. De manier waarop jij ingewikkelde statistische methodes in begrijpelijke taal weet te herformuleren is ongelooflijk, en ik kan met zekerheid stellen dat alle artikelen naar een hoger niveau zijn getild dankzij jouw waardevolle, vlotte commentaar. Leuk dat zelfs jouw kleine Nora betrokken mocht zijn bij mijn promotietraject. Jouw warme, persoonlijke begeleiding enerzijds, en jouw scherpe inzichten anderzijds zijn een unieke combinatie waarvan hopelijk nog vele promovendi mogen profiteren. Dank voor jouw betrokkenheid!

Prof dr. C. de Weerth, Prof dr. J. van Bakel, en Prof dr. M. de Jonge, veel dank voor het zitting willen nemen in de promotiecommissie en uw kritische beoordeling van het manuscript. Overige leden van de promotiecommissie, dr. J. Visser, dr. P. Warreyn en dr. A. de Bildt, hartelijk dank voor het bestuderen van het manuscript en het voeren van de oppositie. Lijkt de verschuiving dan toch plaats te vinden? The world needs science and science needs women...

Gedurende mijn promotietraject heb ik de eer en het genoegen gehad om samen met andere onderzoekers een team te vormen, en daarbij wil ik in het bijzonder Ricarda bedanken. We zijn samen gestart aan dit wetenschappelijk avontuur als onderdeel van het Zebra-project. Ik heb ervan genoten het onderzoeksproject met jou op te zetten
en te coördineren, en het was zeer waardevol onze expertises te kunnen combineren. Dank voor jouw teamspirit en het kunnen delen van vreugde en frustraties. Daarnaast wil ik Loes, Manon, en Yvette danken voor hun grote betrokkenheid bij de gezinnen en hun hoogstaande ‘wegwijsheid’ in alle verzamelde data! Emma, thank you for joining the Zebra-team. I’m wishing you all the best finishing your PhD project, hang in there! Prof dr. Hunnius, Sabine, ik bewonder de manier waarop jij de BabyBRAIN groep leidt waarin openheid en samenwerking een belangrijke rol spelen. Ondanks dat jij niet direct betrokken was bij mijn gedeelte van het Zebra-project, heb ik jouw betrokkenheid op afstand altijd als erg prettig en warm ervaren. Giorgia, thank you for our fruitful collaboration and for sharing your experience with machine learning analyses. Ook de onderzoeksassistenten Karlijn, Renske, Nicole en Annemieke, en verschillende studenten hebben een zeer zinvolle bijdrage gedaan aan het project. Dank daarvoor!

Gezien het internationale karakter van het onderzoeksproject heb ik ook met experts buiten Nijmegen samengewerkt, waaronder onderzoekers bij King’s College London (Verenigd Koninkrijk), Birckbeck University of London (Verenigd Koninkrijk), Universiteit Gent (België), Karolinska Institutet (Zweden), Uppsala Universitet (Zweden), Uniwersytet Warszawski (Polen) en het Universitair Medisch Centrum Utrecht (Nederland). A big thank you to all collaborators at the EU-AIMS and Eurosibs research projects! Prof dr. Charman, Tony, I am very much honored to have had the opportunity to work together with you. I have learned a lot from your contributions to the field of early autism research and during our collaboration I was inspired by your scientific and clinical vision. Emily, thank you for taking the lead in this massive research project and for always having an answer available on both our practical and substantive questions.

Daarnaast in het bijzonder een groot woord van dank aan het onderzoeksteam in Gent, waar ik geregeld verbleef voor de ontwikkeling van het observatieschema gericht op vroege ouder-kind interacties. Chloë, dank voor jouw gastvrijheid! Ondanks de ellenlange observatiesessies die verbonden waren aan mijn bezoeken, was er ook altijd ruimte voor gezelligheid en was onze gedeelde motivatie groot om het observatieschema tot een succes te maken. Petra, dank voor het delen van jouw expertise in de wondere wereld van de observatieschema’s. Ik kijk uit naar onze toekomstige samenwerking en hoop dat we ‘ons kindje’ middels een sensitieve en gestructureerde aanpak verder kunnen doen ontwikkelen.

Natuurlijk kunnen de ‘kippen’ van Karakter niet ontbreken in dit dankwoord. De onderzoeksafdeling van Karakter wordt gerund door een wisselende groep slimme en ambitieuze vrouwen. Fijn om met jullie lief en leed te kunnen delen, dames! Anoek, Annemieke, Annick, Evita, Francesca, Iris, Jennifer, Jolanda, Kirsten, Leonie, Loes, Nienke, Manon, Margreet, Michelle, Mireille, Saskia, Shahrzad, Yvette en Yvonne; dank voor de gezelligheid binnen en buiten kantooruren! Ik hoop en verwacht dat er voor jullie
een mooie toekomst in het verschiet ligt. Daarnaast een groot woord van dank aan Helen; jij speelde een belangrijke rol in het gereed maken van de datasets die vanuit Quest Manager gegeneraliseerd moesten worden. Jouw snelle en nauwkeurige werk heb ik zeer gewaardeerd. Jane Sykes, veel dank voor de kritische en nauwkeurige tekstcorrecties. Gezien het observationele karakter van de studies binnen mijn promotietraject hebben verschillende (inter)nationale studenten tientallen uur geïnvesteerd in het observeren van het gedrag van ouders en kinderen. Met name wil ik bedanken: Aniek, Anouk, Danielle, Denise, Fieke, Joy, Karlijn, Micaela, Monica, Nicole, Stephanie, Suzanne en Wies. Manon, jij raakte ook als student betrokken en vanaf het begin had ik al het gevoel dat jij wel eens een stabiele factor zou kunnen gaan vormen. Inmiddels ben je werkzaam als junior onderzoeker binnen het Kippenhok en werken we samen als collega’s binnen de patiëntenzorg. Fijn om met jou de perikelen van het ‘PhD’ avontuur te kunnen delen, en vooral ook te kletsen over alle andere belangrijke zaken des levens. Ik zal me dan ook gesteund voelen door jou als paranimf aan mijn zijde. Ik hoop dat je een liefdevolle én ontspannen tijd tegemoet gaat in het zuiden!

Ook de klinische afdeling van Karakter heeft veel betekend binnen mijn promotietraject. De collega’s aldaar vormden de brug tussen de onderzoekers en de ouders van de jonge kinderen met ASS. Hun rol was van onschatbare waarde. Inmiddels werk ik met veel plezier binnen het team van het Centrum Jonge Kind en ben ik de collega’s van dit team zeer dankbaar voor hun warme betrokkenheid gedurende de laatste fase van het traject en de prettige samenwerking. Ik hoop dat we als team in de toekomst nog veel voor elkaar zullen krijgen om het zorgaanbod te optimaliseren en er zo nu en dan wat tijd over blijft voor een goede dosis gezelligheid. In het bijzonder wil ik Janne noemen; in het begin van mijn wetenschappelijke loopbaan hebben we menig discussie gehad over de interpretatie van ouder-kind interacties, waardoor ik elke keer opnieuw werd geïnspireerd. Vervolgens had ik de eer om bij de start van mijn klinische loopbaan bij Karakter met jou een koppel te vormen tijdens de intakes, waar ik leerde van jouw enorme kennis over deze doelgroep. Ik zal je missen bij Karakter en ik wens je alle goeds toe! Eric en Martine, dank voor jullie vertrouwen in mij als clinicus en onderzoeker, en de vrijheid die jullie mij daarbij geven.

Daarnaast ben ik de klinische collega’s buiten Karakter zeer erkentelijk voor hun rol in de recruitment van de jonge gezinnen door heel Nederland. Het landelijk expertisenetwerk Autism Jonge Kind (AJK) heeft daarbij een bijzondere rol gespeeld. Met veel plezier maak ik onderdeel uit van dit netwerk en raak ik bij elke bijeenkomst geïnspireerd en gemotiveerd door collega’s vanuit het hele land. Iris, Claudine, en Nienke, dank voor jullie grote inzet in het leiden van deze groep clinici en onderzoekers. Leden van de diagnostiekwerkgroep, dank voor de plezierige samenwerking. Ik kijk ernaar uit samen nog meer werk te verzetten!
Na een lange werkdag was het altijd een genot om het Utrechtse Schimmelplein op te fietsen, dankzij een groep zeer gezellige buren én een aantal fijne vrienden om de hoek. Lieve Floor, jij was er altijd, dank voor het delen, van alles. Zet die wijn maar vast koud augustus, dan kunnen we weer. Pleun, dank voor de vinger. Lieve Juune, dank voor het zijn van de eerste bestie van Cato. Jets, dank voor de energieke work-outs (inclusief de nodige afterdrinks) en de heerlijke etentjes bij Daan en jou thuis. Ik kijk uit naar september! Madelon, ik heb genoten van onze gezamenlijke tijd in Nijmegen én Utrecht, dank voor alle gezellige borrels, de sappige soapverhalen, en je betrokkenheid. Je bent altijd welkom als onze huishond.

Ook ben ik verschillende vrienden buiten Utrecht dankbaar. Fieke, ik ben zo blij met onze onvoorwaardelijke vriendschap en ik kan niet wachten tot het volgende onvergetelijke Fiekmir avontuur. Marlies en Carlijn, op dag 1 van de introductie was ons trio al een gegeven en dankzij jullie voelde ik me thuis in het verre ‘oosten’. Op naar een tweede lustrum in Göteborg! Joanne, dank voor jouw warmte, oprechtheid en liefde; onze vriendschap is mij erg dierbaar. Nikki, Anne-Sophie en Lotte, dank voor jullie steun tijdens onze fijne borrels en etentjes. Vietnamese vrienden, het lot bracht ons bij elkaar in een nachttrein op weg naar Sapa en ik word heel blij van onze gedeelde passie optimaal van het leven te genieten. Ubertolle wintersporters, van kaiserschmarrn tot aperol spritz, het is altijd fantastisch. Dat er nog velen mogen volgen. Anne, niet alleen familair, ook vriendschappelijk verbonden. Dank voor jouw warme betrokkenheid. Iris, dank voor onze vertrouwde vriendschap. Bijzonder om onze meisjes tegelijk te zien opgroeien. Floor en Sil, één van de grote voordelen van onze recente verhuizing is de kortere afstand naar jullie en jullie lieve Fiene en Kobe! Dank voor jullie steun. Ook de rest van de Nijmeegse Obelix crew, dank! Inmiddels trekken velen van ons weer terug naar het vertrouwde oude en ik kijk uit naar alle gezamenlijke (familie)dates. Dit alles was en is van onschatbare waarde. Duizendmaal dank, ook aan degenen die ik hier niet expliciet bij naam heb genoemd.

Lieve families Pijl, Marinussen, van der Velden, en Steenmeijer. Bedankt voor jullie steun en voor wie jullie zijn.

Johan en Lilian, ik had me geen lievere schoonouders kunnen wensen. Dank dat jullie er altijd zijn en met veel liefde voor Cato zorgen. Daan en Emma, dank voor de gezellige etentjes en Arnhemse uitstapjes. Lieve Mees, Anna, en Siem, het is zo bijzonder om jullie nu van dichtbij en naast Cato te kunnen zien opgroeien. Ik voel me vereerd dat ik onderdeel van jullie familie mag zijn.

Marleen en Tom; mijn dierbare zus en broertje. Zus, dank voor alles wat je vanaf het begin voor mij hebt betekend. Je bent er altijd, zowel voor zussenpraat als
doktersadviezen. De vraag of één van ons ooit nog naar de andere kant van het land verhuisd blijft onbeantwoord, en juist daaruit spreekt onze verbondenheid. Hoe dan ook, elke afstand valt door ons te overbruggen. Ik heb grote bewondering voor jouw kracht en hoop dat al jouw wensen mogen uitkomen. Bro, dank voor het zijn van mijn broertje en vriend. Laten we de tweewekelijkse etentjes erin houden, evenals de gezamenlijke vakanties en de (inmiddels sporadische) dronken momenten. Ik houd van jouw gezelligheid, humor, en bescheidenheid, en wens je heel veel liefde in Delft! Ben, het voelt alsof jij er altijd al bent geweest, dank voor jouw zorgzaamheid. Annemiek, je bent nu al onmisbaar. Lieve Ezra, ik geniet van jou!

Pap en mam, in woorden kan ik niet uitdrukken wat jullie steeds opnieuw weer voor mij betekenen. De open gesprekken, de lieve berichtjes, een brandende kaars tijdens spannende momenten, of praktische ondersteuning in de vorm van een oppasdagje op Cato of een klus in ons huis; jullie onvoorwaardelijke liefde en vertrouwen zijn altijd voelbaar. Ik ben trots op wat jullie ons mee hebben gegeven en ben jullie dankbaar voor het warme thuis dat jullie altijd geboden hebben en nog altijd bieden.

Mijn opa’s en oma’s, in liefhebbende herinnering. Zonder hen had ik hier niet gestaan.

Cato, jouw sprankelende ogen, jouw grote schaterlach, jouw enthousiaste kreten, jouw neusje tegen de mijne en jouw armpjes die mijn nek omklemmen. Ik had me nooit kunnen voorstellen hoe bijzonder het is om jouw mama te mogen zijn.

Max. Liefste. Ik ben je dankbaar dat je me altijd motiveert om alles uit het leven (en mezelf) te halen en daarop te vertrouwen, en ik bewonder jouw flexibele meebewegen tijdens mijn overdreven inzinkingen. Je maakt me gelukkig. Ik kijk uit naar augustus, en naar alle avonturen die we samen gaan beleven vanuit onze witte ‘villa’ (villekulla). Het leven is leuk met jou. JAD.
Appendices

PUBLICATIONS

Journal articles


Poster presentations


Oral presentations

Evaluating the BOSCC in an independent Dutch sample: A promising outcome measure? International Meeting for Autism Research, Baltimore (USA), May 2016.


Training

The application of the PInTCI coding scheme. EU-AIMS network, February 2016.

The application of a preliminary version of the BOSCC. Karakter Child and Adolescent Psychiatry University Center, 2014.
ABOUT THE AUTHOR

Mirjam Pijl was born on the 17th of July 1989 in Berkel en Rodenrijs, the Netherlands. After completing high school at the Melanchthon College in Rotterdam, she moved to Nijmegen in 2007. There she started her Bachelor and Master study ‘Pedagogische Wetenschappen’ at the Radboud University. She enjoyed student life, studied for six months in Sweden, did a clinical internship at Praktijk van Waterschoot in Breda, and graduated cum laude in 2012. During her Master study she worked as a research assistant at Karakter University Centre for Child and Adolescent Psychiatry, where the foundations for her interest in early autism and parent-child interaction were laid. After finishing her study, Mirjam combined her clinical work as a psychologist with working as a research-assistant coordinating an international research project, and after 12 months she decided to dedicate herself into a PhD at the Radboud University Medical Centre in Nijmegen. As a PhD candidate, Mirjam worked within the European project EU-AIMS where she studied the early development of infants at familial high and low risk of autism spectrum disorder. The results of this project are described in this thesis and were presented at several international conferences (e.g. INSAR, ESCAP). As part of her PhD Mirjam developed a coding scheme for the observation of early parent-child interactions, in collaboration with the Ghent University in Belgium. In addition, she attended multiple courses, supervised bachelor- and master students, was trained to conduct the Autism Diagnostic Observation Schedule (ADOS) and the Brief Observation of Social Communication Change (BOSCC), and gave training in the application of several coding instruments herself. She also continued her clinical work at the TOPGGz labeled Infants unit at Karakter University Centre for Child and Adolescent Psychiatry, where she currently works as a clinician and senior researcher.