AUTISM ASSESSMENT OF CHILDREN

VOLUME ONE

by

Kaysha Luces

A THESIS SUBMITTED TO THE UNIVERSITY OF BIRMINGHAM

FOR THE DEGREE OF DOCTOR OF CLINICAL PSYCHOLOGY

Department of Clinical Psychology

School of Psychology

The University of Birmingham

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

June 2020

Thesis Overview

This thesis comprises volumes and is submitted by Kaysha Luces for the Clinical Psychology Doctorate. Volume One includes two chapters. The first chapter is a meta-analysis of studies reporting on the diagnostic accuracy of autism screening tools in children born preterm. Eight studies were included for meta-synthesis. The key finding was that autism screening tools demonstrate poorer diagnostic accuracy in a preterm population compared to the general population. The second chapter is a pilot study that estimates the effect size of the correlation between shared intentionality and social competence as assessed by the ADOS-2^T Social Affect domain. A very small negative correlation was found. A power analysis recommended a sample size of 1716 would be required for a full study, were this the true size of the effect.

Volume Two includes four chapters, each containing one Case Practice Report. Chapter One details the formulation of a 23-year old female with a learning disability, referred for her experience of anger outbursts, formulated from CBT and psychodynamic perspectives. Chapter Two details a service evaluation exploring non-attendance of first psychology appointments and referral to treatment times. Chapter Three details a CBT intervention with 66-year old women experiencing generalised anxiety. Chapter Four reports a single case experimental design of a 6-year old boy with social anxiety. Chapter Five is an abstract of an integrated case study for 32-year old women experiencing complex post-traumatic stress disorder.

Dedication

For my son, Ezra.

My reason for everything.

Be whomever you wish to be. I will love you all the same.

Acknowledgements

First and foremost, I'd like to thank God for giving me a purpose and being an everlasting source of strength. Without his blessings my accomplishments would not have been possible.

Many thanks to my research supervisors, Dr Caroline Richards and Dr Andrew Surtees. Caroline has been a boundless source of inspiration. It has been a pleasure to work with another woman of colour who is so accomplished in her field. Her heartfelt support and guidance have extended beyond my thesis, and for this I am grateful. Andrew has provided continued honesty, patience and humour. His manner enabled me to ask any question, no matter how small safe in the knowledge that he would endeavour to guide me without judgement. Both Caroline and Andrew have been incredibly generous with their time. Their knowledge and words of encouragement transformed an initially daunting assignment into a valued and enjoyable experience. I thank them for holding me to such high expectations. This has enabled me to produce a quality of work beyond my own expectations. I'd also like to thank Catherine Laverty who has been a delightful research partner, consistently organised and attentive. Many thanks to Dr Christopher Jones for his continued support and guidance throughout.

I'd like to thank to all my clinical supervisors. They have all supported and challenged me, playing an active role in helping me on my journey to becoming a qualified Clinical Psychologist. A special thank you to all the clients and families I have worked alongside. They have taught me much about humanity and I am privileged and humbled that we have been able to learn from one another.

My acknowledgment would be incomplete without thanking my family. To my father who has demonstrated a lifelong passion to supporting my education. His limitless dreams instilled in me a belief that I can do all things through Christ who strengthens me. Undoubtedly, without him I would not be where I am today. To my mother who keeps me grounded. Her unconditional love has given be a safe base. With her I can share my failures, successes, fears and dreams. To my husband who is always proud of my achievements. Thank you for taking the lion's share of family duties. Finally, to my siblings who always provide laughter in times of stress.

TABLE OF CONTENTS

VOLUME ONE

CHAPTER ONE

SCREENING FOR AUTISM IN CHILDREN BORN PRETERM: A LITERATURE REVIEW AND META-ANALYSIS OF DIAGNOSTIC ACCURACY

Abstract 1

Introduction 2

Methodology 6

- Search Strategy 6
- Data extraction 13
- Assessment of Methodological Quality 15
- Data synthesis and statistical analysis 15
- Investigation of heterogeneity and identification of appropriate studies for analysis 17
- Bivariate measures of meta-analytic psychometric properties 17

Results 19

- Study selection 19
- Characteristics of included studies 19
- Methodological quality of included studies 20
- Univariate analysis of sensitivities and specificities 22
- Bivariate indicators of screening tool accuracy 30
- Comparisons of results in preterm children and the general population. 33

Discussion 35

- Overall diagnostic utility in preterm children 36
- Diagnostic utility in preterm children compared to the general population 36
- Exploration of difference in psychometric properties 37

Differences in psychometric properties between screens 37
Methodological strengths and limitations 38
Clinical implications 39
Conclusion 41
References 42

CHAPTER TWO

A PILOT STUDY OF THE RELATIONSHIP BETWEEN SHARED INTENTIONALITY AND SOCIAL COMPETENCE IN TYPICALLY DEVELOPING PRESCHOOL CHILDREN

Abstract 53
Introduction 54
Social Cognition and Theory of Mind 54
Scalable Measures of ToM 55
ToM precursors 56
The Early Social Cognition Scale (ESCogS) 57
Theory of mind and Clinical measurement of Social Communication 57
Theory of mind and social affect 58
Rationale 60
Aim 61
Method 61
Recruitment 61
Participants 62
Changes to recruitment strategies due to the COVID-19 pandemic 62
Participant Characteristics 63
Procedure 64
Measures 65
Statistical analysis 70

Results 71
Task Performance on the ESCogS 71
ADOS-2^{T} performance 73
The correlation between social competence and intentionality 74
Power analysis 75
Discussion 76
Comparison with previous findings 76
Threats to validity 77
Pilot studies and small sample sizes 77
Novel assessment tools 78
Strengths and Limitations 79
Clinical Considerations 80
Research considerations 80
Recommendations for future study 81
Summary 82
References 83

CHAPTER THREE

PRESS RELEASES

Autism screening tools are not as accurate in detecting autism in children born preterm. 92
References 94
Gold standard autism diagnostic assessment does not capture important autistic traits 95
Appendices 97
Volume II 102

VOLUME TWO

CHAPTER ONE

CASE FORMULATION OF A 23-YEAR OLD FEMALE WITH A LEARNING DISSABILITY, REFERRED FOR HER EXPERIENCE OF ANGER OUTBURTS: CBT AND PSYCHODYNAMIC PERSPECTIVES

Abstract 2

Assessment 3

Presenting Difficulties 3

Assessment Method 3

Assessment of the presenting difficulties 5

Therapeutic Relationship 7

Personal History and Circumstances 9

Cognitive Behavioural Therapy Formulation 10

Cognitive Behavioural Therapy Perspective 10

Longitudinal Formulation 11

Early Experience 11

Core Beliefs and Negative Assumptions 12

Negative Automatic Thoughts 14

Psychodynamic Formulation 18

Psychodynamic Perspective 18

Transference and Countertransference 19

Dynamic Hypotheses 21

Reflections 25

Reflections of Cognitive Behavioural Therapy Formulation 25

Reflections on Psychodynamic Formulation 25

References 28

CHAPTER TWO

A SERVICE EVALUATION EXPLORING NON-ATTENDANCE OF FIRST PSYCHOLOGY APPOINTMENTS AND REFERRAL TO TREATMENT TIMES

Abstract 31

Rationale: 31

Method: 31
Results: 31
Introduction 32
The audit cycle 32
Stage 1 of the audit cycle 32
Local policy context 33
Organisational factors in the service and rationale for the audit 33
Stage 2 of the audit cycle 34
The organizational impact of non-attendance 34
Reasons for non-attendance 35
Increasing attendance of first appointments 35
Patient choice 36
Organisational processes 36
Aims of the service evaluation 36
Method 37
Participants 37
Procedure 37
Stage 3 of the audit cycle 37
Sample 39
Ethical considerations 39
Measure 40
Method of analysis 40
Stage 4 of the audit cycle 40
Results 41
Attendance rates 41
Questionnaire results 43
1. Service-users' expectations regarding waiting times 43
2. Barriers to accessing psychological services 43
3. Client recommendations 44
Discussion with the LDCS-PD 44

The Team's responses to the above questions: 45
Discussion 47
Summary of findings 47
Recommendations 48
Stage 5 of the audit cycle 48
6. Continued monitoring 51
Stages 4 and 5 of the clinical audit 51
Strengths and limitations of the service evaluation 51
Reflections 53
References 54

CHAPTER THREE

A CASE STUDY OF COGNITIVE BEHAVIOURAL THERAPY INTERVENTION WITH A 66--YEAR OLD WOMAN EXPERIENCING GENERALISED ANXIETY

Abstract 58
Background 60
Referral information 60
History 60
Assessment 61
Initial Assessment 61
Standardised measures 61
The Geriatric Anxiety Scale (GAS) 61
The Penn State Worry Questionnaire (PSWQ) 62
Presenting factors 64
Frequency and severity 64
Formulation 64
Beliefs about worry 64
Precipitating factors 66
Predisposing factors 67

Relationships 69
Goals 69
Choice of theoretical framework 70
Intolerance of uncertainty 72
Positive beliefs about worry 72
Negative problem orientation 72
Cognitive avoidance 73
Relationships 73
Goals 74
Longitudinal formulation 74
Rationale 74
Conceptualisation 75
Intervention 78
Intervention rationale 78
Person centred approach 78
Aim of intervention 79
Socratic dialogue and Guided discovery 80
Psychoeducation 80
Between-session activities 81
The downward arrow technique 82
The continuum technique 82
Evaluation 83
Relapse prevention 83
Diane's feedback 83
Standardised measures 83
Reflections 84
Professional development 84
Implications for practice 84

References 86

CHAPTER FOUR
A SINGLE CASE EXPERIMENTAL DESIGN OF A 6-YEAR OLD BOY WITH SOCIAL ANXIETY

Referral 91

Initial Assessment 91

Background information 92

Early experiences 92

Family relationships 93

Presentation within sessions 94

Identified reinforcers 94

Functional behavioural assessment 94

Natural observation 94

Semi-structured interview 95

Beliefs about handwashing 95

Possible cognitive-developmental level 96

Onset 96

Topography [In-vivo handwashing] 96

Severity 97

Frequency 97

Developmental norms 97

Antecedent-Behaviour-Consequence Charts 97

Assessment measures 98

Subjective Units of Distress 98

The Preschool Anxiety Scale 98

Formulation 99

Behavioural formulation 99

Setting Events 99

Establishing Operations 100

Functions of behavior 100

Maintaining consequences 100

Social Anxiety 101
Treatment hypothesis 102
Intervention 102
Developmentally sensitive treatment choice 102
Structure of the intervention 103
Differential Reinforcement of Incompatible behaviour 103
Developmentally sensitive treatment 104
Method 105
Experimental design 105
Phase A (baseline) 105
Phase B (intervention) 105
Results 106
Visual analysis 106
Trend 106
Stability 107
Level 107
Statistical analysis 108
Autocorrelation 108
Measure of trend 108
Inferential testing 109
Discussion 110
Limitations 110
A-B design 110
Therapeutic approach 111
Multiple dependent variables 112
Strengths 113
Reflections 113
References **114**

CHAPTER FIVE

CASE STUDY OF A 32-YEAR OLD WOMAN WITH COMPLEX POST-TRAUMATIC STRESS DISORDER

Abstract 120

Appendices for Volume Two 121

Appendices for CPR 1 121

Appendices for CPR 2 127

Appendices for CPR 3……**.....140**

LIST OF FIGURES

VOLUME I

CHAPTER I

Figure 1.1.1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart detailing papers included and excluded at each stage of screening. 12

Figure 1.1.2. Proportion of studies with low high and unclear risk of bias (%) 21

Figure 1.1.3. Proportion of studies with low, high and unclear concerns regarding applicability (%) 21

Figure 1.1.4. Forest plot on study level and pooled sensitivities 22

Figure 1.1.5. Forest plot on study level and pooled specificities 23

Figure 1.1.6. Quantile-Quantile plot using both the random effects and fixed model of the distribution of study level sensitivity 24

Figure 1.1.7. Quantile-Quantile plot using both the random effects and fixed model of the distribution of study level specificity 24

Figure 1.1.8. Leave-one-out analysis for sensitivity (the effect of omitting each study on the pooled sensitivity) 25

Figure 1.1.9. Leave-one-out analysis for specificity (the effect of omitting each study on the pooled specificity) 26

Figure 1.1.10. Corrected funnel plot of study level sensitivity associated with publication bias using the trim and fill procedure (Duval &Tweedie, 2000) 27

Figure 1.1.11. Corrected funnel plot of study level specificity associated with publication bias using the trim and fill procedure (Duval & Tweedie, 2000) 28

Figure 1.1.12. Subgroup analysis of sensitivities across studies by screening tool used 29

.....30

Figure 1.1.13. Subgroup analysis of specificity across studies by screening tool used.....30

Figure 1.1.14. Summary Receiver Operating Curve for all screening tools.....31

CHAPTER II

Figure 1.2.1. Four-stage scale of early social cognitive skills taken from Ellis et al. (2020).....67

Figure 1.2.2. Individual participant performance on the Early Social Cognition Battery(ESCogS) and social competence assessed by the Autism Diagnostic Observation Schedule – Toddler module (ADOS-2^T) 75

VOLUME II

CHAPTER I

Figure 2.1.1. Genogram depicting violence (red zig zag line), separartion (red diagonal line), estranged (broken line) and mental health disorder (half black and white square) 10

Figure. 2.1.2. Longitudinal CBT Formulation based on, Emery and Greenberg (1985).....11

Figure 2.1.3.. Claire's formulation adapted from Beck's (1976) longitudinal formulationMaintenance Cycle.....16

Figure 2.1.4. Based on Beck, Emery and Greenberg's (1985) longitudinal CBT model 17

Figure 2.1.5. Malan's two triangles; triangle of conflict (left) and triangle of person 18

Figure 2.1.6. Claire's formulation (tentative) of her experience using the Triangle of Conflict from Malan (1975).....23

Figure 2.1.7. Claire's formulation (tentative) of her experience, using the Triangle of Person from Malan (1975).....24

CHAPTER II

Figure 2.2.1. Audit cycle taken from Leroux, McPherson and Richardson (1993).....32

Figure 2.2.2. Timeline of audit procedure 39

Figure 2.2.3. Monthly attendance rates of first psychology appointments during audit period 41

Figure 2.2.4 Percentage of clients seen within and beyond the 15working days 42

Figure 2.2.5. Barriers to attending appointments 43

Figure 2.2.6. Current reality tree adapted from Goldratt (1990) 46

Figure 2.2.7. Transition Tree adapted from Goldratt (1990) 47

CHAPTER III

Figure 2.3.1. Diane's genogram 63

Figure 2.3.2. Metacognitive formulation of worry associated with night waking, adapted from Wells (1995) 65

Figure 2.3.3. Diane's depiction of the timeline of her worry 68

Figure 2.3.4. Conceptualisation of Diane's anxiety related to night waking (adapted from Dugas & Robicaud, 1998) 71

Figure 2.3.5. Longitudinal formulation of Diane's presenting difficulties adapted from beck et al., 1979 77

CHAPTER IV

Figure 2.4.1.Genogram key based on McGoldrick and Gerson (1985). 93

Figure 2.4.2. Formulation of Flynn's handwashing based on Carr (2006) 101

Figure 2.4.3. Frequency of handwashing by day across baseline and intervention phase. 106

Figure 2.4.4. Frequency of handwashing by day across baseline and intervention phase 107

LIST OF TABLES

VOLUME I

CHAPTER I

Table 1.1.1. Search strategy carried out on PubMed in April 2019 7

Table 1.1.2. Stage one inclusion and exclusion criteria for screening titles and abstracts 10

Table 1.1.3. Stage 2 inclusion and exclusion criteria for screening full text 11

Table 1.1.4. Summary of Individual study characteristics: assessment tools, diagnostic outcomes, and QUADAS-2 quality criteria ratings 14

Table 1.1.5 Test performance indicators 16

Table 1.1.6. Summary of bivariate meta-analytic psychometric properties of screening tools in preterm population 33

Table 1.1.7.Summary of bivariate meta-analytic psychometric properties of screening tools in preterm and the general population 34

CHAPTER II

Table 1.2.1. Individual child demographics and characteristics 64

Table 1.2.2. Group scaling (individual performance at each stage of the ESCogS) and participant characteristics (age and Mullen composite score) 72

Table 1.2.3. Other observed patterns on the ESCogS 73

Table 1.2.4. Individual performance on the ADOS-2^{T} 74

Table 1.2.5. ESCogS scoring criteria 101

VOLUME II

CHAPTER I

Table 2.1.1. Claire's cognition themes adapted from Millings and Carnelley (2015)..... 13

CHAPTER II

Table 2.2.1 Reasons for non-attendance of first psychology appointments 42

CHAPTER III

Table 2.3.1. Diane's scores on standardised measures 84

CHAPTER IV

Table 2.4.1 An overview of the assessment phase..... 92

Table 2.4.2. Summary of information obtained by ABC charts 98

Table 2.4.3. Flynn's scores on the Preschool Anxiety Scale 99

Table 2.4.4. An overview of the intervention phase 103

Table 2.4.5. Number of data points below and above the trend line and in the first and second half of baseline, and comment on trend 109

Table 2.4.6. Mean frequency of handwashing during baseline and intervention, and t-test results 109

VOLUME ONE: RESEARCH COMPONENT

CHAPTER ONE

SCREENING FOR AUTISM IN CHILDREN BORN PRETERM: A SYSTEMATIC REVIEW AND META-ANALYSIS OF DIAGNOSTIC ACCURACY

Abstract

Background. Screening tools play a pivotal role in identifying children at risk of autism. However, research examining the utility of screening tools in high-risk populations such as children born preterm has not been synthesised.

Method. This meta-analysis aims to evaluate the diagnostic accuracy of multiple screening tools for autism when applied to preterm populations and compare this to published findings reporting the diagnostic accuracy of such tools in the general population. Ovid Embase, Ovid Medline, Ovid PsychINFO, CINAHL and PubMed were searched for English- written studies (from database creation to April 2019) reporting the diagnostic accuracy of autism-specific screening tools. We used the bivariate effects model to consider the effects of sensitivity and specificity by means of a receiver operator characteristic (ROC) analysis. Posterior prediction for sensitivity, specificity, likelihood ratios and diagnostic odds ration were also calculated.

Results. Screening tools were found to have just below acceptable diagnostic accuracy in detecting autism in preterm populations (AUC = 0.694). Compared to the general population, screening tools in preterm populations demonstrated lower sensitivity (preterm= 0.67, 95% CI 0.56 - 0.77, general population = 0.77, 95% CI 0.69-0.84), lower specificity (preterm= 0.82, 95% CI. 0.74 – 0.87, general population = 0.77, 95% CI 0.69-0.84), and lower diagnostic odds ratio (preterm= 9.73, 95% CI 5.55 – 17.05, general population = 596.09, 95% CI 174.32 – 2038.34).

Conclusion. Autism-specific screening tools demonstrated some utility for use in preterm populations. However, caution should be taken when interpreting such results, as diagnostic accuracy is lower than in the general population. Specifically, specificity and the overall diagnostic odds ratio are considerably poorer in preterm populations compared to the general population.

Clinicians need to be aware that traditional screening measures may present a high number of false positives (compared to the general population) in the pre-term population.

Introduction

Autistic people [1] present with atypical social communication and interaction, and restricted repetitive behaviour (American Psychiatric Association, [APA], 2013). Difficulties in these domains negatively impact adaptive functioning to varying degrees, with preschool autistic children demonstrating deficits in social and practical skills (Milne, McDonald, & Comino, 2013). The aetiology of autism remains unclear. However, multiple epidemiological studies have associated genetic and endogenous perinatal environmental risk factors with an increased likelihood of autism (Hisle-Gorman et al., 2018). Preterm birth is a well-established risk factor for autism (Agrawal, Rao, Bulsara, & Patole, 2018), with prevalence rates of autism as high as 7% (Agrawal et al., 2018) in preterm groups, compared to 1.47% in the general population (Centre for Disease Control and Prevention [CDC], 2014).

Preterm births are births that occur prior to 37 weeks gestational age. Such births may be further classified as extremely, very, moderate and late preterm (< 28, 28-32, 32-34, and 34-37 weeks respectively; World Health Organization, 2010). Developments in obstetric interventions and neonatal care, have increased the survival rates of infants born preterm (Iams, Romero, Culhane, &

[1] As advised by Autistica guidelines the current paper will use the descriptors ‘autism’ and ‘autistic children/people/individuals’. https://www.autistica.org.uk/about-us/media-communications-guide

Goldenberg, 2008; Office for National Statistics [ONS], 2019). With the reduction of preterm neonatal mortality rates, consequent medical and neurodevelopmental complexities have increased (Manuck et al., 2016; Wilson-Costello, Friedman, Minich, Fanaroff, & Hack, 2005), including intellectual disabilities, cerebral palsy, hearing loss and visual impairments, and Attention Deficit Hyperactivity Disorder (Broring et al., 2018; Himmelmann, Hagberg, & Uvebrant, 2014; Hirvonen et al., 2017; Hirvonen et al., 2018). These neurodevelopmental consequences are inversely associated with gestational age (Broring et al., 2018; Himmelmann, Hagberg, & Uvebrant, 2014; Hirvonen et al., 2017; Hirvonen et al., 2018). However, whilst a recent meta-analysis revealed high prevalence rates of autism in preterm children, higher rates were not associated with lower gestational age (Agrawal et al., 2018). Due to the increasing neurodevelopmental complexities within preterm populations, it is important that preterm children receive the appropriate assessments for timely identification or monitoring.

As well as heightened rates of broad neurodevelopmental variations, specific neurodevelopmental variations in verbal IQ, phonological fluency, visual-motor integration, executive functioning and somatosensory registration have been identified within preterm populations (Broring et al., 2008; Guarini et al., 2019; Luu, Vohr, Allan, Schneider, & Ment, 2011). Furthermore, atypical social development has also been associated with preterm birth, including specific variations in empathic development and social reasoning abilities (Campbell et al., 2015; Lejeune et al., 2016). The wide range of specific neurodevelopmental variations present in preterm populations may make formally diagnosing autism more complex, due to overlapping difficulties such as phonological fluency and social reasoning ability.

Considering the prevalence of neurodevelopmental difficulties identifiable in children born preterm, it is important that developmental screening and assessment tools are able to facilitate the identification of atypical development and accurately classify neurodevelopmental disorders within

this group. There are a variety of autism-specific screening and diagnostic tools available to assist in the identification of autism in young children, using methods such as observation and informant reports (Ibanez, Stone, & Coonrod, 2014). Screening tools may be categorized as level 1 or level 2 (Yong-Hwee, Young, Brewer, & Berlingeri, 2014). Autism-specific level 1 screening tools are administered to the general population in order to identify risk of autism (Yong-Hwee et al, 2014). Autism-specific level 2 screening tools are administered to children identified for specific risk of developmental delay and facilitate in differentiating between autism and other conditions such as speech delay (Yong-Hwee et al, 2014). Diagnostic tools are more detailed assessments that aid in the identification of diagnostic categories such as autism. Due to the importance of such level 1 and level 2 screening tools in identifying children who require further diagnostic assessment, it is important that we have a good understanding of their psychometric properties, particularly in high-risk groups, such as children born preterm.

A recent meta-analysis summarized evidence pertaining to the diagnostic accuracy of autism-specific level 1 screening tools in children within the general population (Sanchez-Garcia, Galindo-Villardon, Nieto-Librero, Martin-Rodero, & Robins, 2019). They found that screening tools consistently demonstrated high levels of diagnostic accuracy, with good accuracy in detecting autism in the general population (pooled sensitivity = 0.72; 95% CI: 0.61–0.81) and very good accuracy in correctly identifying an absence of autism (pooled specificity = 0.98; 95% CI: 0.97–0.99). Consequently, this meta-analysis evidenced good Diagnostic Odds Ratio (596.09;95% CI: 174.32–2038.34), pooled positive likelihood ratio (131.27;95% CI: 50.40–344.48) and negative likelihood ratio (0.22: 95% CI: 0.13–0.45) for screening tools in the general populations. However, as estimates were pooled across multiple screening tools, firm conclusions of the diagnostic accuracy of individual screening tools remains unclear.

Although the meta-analysed screening psychometrics produced by Sanchez-Garcia et al. (2019) are promising, sensitivity and specificity of tools can vary alongside prevalence rates, due to clinical variability in subgroup symptoms and severity (Leeflang, Rutjes, Reitsma, Hooft, & Bossuyt, 2013). For example, a meta-analytic review of developmental assessments found that, when applied to a preterm group, the pooled sensitivity of tools was low, with almost 50% of children with deficits in cognitive functioning being classified as having typical neurodevelopmental functioning (Wong, Santhakumaran, Cowan, & Modi, 2016). Poor sensitivity is likely to result in children with atypical development being incorrectly categorized as typical, and therefore denied further diagnostic assessments and potentially promising interventions. In contrast, poor specificity is likely to result in typically developing children being classified as atypical, causing misdiagnosis (Van Den Heuvel et al., 2016). Clinically, such false screening results may exacerbate the already heightened levels of distress within this population (Tu, He, Chen, Shi, & Li, 2012; Gurian, Kinnamon, Henry, & Waisbren, 2006). Therefore, in order to ensure timely and appropriate access to autism assessment and minimise parental distress, it is essential that the psychometric properties of autism screening tools are established for preterm populations.

In summary, children born preterm are at a heightened risk of autism. There are a variety of autism-specific screening tools applicable to the general population, with high diagnostic accuracy that facilitate the timely diagnosis of autism and allow for early intervention. However, the diagnostic accuracy of screening tools in preterm populations has historically received less attention. More recently, epidemiological studies have explored the predictive value of developmental tools when applied to preterm populations, with some authors reporting poorer diagnostic accuracy than in the general population. However, to date the psychometric properties of autism-specific screening tools have not been synthesised. Increased understanding of this area will

provide clinically useful information that may be used to guide autism-specific assessments in preterm born children.

In this meta-analysis, we will synthesise the available empirical data to establish the diagnostic accuracy of screening tools in detecting autism in children born preterm. We will:

1. Conduct a systematic search to identify all studies that have examined the diagnostic accuracy of autism-specific screening tools.
2. Synthesise existing empirical data from appropriate studies to produce pooled estimates of sensitivity, specificity, positive and negative likelihood ratios and a Diagnostic Odds Ratio.
3. Compare obtained estimates to published estimates from the children in the general population (Sanchez-Garcia et al., 2019).

Methodology

Search Strategy

The databases Ovid Embase, Ovid Medline, Ovid PsychINFO, CINAHL and PubMed were searched from database creation to the second week of April 2019. Search terms were included under the categories of; 1) autism 2) preterm 3) diagnosis and 4) screening. Search terms for autism, diagnosis and screening were adapted from Sanchez-Garcia et al.'s (2019) recent meta-analysis of the diagnostic accuracy of autism screening tools in the general population to allow comparison between the obtained estimates. Search terms for prematurity were adapted from Pascal et al.'s (2018) recent meta-analysis of neurodevelopmental outcomes in preterm and very low birth weight

infants (Pascal et al., 2018). These terms were combined using the 'AND' function (see Table 1.1.1 for a list of the terms used in PubMed, and Appendix A for terms used in Ovid Embase, Ovid Medline, Ovid PsychINFO, and CINAHL). The complete search strategy for all databases is documented in appendix A as per PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). All papers were exported and screened.

Table 1.1.1

Search strategy carried out on PubMed in April 2019

Search terms	
#1 Autism	"Autistic Disorder" (exp/TI/A) OR "Autistic Disorders" (/TI/A) OR "Autism" (TI/A) OR ASC (TI/A) "Child Development Disorder" (TI/A) OR "Child Developmental Disorders" (TI/A) OR "Pervasive Developmental Disorders" (TI/A) OR "Pervasive Developmental Disorder" (TI/A) OR "Autistic Spectrum Disorder" (TI/A) OR "Autistic Spectrum Disorders" (TI/A) OR "ASC Spectrum Disorder" (TI/A) OR "ASC Spectrum Disorders" (TI/A) OR "Autistic Spectrum Condition" (TI/A) OR "Autistic Spectrum Conditions" (TI/A) OR "ASC Spectrum Condition" (TI/A) OR "ASC Spectrum Conditions" (TI/A) OR ASC (TI/A) OR ASD (TI/A)
#2 Preterm	"Premature (exp) OR preterm (TI/A) OR pre-mature (TI/A) OR pre-term (TI/A) OR "nearly term" (TI/A) OR mid-term (TI/A) OR late-preterm (TI/A) OR "low birth weight" (TI/A) OR "Extremely low birth weight" (TI/A) OR "Very low birth weight" (TI/A) OR LBW (TI/A) OR ELBW (TI/A) OR VLBW (TI/A) OR "Extremely preterm" (TI/A)
#3 Diagnosis	Diagnosis (SH/TI/A) OR "Early diagnosis" (/SH/TI/A) OR Detection (TI/A) OR "Early detection" (TI/A) OR "Early Identification" (TI/A) OR "delayed identification" (TI/A) OR "Early Intervention" (TI/A) OR "Early Prediction" (TI/A) OR "predictive validity" " (TI/A) OR "false positive" " (TI/A) OR "false negative" " (TI/A)
#4 Screening	Screening (TI/A) OR "Early Screening" (TI/A) OR "Mass Screening" (SH/TI/A) OR "Screening Tool" (SH/TI/A) OR "Screening Tools" (SH/TI/A) OR "Screening Test" (SH/TI/A) OR "Screening Tests" (SH/TI/A) OR "Screening Instrument' (SH/TI/A) OR "Screening Instruments" (SH/TI/A) OR Checklist (TI/A) OR Checklists (TI/A) OR Follow-up (TI/A)

Note: TI title, A abstract, exp explode, SH subject heading

For inclusion, papers had to report the results of an autism screen and either a research[2] or clinical diagnosis[3] of autism in children born preterm. Where studies included complete data sets on multiple screening tools, these were all included. Studies that included duplicate data (using identical screening tools in the same sample of participants) were excluded to prevent any unrepresentative increase in homogeneity. Studies that used duplicate samples, but different screening tools were included in order not to lose relevant data. However, it should be noted that including multiple outcomes from the same study will increase the apparent sample size and decrease the confidence intervals for the weighted average effect.

The systematic search aimed to find papers that focused on the diagnostic accuracy of autism. Therefore, all papers that included sufficient data to construct a 2x2 contingency table (screening (positive/negative) x diagnosis (positive/negative)) were included. In the event that papers reported only the percentage of individuals who screened positive and the percentage of

---

[2] *Research diagnoses are used only for the purpose of research. An autism diagnosis is assigned if the child reaches a stipulated criterion on a validated diagnostic assessment tool.*

[3] *Clinical diagnosis is used for the purpose of diagnosis. Clinical diagnosis of autism involves qualified professionals such as Psychiatrists and Psychologists interpreting data from diagnostic assessments tools, developmental histories and clinical judgement. These assessments take place in clinical settings such as community clinics. Data on diagnosis may be collected retrospectively from children's clinical files.*

individuals who met criteria for an autism diagnosis, authors were contacted to request rates of true positives and negatives, and false positives and negatives. In cases where papers only assessed for autism in those with a positive screen, data were extracted for review but not included in the meta-analysis.

A manual search of the reference lists of all included studies was conducted. The reference list of a recent meta-analysis exploring the prevalence of autism rates in children born preterm (Agrawal et al., 2018) was also manually searched, however no further studies were identified. Tables 1.1.2 and 1.1.3 document the inclusion and exclusion criteria at title/ abstract screening (stage one) and full text screening (stage two). Each study was examined at stage one to establish if it met the established criteria. When all inclusion criteria were met at stage one, the study was then examined at stage two. In the event that the study met any one of the exclusion criteria, it was excluded from further examination. In the event that a study contained unclear or missing information at stage one it was examined at stage two of screening. A complete flow chart of the screening process can be seen in Figure 1.1.1.

Table 1.1.2

Stage one inclusion and exclusion criteria for screening titles and abstracts

Inclusion Criteria	Exclusion Criteria	Justification
Empirical studies	Literature reviews, book chapters, and conference abstracts	Raw data were needed to analyse the psychometric properties of screening tools
Available in English	Not available in English	No resources to translate papers were available
Human participants	Animal subjects	The topic was not applicable to animal samples
Refers to autism	Does not refer to autism	A diagnosis of autism was required to calculate the necessary psychometric properties of screening tools
Study included ex preterm child or adult outcomes	Study included maternal, paternal or neonatal factors	This study was not focused on the factors that contribute to autism
	Single case design	Single case designs did not include enough data to calculate psychometric properties of screens

Table1.1. 3

Stage 2 inclusion and exclusion criteria for screening full text

Inclusion Criteria	Exclusion Criteria	Justification
Prematurity, gestational age or birth weight reported	Prematurity, gestational age or birth weight not reported	Topic focused on preterm populations
Data not duplicated in an included paper with more complete data	Data duplicated to a less complete standard than an included paper	To prevent over inflation of homogeneity results, and inaccurate pooled estimates
The exact number or participants were reported at each stage of the study	The exact number or participants were not reported at each stage of the study	Sample numbers were required to calculate psychometric properties
Reported confirmatory autism diagnoses using a review of medical notes, diagnostic tool, or by either a Psychologist or Psychiatrist	Did not report confirmatory Autism diagnosis by review medical notes, diagnostic tool, or by either a Psychologist or Psychiatrist	Diagnosis results required to calculate psychometric properties
Autism diagnoses were linked with screens sufficiently to identify true positive screens	Autism diagnoses were not linked with screen sufficiently to identify true positive screens	Sufficient data were required to provide pooled estimates of psychometric properties
Or used at least one of the following measures of accuracy: positive or negative predictive values, sensitivity and/or specificity, and area under the curve	Or did not use any of the following measures of accuracy: positive or negative predictive values sensitivity and/or specificity, and area under the curve	
	Participants were included in the study due to medical condition e.g. diabetes or surgical/medical procedure	Including a sample consisting of a high proportion of children with a medical condition may have resulted in an unrepresentative pooled sample.

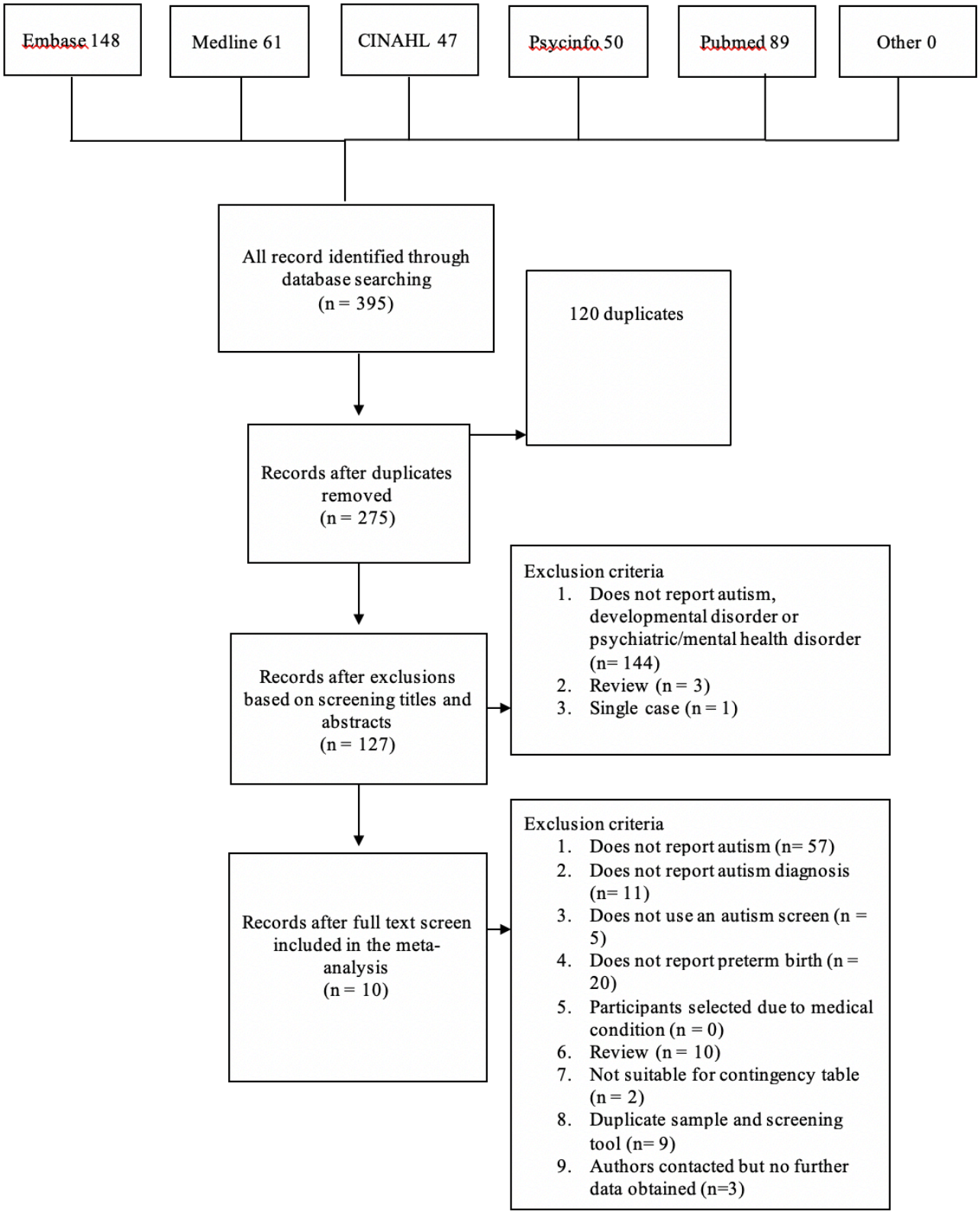


Figure 1.1.1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart detailing papers included and excluded at each stage of screening.

Data extraction

The first author extracted the following data from each study: study location, screening tool and diagnostic tool used, country of origin, sex of participants, age at assessment and average gestational age. Rates of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) were also extracted and are shown in Appendix B. In the event that these rates were not directly available from the paper, they were calculated using sensitivity, specificity and number of participants. Table 1.1.4 shows the studies in alphabetical order. Studies that only followed up positive screens with diagnostic assessment did not provide sufficient data to be included in the meta-analysis. As this meta-analysis calculated and synthesised psychometric properties, the rates of true and false negatives were required alongside rates of true and false positive. Where negatives screens were not followed up with diagnostic assessment, rates of true and false negatives were not available.

Table 1.1.4

Summary of Individual study characteristics: assessment tools, diagnostic outcomes, and QUADAS-2 quality criteria ratings

Studyr	Country	N	Sex		Av. Age (months)	Av. GA	Screening tool	Diagnostic method	Prevalence of autism	Risk of bias				Concerns regarding applicability		
			M	F						Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
*Beranova et al (2017)	Czech Republic	131	88	69	NR	28	CSBSD	ADOS-G	10%							
*Boon et al (2018)	USA	528	324	231	32	27	PDDST-II-DCS	ADOS-2, CARS 2 & ARD2	4%							
*Dudova et al (2014)	Czech Republic	134	88	69	NR	28	MCHAT	ADOS-G	10%							
Gray et al. (2015)	Australia	14	-	-	-	-	MCHAT & MCHAT-FI	Clinical assessment	-							
*Johnson et al. (2010)	UK & Ireland	173	78	95	NR	-	SCQ	DAWBA	6%							
*Kim et al (2016)	USA	831	-	-	24	-	M-CHAT	ADOS-2 & ADI-R	7%							
*Padilla et al (2017)	Sweden	84	46	38	78	-	SRS	Clinical assessment	12%							
Pritchard et al (2016)	Australia	15	-	-	-	<29	MCHAT-FI	ADOS-G	-							
*Verhaeghe et al (2016)	Belgium	47	32	21	151	-	SRS	ADOS-2 & OR ADI-R	45%							
*Verhaeghe et al (2016)	Belgium	45	32	21	151	-	SCQ	ADOS-2 & OR ADI-R	44%							
*Yaari et al (2016)	Israel	99	60	50	8	-	AOSI	ADOS-2^{T}	8%							

Note. included in the meta-analytic sample,*- not available from paper, Av. GA average gestational age, SCQ Social Communication Questionnaire, SRS Social Responsiveness Scale PDDST-II-DCS Pervasive Developmental Disorders Screening Test - II, Stage 2, Developmental Clinic Screener, M-CHAT Modified-Checklist for Autism in Toddlers, M-CHAT-FI Modified-Checklist for Autism in Toddlers-Follow-up Interview, CSBSD Communication an d Symbolic Behavior Scales Development, AOSI Autism Observation Scale for Infants. ADOS-G Autism Diagnostic Observation Schedule-Generic, ADOS-2 Autism Diagnostic Observation Schedule- 2nd edition, ADOS-2^{T} Autism Diagnostic Observation Schedule- 2nd edition Toddler Module , CARS Childhood Autism Rating Scale, ASRS, Autism Spectrum Rating Scale, - data not reported , low risk, unclear risk. high risk

Assessment of Methodological Quality

In concordance with Sanchez-Garcia et al. (2019), and as recommended by the Cochrane Collaboration (Reitsma et al., 2019) and NICE guidelines (National Institute for Health and Care Excellence, 2012), the included studies were assessed for quality using the QUADAS-2 Tool (Quality Assessment of Diagnostic Accuracy Studies-2, Whiting et al., 2011). The QUADAS-2 is a 14-item quality checklist that encompasses four domains of bias that may occur in studies exploring diagnostic accuracy. The four domains in which bias may be introduced are selection of participants, conduct or interpretation of the screening tool, conduct or interpretation of the diagnostic assessment, and flow and timing between the screening tool and diagnostic tools being administered. Two authors rated each study on all of these criteria using the recommendations described by Whiting et al. (2011). Inter-rater reliability was calculated using Cohen's k . There was moderate agreement between the two raters ($k = 0.65$, $p < .0005$). All discrepancies were discussed until an agreement was made.

Data synthesis and statistical analysis

The data analysis strategy follows the guideline for the Centre for Applied Psychology, University of Birmingham, and are paraphrased below.

Each of the screening tools included for analysis have applied a cut-off value to classify results into positive or negative for high risk of autism. This dichotomisation enabled us to represent the comparison between screening and diagnostic tests similar to Sanchez-Garcia et al. (2019). Common indicators of screening tool performance are; test sensitivity and specificity, and positive and negative likelihood ratios. In this meta-analysis univariate test performance indicators for sensitivity and specificity were calculated from true positives,

false positives, true negatives and false negatives rates (described in Table 1.1.5). Such indicators allow for exploration of the heterogeneity of included studies.

Ideally meta-analyses of diagnostic accuracy would pool estimates across multiple studies examining the same screening tools. However, as the amount of literature exploring diagnostic accuracy in preterm populations is limited, this meta-analysis pooled estimates across multiple studies using a variety of screening tools. This is consistent with the approach taken Sanchez-Garcia et al. (2019).

Table 1.1.5

Test performance indicators

	Formula	**Definition**
Sensitivity (true positive rate, TPR)	**TP/(TP + FN)**	**Proportion of positive test results among autistic subjects. High scores reflect better sensitivity**
Specificity (true negative rate, TNR)	**TN/(TN + FP)**	**Proportion of negative test results among non-autistic subjects. High score reflects better specificity**
Likelihood ratio of a positive test result (LR+)	**sensitivity/(1-specificity)**	**Ratio of a positive test result among autistic subjects to the same result in the "healthy". High scores reflect better predictive accuracy.**
Likelihood ratio of a negative test result (LR-)	**(1-sensitivity)/specificity**	**Ratio of a negative test result among non-autistic subjects to the same result in the "healthy". Low scores reflect better predictive accuracy.**
Diagnostic Odds Ratio	$\frac{\frac{TP}{FN}}{\frac{FP}{TN}}$ $=\frac{\left(\frac{PPV}{(1-PPV)}\right)}{\left(\frac{(1-NPV)}{NPV}\right)}$ $=\frac{LR+}{LR-}$	**Ratio of a test is the ratio of the odds of positivity in autism relative to the odds of positivity in the non-autistic. Alternatively, the DOR can be viewed as the ratio of the odds of autism, in test positives relative to the odds of autism in test negatives.** **The diagnostic odds ratio ranges from zero to infinity, although for useful tests it is greater than one, and higher diagnostic odds ratios are indicative of better test performance.**

Note. The abbreviations TP, FP, FN, and TN denote true positives, false positives, false negatives, and true negatives.

Investigation of heterogeneity and identification of appropriate studies for analysis

Diagnostic accuracy meta-analyses examine the heterogeneity of sensitivity and specificities separately. In line with this method, the levels of heterogeneity were examined using Higgins I^2 (values below 50% indicate heterogeneity) and Tau^2 (values of $p < 0.05$ indicate significant heterogeneity) for sensitivities and specificities. As Higgins I^2 and Tau^2 indicate unacceptable levels of heterogeneity across study sensitivities and specificities, a "leave-one-out" analysis was also conducted. Publication bias was assessed using visual and statistical analysis of Quantile-Quantile (QQ) plots, and a trim and fill procedure (Duval & Tweedie, 2000). Finally, subgroup analyses were performed for screening tools used.

Bivariate measures of meta-analytic psychometric properties

Using only univariate test performance indicators may lead to biased results, as they do not validly represent a test's discriminatory performance. Univariate indicators of sensitivity and specificity cannot be examined in isolation as each accounts for only part of the discriminatory evidence; describing only the classification of high risk or low risk of autism, in that as one variate increases the other decreases (Deeks, 2001). Therefore, a bivariate meta-analytic model is preferable (Zwinderman & Bossuyt, 2008).

The bivariate model can be used to consider the effects of sensitivity and specificity simultaneously by plotting the true positive rate as a function of the false positive rate (specificity) by means of the summary Receiver Operator Characteristic (sROC). The sROC model has been shown by Harbord and Whiting (2009) to be equivalent to Rutter and Gatsonis' (2001) HSROC linear mixed effect model used by Sanchez-Garcia et al. (2019). The sROC curve is calculated by plotting the point estimate of the pair of sensitivity (y-axis) and false positive rates (x-axis) together with a 95% confidence interval for both axes. The

average trajectory of these point estimates (weighted by study precision measure) across the range of the included data was calculated and the trajectory was extrapolated beyond this range and represented visually. The relationship between sensitivity and specificity was also measured quantitatively by obtaining the Area Under the Curve (AUC). The AUC is a global indicator or test performance. The AUC has several interpretations, including (a) the probability that a randomly selected patient autistic person will have a higher test result than a randomly selected control non-autistic person and (b) the average sensitivity at all levels of specificity.

Likelihood ratios can only be interpreted with reference to the base rate of the condition that is being screened (Zwinderman & Bossuyt, 2008). Therefore, we require an aggregation procedure to combine sensitivity and specificity, and likelihood ratios into one global measure of test performance. The Diagnostic Odds Ratio (DOR) provides such a measure of global test performance that is not affected by base rates (Glas, Lijmer, Prins, Bonsel, & Bossuyt, 2003). The DOR can be interpreted as the odds ratio of a positive test result from an autistic person relative to the odds of a positive result from a non-autistic person. Although a useful measure of global test performance, the DOR cannot be used to quantify the true positives and true negative rates of tests. Therefore, two tests with the same DOR may have different values for sensitivity and specificity.

In summary, heterogeneity was examined using univariate indicators of sensitivity and specificity. The overall meta-analytic psychometric properties for sensitivity, specificity, likelihood ratios, DOR and AUC were calculated from the Reitsma et al. (2005) bivariate model.

Results

Study selection

The first aim of this meta-analysis was to conduct a systematic search to identify all studies that have examined the diagnostic accuracy of autism-specific screening tools. The initial systematic search identified 395 papers. This included 120 duplicates which were eliminated, leaving 275 non-duplicated papers. Inclusion/exclusion criteria were applied to review each title and abstract for suitability; 148 were excluded and 127 remained. Following full text screen 114 were excluded leaving 13. Three of these papers screened and diagnosed participants but did not display rates of diagnosis. Instead, they reported the percentage of those screened as positive and negative and the percentage of those diagnosed. As the authors were unable to provide the required information or did not respond, these papers were also eliminated. Following elimination, ten papers remained for inclusion. Of these ten studies, two (Gray, Edwards, O'Callaghan, & Gibbons, 2015; Pritchard et al., 2016) did not include sufficient information to create 2x2 contingency tables as they only assessed those who screened positive on screening tests, meaning the rate of TN and FN are unknown. This left eight papers for meta-analytic synthesis.

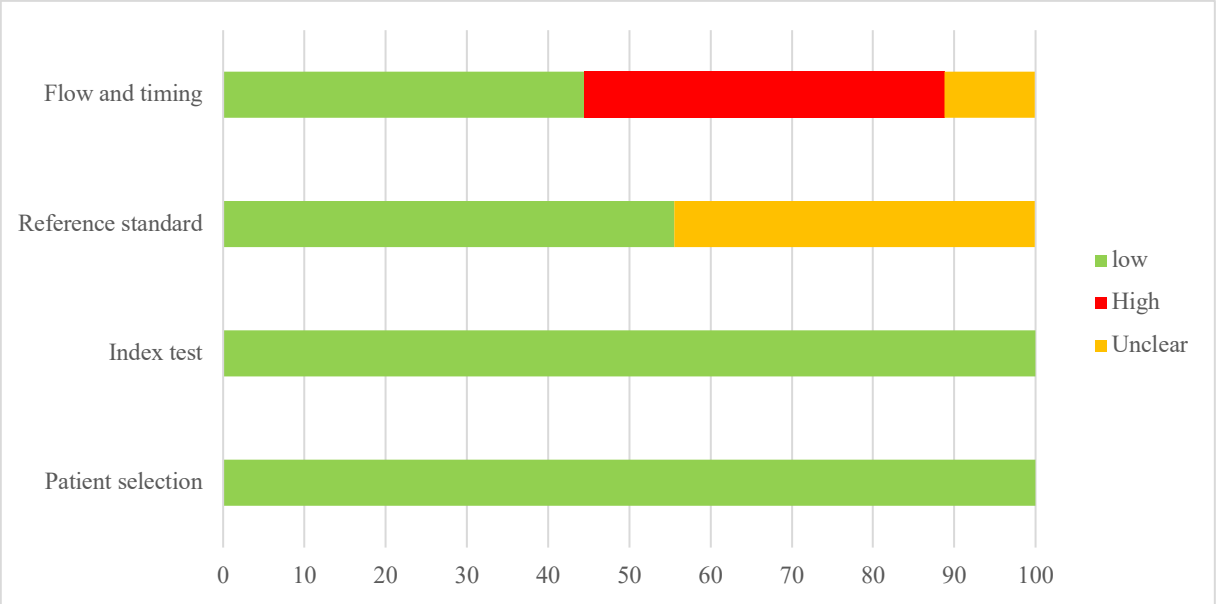
Characteristics of included studies

Eight studies were eligible for complete meta-analysis. One of these articles examined two screening tools (Verhaeghe et al., 2016), the Social Responsiveness Scale (SRS; Constantino et al., 2003) and the Social Communication Questionnaire (SCQ; Bailey & Lord, 2003). Data from both screening tools were included in this meta-analysis. Two of the included papers used the same sample pool and different screening tools (Beranova et al., 2017; Dudova et al., 2014). The meta-analytic sample consisted of 1896 participants. Based

on study reports, 23% of unique participants were from Europe, 5% from Asia and 72% from North America. Participants' gestational age ranged from 24-34 weeks, and age at testing ranged from 2-15 years. Of the included papers, eight used level 1 screening tools, and one used a level 2 screening tool. As described above, level 1 screening tools are developed for use within the general population regardless of level of risk for any developmental conditions (Towle & Patrick, 2016). Level 2 screening tools are developed for use in at risk populations and so aim to distinguish between developmental conditions (Towle & Patrick, 2016).

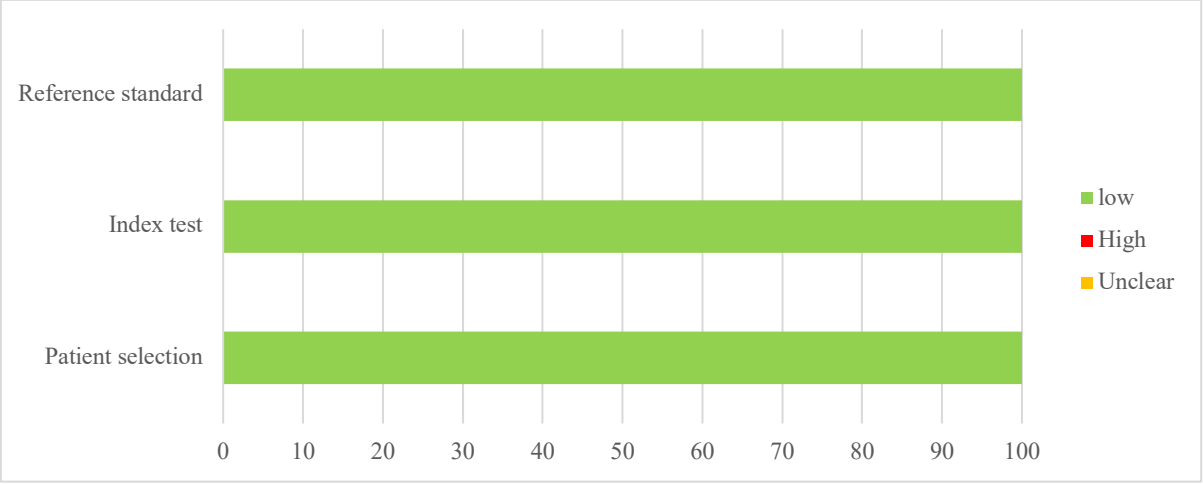
Methodological quality of included studies

Using the QUADAS-2 (Whiting et al., 2011), the overall bias within studies was rated as low, however some risk of bias was identified. Due to poor reporting of the reference standards used, the risk of bias within five studies was unclear. Risk of bias due to the flow and timing of screening tools and diagnostic assessments was rated as high in four studies and unclear in one study. All studies were rated as low risk of bias for the index test and patient selection, and low for concerns regarding the applicability of the reference test, index test and patient selection. Study level quality criteria were shown in Table 1.1.4, quality criteria ratings are included in Figures 1.1.2 and 1.1.3.



Note. Risk of bias is shown for all studies included in the meta-analysis. Ratings were allocated using guidelines from the QUADAS-2 Tool (Whiting, et al., 2011)

Figure 1.1.2. Proportion of studies with low high and unclear risk of bias (%)



Note. Risk of concerns regarding applicability are shown for all studies included in the meta-analysis. Ratings were allocated using guidelines from the QUADAS-2 Tool (Whiting, et al., 2011)

Figure 1.1.3. Proportion of studies with low, high and unclear concerns regarding applicability (%)

Univariate analysis of sensitivities and specificities

Sensitivity and specificity were pooled using the 'Mada' package (Doebler & Holling, 2015) of the R language (R Core Team, 2018). Individual study and pooled sensitivity and specificity are displayed in Figure 1.1.4 and 1.1.5 The pooled sensitivity was 0.71 (95% CI = 0.60 - 0.83) and the pooled specificity was 0.82 (95% CI = 0.76 – 0.88). The confidence intervals of all studies fell within the expected range based on the confidence intervals of pooled sensitivity (CI = 0.60-0.83) and specificity (CI = 0.76-0.88).

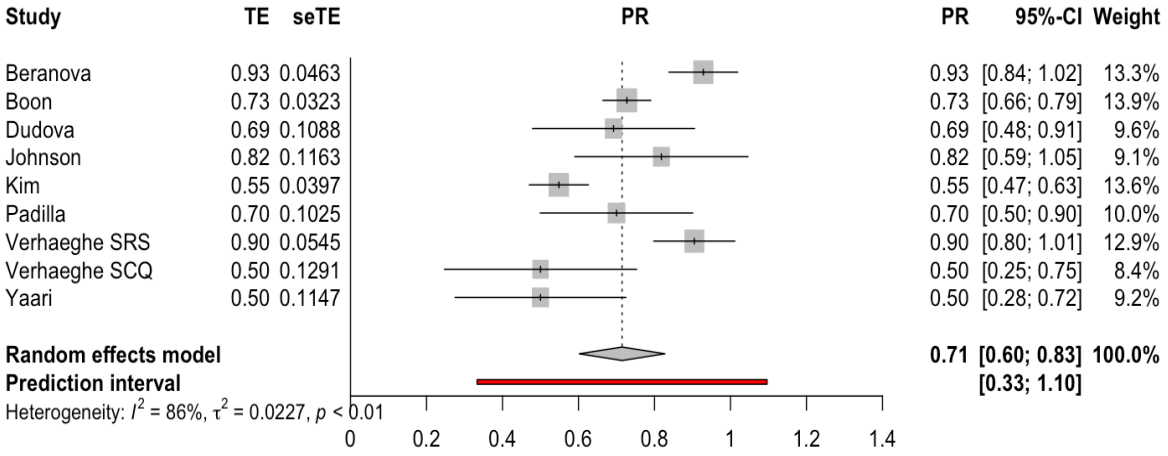


Figure 1.1.4. Forest plot on study level and pooled sensitivities

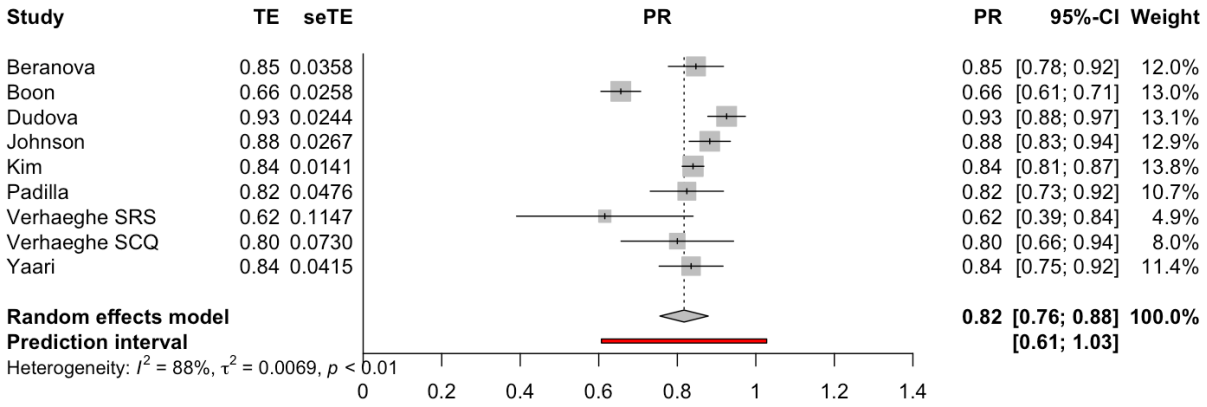


Figure 1.1.5. Forest plot on study level and pooled specificities

Random effect estimation using the DerSimonian-Laird estimator and Quantile-Quantile plots. We explored the heterogeneity of data in order to identify studies not appropriate for inclusion into the meta-analytic synthesis. Normality of data distribution was measured using DerSimonian-Laird estimated and visual inspection of Quantile-Quantile (QQ) plots. A high level of heterogeneity was observed for sensitivities (I^2 = 86.1, tau^2 = 0.02*27*, p < 0.01) and specificities (I^2 = 88.4, tau^2= 0.0069, p < 0.01), with sensitivities ranging between 0.50 and 0.93, and specificities ranging between 0.62 and 0.93. QQ plots compare the observed distribution to the expected distribution (normal distribution) using both random and fixed effect models. QQ plots for sensitivity and specificity are shown in Figures 1.1.6 and 1.1.7. Visual inspection suggests that the data for both sensitivity and specificity are normally distributed using the random effects model. Therefore, the random effects model was deemed most appropriate.

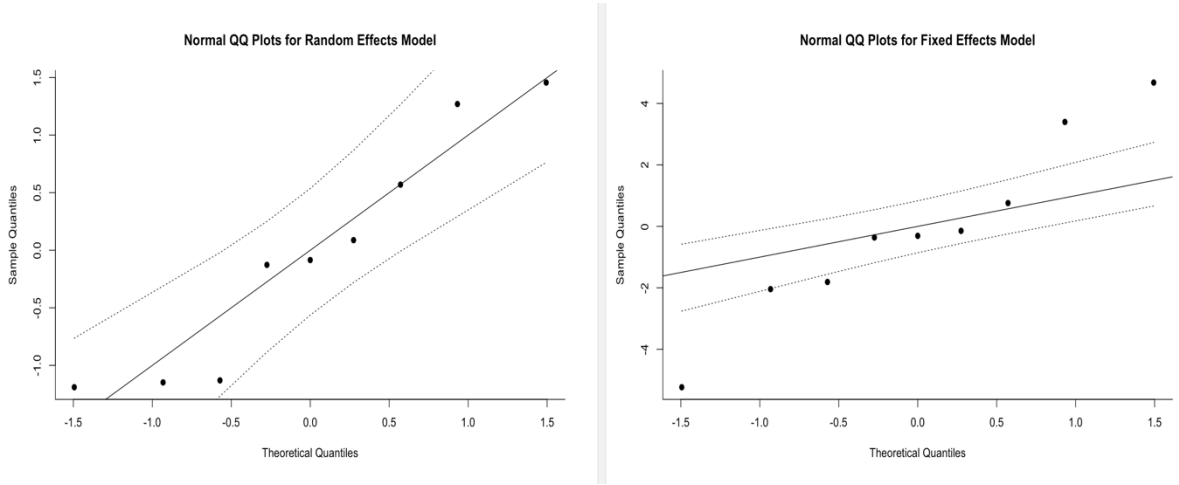


Figure 1.1.6. Quantile-Quantile plot using both the random effects and fixed model of the distribution of study level sensitivity

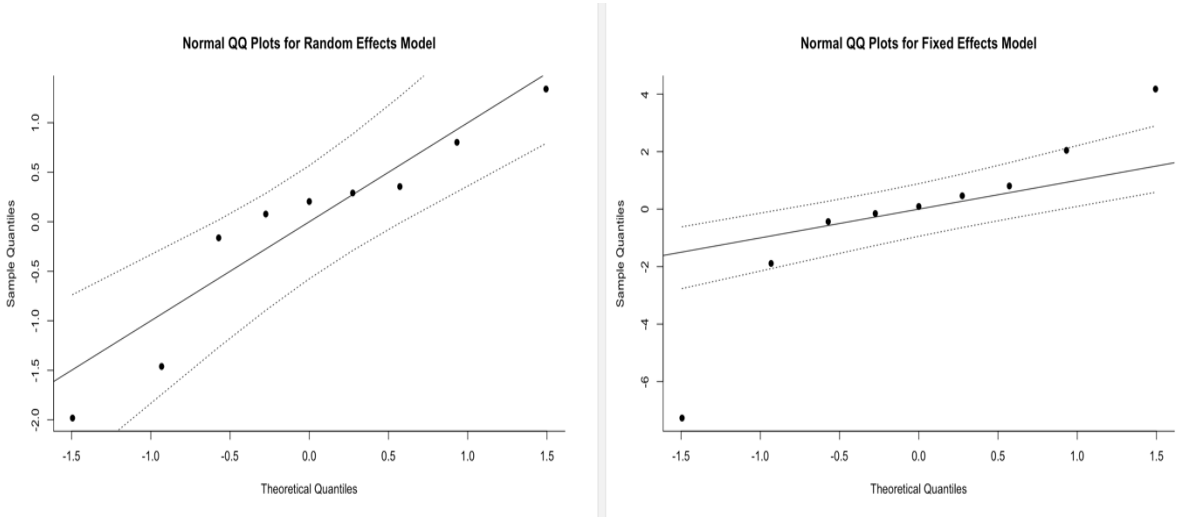


Figure 1.1.7. Quantile-Quantile plot using both the random effects and fixed model of the distribution of study level specificity

The impact of influential studies. As a high level of heterogeneity was observed, a "leave-one-out" analysis was conducted in order to measure the impact of disproportionately influential studies. Here the random effects model was calculated with each study removed in turn for sensitivity and specificity. Figure 1.1.8 shows a forest plot of leave-one out analysis for sensitivity. If the 95% confidence interval for a removed study does not include a value contained within the confidence interval of the pooled sensitivity, it may be inferred that its removal results in a quantitatively different outcome, as the study is exerting excessive influence. Figure 1.1.8 shows that no individual study is exerting excessive influence on pooled sensitivity. Equally, figure 1.1.9 shows that no individual study is exerting excessive influence on pooled specificity.

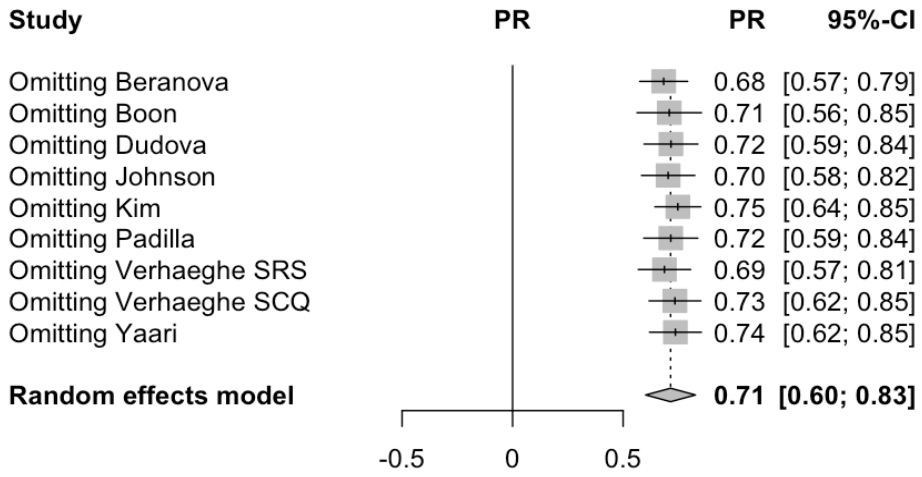


Figure 1.1.8. Leave-one-out analysis for sensitivity (the effect of omitting each study on the pooled sensitivity)

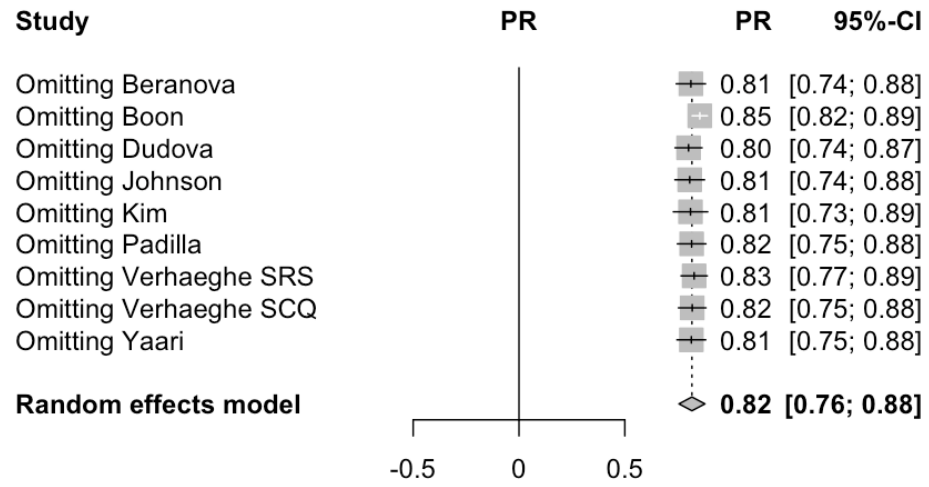


Figure 1.1.9. Leave-one-out analysis for specificity (the effect of omitting each study on the pooled specificity)

Publication bias. Publication bias is assessed using funnel plots as visual aids for detecting systematic heterogeneity. Each funnel plot presents a scatterplot of individual study sensitivities and specificities against a measure of study precision. If publication bias is present the funnel plot will appear asymmetrical. If publication bias is absent the funnel plot will be more evenly distributed around the average. A trim and fill procedure (Duval & Tweedie, 2000) uses an iterative algorithm to remove extreme small studies causing asymmetry, re-computing sensitivities and specificities at each iteration until the funnel plot is symmetric. In theory, this yields an unbiased estimate of sensitivity and specificity. Removed studies are then reintroduced and a mirror images imputed. Figure 1.10 shows the corrected funnel plot for sensitivities. Using the trim and fill procedure two studies were estimated to be added to the bottom right of the funnel plot, demonstrating that there may be a publication bias towards studies reporting low sensitivities within studies. The uncorrected sensitivity was 0.71 (95% CI = 0.60-0.83), and the corrected sensitivity was 0.76 (95 % CI = 0.65-0.87). Based on these confidence intervals these uncorrected and corrected sensitivities

are comparable. Figure 1.1.11 shows the corrected funnel plot for specificities. The trim and fill procedure did not impute any further studies, showing that the data conform to normal expectations and there is no evidence of substantial publication bias. The uncorrected specificity was 0.82 (CI = 0.75-0.88, and the corrected specificity was 0.82 (CI = 0.76-0.88).

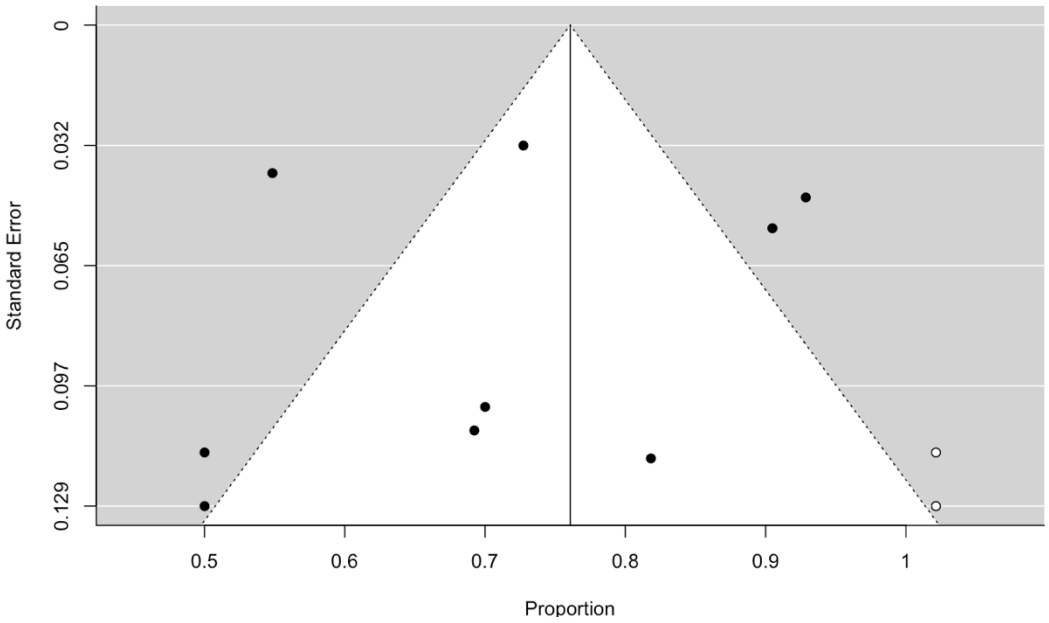


Figure 1.1.10. Corrected funnel plot of study level sensitivity associated with publication bias using the trim and fill procedure (Duval &Tweedie, 2000)

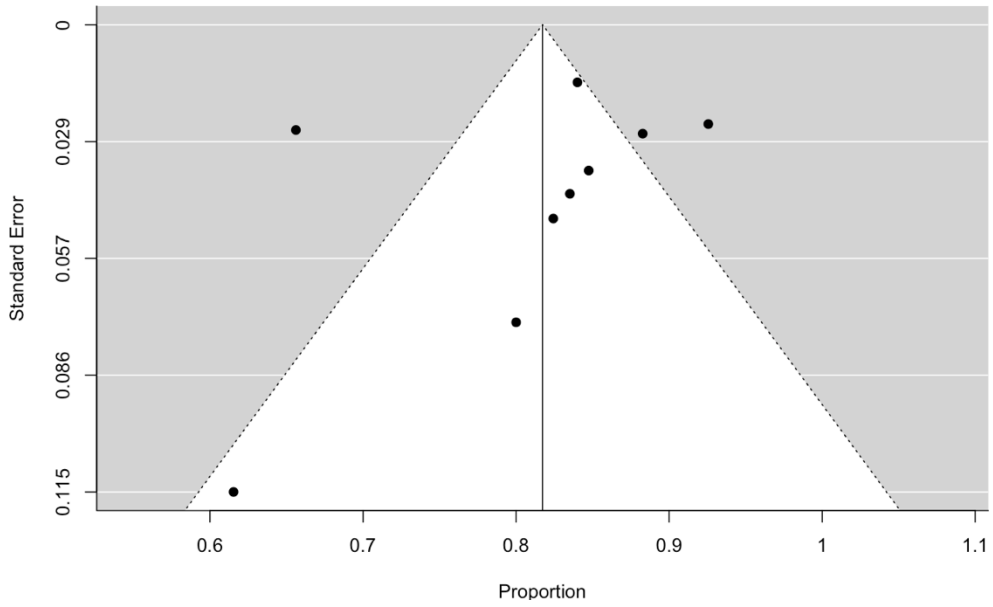


Figure 1.1.11. Corrected funnel plot of study level specificity associated with publication bias using the trim and fill procedure (Duval & Tweedie, 2000)

Subgroup analysis. Across the eight studies that reported diagnostic accuracy, six screening tools were evaluated: five level 1 screens; Social Communication Questionnaire, Social Responsiveness Scale (SRS), Modified Checklist for Autism in Toddler (M-CHAT; Robins, Fein, Barton, & Green, 2001), Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP; Wetherby & Prizant, 2002) and Autism Observation Scale for Infants (AOS1; Bryson, Zwaigenbaum, McDermott, Rombough, & Brian, 2008) and one level 2 screen; Pervasive Developmental Disorders Screening test-II Stage 2-Developmental Clinic Screener (PDDST-II-DCS; Siegel, 2004). Subgroup analysis of studies by screening tool was conducted. Figure 1.1.12 shows pooled sensitivities for each screening tool. The AOSI produced the poorest sensitivity followed by the MCHAT. The highest sensitivity was produced by CSBSD. A significant difference between subgroups based on screening tools was found ($p < 0.01$). Figure 1.1.13 shows pooled specificities for each screening tool and

each level of screening tools. The PDDST-II-DCS produced the lowest specificity followed by the SRS. All other screening tools produced comparable specificities. A significant difference between subgroups based on screening tools was found ($p < 0.01$).

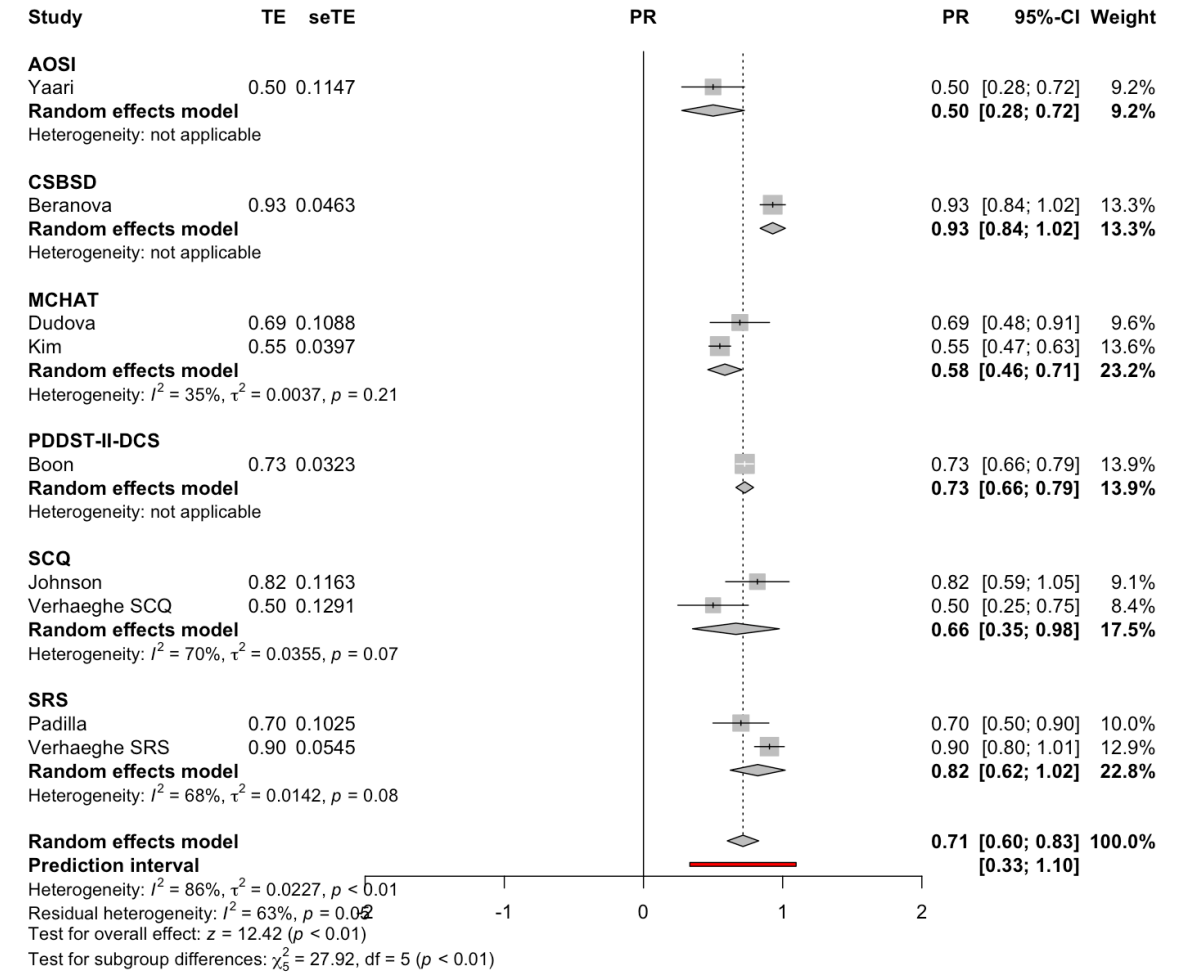


Figure 1.1.12. Subgroup analysis of sensitivities across studies by screening tool used

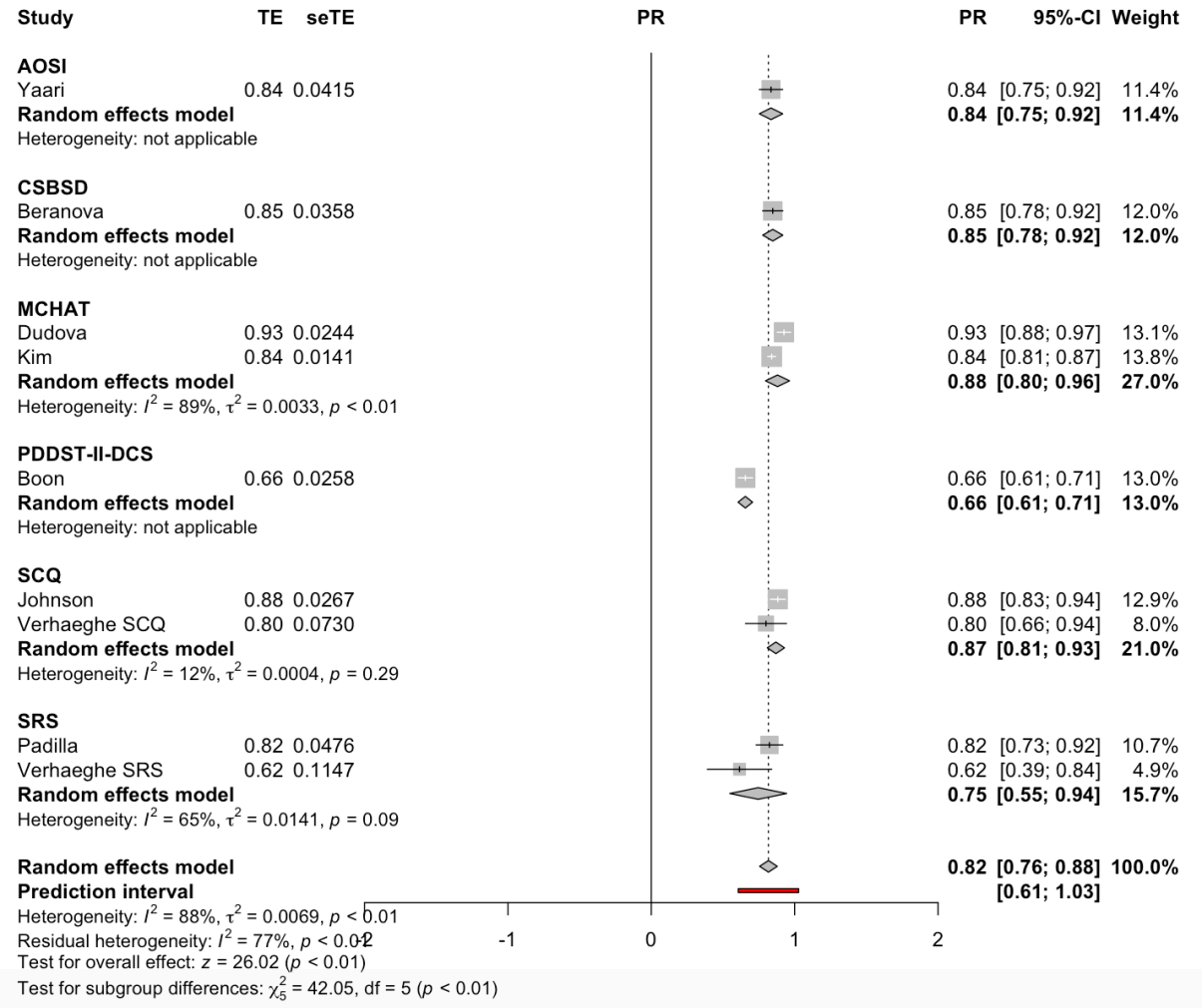


Figure 1.1.13. Subgroup analysis of specificity across studies by screening tool used

Bivariate indicators of screening tool accuracy

The second aim of this meta-analysis was to synthesise existing empirical data from appropriate studies to produce pooled estimates of sensitivity, specificity, positive and negative likelihood rations and a DOR from the bivariate model.

Univariate analysis of study level sensitivities and specificities indicated that all studies were appropriate for meta-analytic synthesis. Bivariate test indicators provide a more

accurate estimate of psychometric properties than univariate test indicators by considering each property with its counterpart (i.e. considering sensitivity alongside specificity). In order to calculate bivariate measures a summary receiver operating curve (SROC) must first be computed. Figure 1.1.14 displays the results of individual studies in ROC space (the summary ROC (SROC).

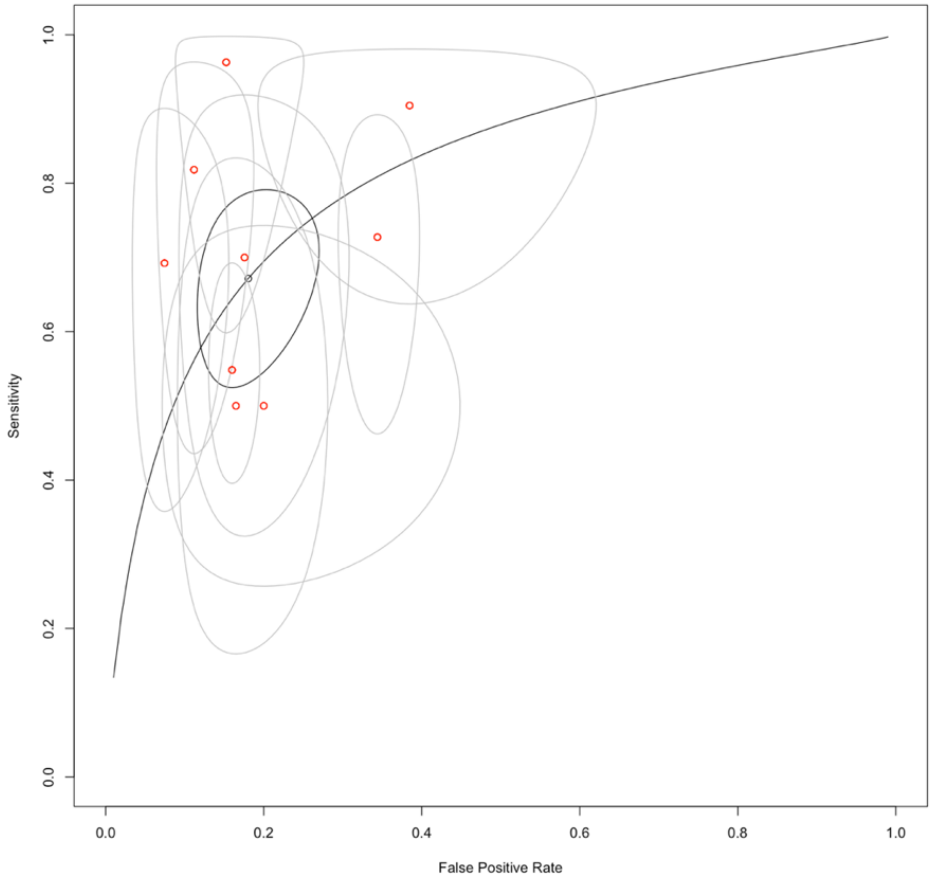


Figure 1.1.14. Summary Receiver Operating Curve for all screening tools

Each study is plotted as a point estimate, showing sensitivity (true-positive rate) across specificity (false-positive rate). The AUC provides a single summary statistic from the sROC curve and is equal to the probability that if an autistic and non-autistic child were selected at random, the autistic child would have a higher score on screening tools compared

to the non-autistic child. Meta-analysed studies indicated an AUC value of 0.808 (extrapolated across all possible false positive rates) and a partial AUC of 0.694 (restricted to the observed false positive rates). Macaskill et al, (2010) recommend the use of partial AUCs in meta-analyses of diagnostic accuracy.

According to Jones and Athanasiou (2005), a random test (allocating positive results 50% of the time) would produce an AUC of 0.5 and a perfect test would theoretically produce an AUC of 1. Jones and Athanasiou (2005) continue to recommend that an acceptable test would produce an AUC = 0.7 – 0.8, and an excellent test an AUC of 0.8 - 0.9). Using these criteria, overall screening tools demonstrated just below acceptable accuracy (0.694; Mandrekar, 2010). The sROC curve can be used to generate bivariate measures of sensitivity, specificity, positive and negative likelihood ratios and DOR that provide a more accurate estimate than univariate analysis (Reitsma et al., 2005). This method of calculating psychometric properties in recommended when there is no standard cut-off score that defines a positive (The Joanna Briggs Institute, 2015), and has been used in meta-analyses comparing the diagnostic accuracy of different screening tools (Van der Lee et al., 2015; Wong et al., 2016; Sanchez-Garcia et al., 2019). All bivariate psychometric properties are displayed in Table 1.1.6.

Table 1.1.6

Summary of bivariate meta-analytic psychometric properties of screening tools in preterm population

Se	95% CI	Sp	95% CI	LR+	95% CI	LR-	95% CI	DOR	95% CI
0.67	[0.56 – 0.77]	0.82	[0.74 – 0.87]	3.74	[2.67 – 5.15]	0.45	[0.29 – 0.54]	9.58	[5.35 – 15.90]

Note: Se sensitivity, CI confidence interval, LR+ positive likelihood ratio, LR- negative likelihood ratio, DOR diagnostic odds ratio.

The pooled sensitivity was .67 (95% CI = 0.56 - 0.77) and the pooled specificity was .82 (95% CI = 0.74 – 0.87). The pooled positive likelihood ratio was 3.74 (95% CI = 2.67 – 5.15) and the negative likelihood ratio was .45 (95% CI = 0.29-0.54). The DOR was 9.58 (95% CI = 5.35 – 15.90). This DOR suggests that the odds of a positive screen in an autistic child is 9.58 times higher than the odds of positive screen in a non-autistic child.

Comparisons of results in preterm children and the general population.

The third aim of this meta-analysis was to compare the synthesised psychometric properties to those obtained by Sanchez-Garcia et al. (2019). This allows comparison of the clinical utility of screening tools to be made between preterm and the general population children. A summary of pooled psychometric properties from both meta-analyses (using bivariate models) are outlined in Table 1.1.7.

Table 1.1.7

Summary of bivariate meta-analytic psychometric properties of screening tools in preterm and the general population

Meta-analysis	Se	95% CI	Sp	95% CI	LR+	95% CI	LR-	95% CI	DOR	95% CI
General	0.72	[0.61 - 0.81]	0.98	[0.97 - 0.99]	131.27	[50.40 - 344.48]	0.22	[0.13 - 0.45]	596.09	[174.32 - 2038.34]
Preterm	0.67	[0.56 – 0.77]	0.82	[0.74 – 0.87	3.74	[2.67 – 5.15]	0.45	[0.29 – 0.54]	9.58	[5.35 – 15.90]

Note. General refers to the meta-analysis published by Sanchez-Garcia et al (2019). Preterm refers to results synthesised in this meta-analysis. Se sensitivity, CI confidence interval, LR+ positive likelihood ratio, LR- negative likelihood ratio, DOR diagnostic odds ratio.

Although sensitivity is lower in preterm populations (preterm population = 0.67; 95% CI = 0.56 - 0.81; general = 0.71; 95% CI = 0.60 - 0.83), it does fall within the confidence interval of the general population sample. This demonstrates that sensitivity is comparable across populations. However, specificity is considerably lower in the preterm population (0.82; 95% CI = 0.74 - 0.87), compared to the general population (0.99; 95% CI = 0.97 - 0.99), with no overlap of the preterm specificity estimate and general population confidence interval. This demonstrates that specificity is much lower in preterm populations. In the general population the positive likelihood ratio estimate is high (131.27; 95% CI= 50.40 - 344.48) indicating excellent accuracy in detecting autism in autistic compared to non-autistic individuals. In the preterm population the positive likelihood ratio estimate is closer to zero (3.74; 95% CI= 2.67 - 5.15), indicating low accuracy in this area. Furthermore, the preterm positive likelihood ratio estimate of 3.74 does not fall within the general population confidence interval (50.40 - 344.58), indicating a large difference across populations. The

lower the negative likelihood ratio, the more accurate tests are in identifying those without autism compared to those with autism. Therefore, screening tools were twice as accurate in ruling out those without autism in the general population (0.22; 95%; CI = 0.13 - 0.45) compared to the preterm population (0.45; 95% CI = 0.29 - 0.54). However, the preterm negative likelihood ratio estimate (0.45) does fall within the general population confidence interval (0.13 - 0.45), and so this should be interpreted with caution.

Overall, Sanchez-Garcia et al. (2019) demonstrated that within the general population, screening tools are highly accurate in detecting autism (DOR = 596.09; 95% CI = 174.32 - 2038.34). Specifically, the odds of obtaining a positive screen result when autism is present relative to the odds in those without autism are high. In comparison, the DOR in the preterm population was low (9.58; 95% CI = 5.35 – 15.90), showing poorer odds of obtaining a positive screen in those born preterm when autism is present relative to the odds in those without autism.

Discussion

Over the past decade there has been growing interest in the diagnostic utility of autism-specific screening tools when applied to preterm populations. Meta-analysis allows for available data on the diagnostic utility of such tools to be synthesized and interpreted for both research and clinical practice. Sanchez-Garcia et al. (2019) meta-analysed the diagnostic utility of screening tools in the general population. However due to the neurodevelopmental complexities that occur at high rates in preterm population, their results cannot reliably be extrapolated to preterm populations. This meta-analysis aimed to answer a clinically relevant question: How informative are screening tools when assessing for autism in preterm children in comparison to children within the general population?

Overall diagnostic utility in preterm children

The review included eight studies examining the accuracy of six screening tools for detecting autism in preterm populations. Glascoe. (2005) suggests that reasonable screening tools should demonstrate a sensitivity of at least 70-80% and specificity of around 80%. The results of this meta-analysis demonstrate that the sensitivity was marginally below this recommendation (0.67) and specificity was at the suggested level (0.82). The AUC was 0.694, just below the recommended cut-off for acceptable accuracy (Jones and Athanasiou, 2005). As the results show acceptable specificity, and marginally below acceptable sensitivity and accuracy, autism-specific screening tools demonstrate limited validity in preterm populations. Whilst provided informative data, caution should be taken when using such tests in clinical practice.

Diagnostic utility in preterm children compared to the general population

When applied to the general population (compared to the preterm population), autism screening tools demonstrate better psychometric properties, and so are more informative when identifying those who require diagnostic autism assessments, than when applied to preterm populations. Although the difference in sensitivities and negative likelihood ratios were small, better accuracy in the general population was demonstrated for specificity, positive likelihood ratio, and DOR. These findings are consistent with previous findings that psychometric properties of screening tools are less accurate in preterm populations (Simard, Luu, & Gosselin, 2012). Although reasonably accurate in preterm children, screening tools are more accurate in detecting autism in the general population.

Exploration of difference in psychometric properties

The poorer diagnostic accuracy of screening tools in preterm populations may be due to the neurodevelopmental complexity of this population. As stated previously, preterm children display variations in phonological fluency, verbal IQ, visual-motor integration, executive functioning and somatosensory registration, as well as atypical empathic development and social reasoning abilities (Broring et al., 2008; Campbell et al., 2005; Guarini et al., 2019; Lejeune et al., 2016; Luu, Vohr, Allan, Schneider, & Ment, 2011). Atypical development in any of these areas may result in such children scoring highly on autism screening tools, with specific difficulties arising from limited capabilities of autism-specific screening tools to discern between those with and without autism in preterm populations.

Base rates of autism are higher in preterm populations compared to the general population. Indeed, studies included in this paper had prevalence rates of 4 - 45%, compared to approximately 1% in Sanchez-Garcia et al's (2019) sample. Whilst base rates impact on sensitivity, specificity, and likelihood ratios, they do not impact on DOR. Therefore, screening test performance is poorer in preterm populations despite high base rates. A possible variable impacting false screen rates may be parental over or under-reporting. However parental reports of developmental difficulties have proven to be valid in preterm populations (Johnson et al., 2004; Perra et al., 2015).

Differences in psychometric properties between screens

We explored differences in sensitivity and specificity between screening tools used. However, due to limited quantity of research examining the diagnostic accuracy of autism-

specific screening tools in preterm population, each subgroup consisted only of one or two studies.

Glascoe (2005) suggests that reasonable screening tools should demonstrate a sensitivity of at least 70-80% and specificity of around 80%. Subgroup analysis of sensitivities and specificities between screening tools revealed that some screens fall short of this. The AOSI, M-CHAT, and SCQ had sensitivities below 0.67. The PDDST-II—DCS and the SRS had specificities below 0.76. Therefore, only the CSBSD has psychometric properties within a reasonable range. In comparison with included screens, the CSBSD demonstrated good psychometric properties in all areas. However, this should be interpreted with caution as it is based on only one study. Furthermore, confidence intervals for sensitivities and specificities were large, and so only limited conclusions can be drawn from these data.

Methodological strengths and limitations

Any interpretations of the data should be made in the context of the study strengths and limitations. We used methods comparable to those used by Sanchez-Garcia et al. (2019), to enable comparisons of findings between meta-analyses. The similarities in methodology allow for direct comparison between preterm groups and the general population. Sanchez-Garcia et al. (2019) employed the HSROC model to "estimate screening parameters and a summary in each study as functions of an underlying bivariate model" (p. 1847). We used the bivariate random effects model using the procedures described by (Reitsma et al., 2005). This model has been shown by Harbord and Whiting (2009) to be equivalent to the HSROC model used by Sanchez-Garcia et al. (2019). Although the sample was small, we believe that the search strategy was robust in identifying all relevant papers. Furthermore, the overall quality of studies included was observed to be good, by two independent raters. The AUC

demonstrated just below acceptable diagnostic accuracy of screening tools in preterm populations. These results must be interpreted with caution. Jones and Athanasiou (2005) recommend that ten or more data points are needed to produce a normal distribution. With only eight studies, this meta-analysis falls short of this guideline. Furthermore, such results may only be meaningful when examining results from multiple studies using a single screening tool, as opposed to this meta-analysis exploring multiple tools. Unfortunately, due to the small amount of literature in this area, larger data synthesis was not possible. Despite the small sample size analysed, included studies were all published within the last ten years, increasing the likelihood that the meta-analysed sample are representative of the current preterm population. The number of studies included within this meta-analysis was comparable to the meta-analysis by Sanchez-Garcia et al. (2019, nine data sets in each meta-analysis). The overall quality of studies included was rated higher in all areas, except risk of bias for flow and timing and reference standard. Unfortunately, subgroup analysis between screening tools was limited to one to two studies per subgroup, making any conclusions drawn limited.

Another important factor is that not all studies extracted were able to be included in the meta-analytic synthesis due to missing data (Gray, Edwards, O'Callaghan, & Gibbons, 2015; Pritchard et al., 2016). These studies would have contributed 20% of the study sample. It is recommended that researchers follow up all participants included in their studies regardless of the outcome of the screening tool. This will ensure the rates of false and true negatives can be quantified and included in future meta-analytic synthesis.

Clinical implications

Due to the high prevalence rate of autism in children born preterm (Agrawal et al., 2018), positive results on autism-specific screening tools warrant further investigation.

However, when considering diagnosing autism, considerations should be made. Screening tools in preterm children show poorer accuracy compared to the general population in discriminating between autistic and non-autistic children, with a very low DOR (9.73 95% CI: 5.55 – 17.03), compared to in the general population. This may be due to the neurodevelopmental complexity within this population. However, there is also evidence that autistic children born preterm may display different clinical expressions than their full-term peers. Termed the "preterm behavioural phenotype", this clinical expression may be primarily based on socialisation difficulties (Bowers, Wink, Pottenger, McDougle, & Erickson, 2015; Johnson, et al., 2004), potentially resulting in atypical profiles of strengths and weaknesses in autism-specific screening and diagnostic tools. Due to high rates of social-communication difficulties in preterm populations, it is unclear how accurately screening and diagnostic tools are able to discern whether such difficulties are indicative of autism or not. In preterm children, other neurodevelopmental sequelae such as this should be considered, in order to determine if there are any overlap of autistic traits, that may be similar to but not indicative of autism.

Clinicians should remain open about the possibility of false screen results and be mindful of the distress a false positive may cause for families. Tu et al. (2012) suggest that clinicians can help to reduce this stress by communicating with parents the possibility of a false positive screen. Therefore, it is important that when assessing children born preterm for autism, clinicians use preterm specific estimates such as those reported within this meta-analysis as opposed to estimates for the general population. Overall, more caution should be taken when interpreting screen results of children born preterm. In line with NICE guidelines, enhanced developmental surveillance may be necessary to monitor the trajectory of development over time.

Conclusion

This meta-analysis examined the diagnostic accuracy of autism-specific screening tools in preterm populations. The overall diagnostic accuracy of autism-specific screening tools is poorer in preterm populations compared to the general population. However, they are still have some validity in providing informative data to clinicians when discerning whether or not a child should receive a full diagnostic assessment for autism. Clinicians should interpret any information obtained from screening tools with caution together with information regarding the neurodevelopmental sequelae observed in those born preterm. Given the neurodevelopmental complexity of preterm children, accurate screening tools are essential in supporting clinicians in understanding each child's developmental profile. This will facilitate in the early identification of children with the highest needs.

References

Agrawal, S., Rao, S. C., Bulsara, M. K., & Patole, S. K. (2018). Prevalence of Autism Spectrum Disorder in Preterm Infants: A Meta-analysis. *Pediatrics, 142*(3), 1-15.

Autistica. (unknown). *Media communications guide*. Retrieved from www.autistica.org.uk: https://www.autistica.org.uk/about-us/media-communications-guide

Bailey, A. L., & Lord, C. (2003). *The Social Communication Questionnaire.* Los Angeles, CA: Western Psychological Services.

Beranova, S., Stoklasa, J., Dudova, I., Markova, D., Kasparova, M., Zemankova, J., . . . Hrdlicka, M. (2017). A possible role of the infant/toddler sensory profile in screening for autism: A proof-of-concept study in the specific sample of prematurely born children with birth weights <1,500 g. *Neuropsychiatric Disease and Treatment, 13*, 191-200.

Bowers, K., Wink, L. K., Pottenger, A., McDougle, C. J., & Erickson, C. (2015). Phenotypic differences in individuals with autism spectrum disorder born preterm and at term gestation. *Autism, 19*(6), 758-763.

Broring, T., Konigs, M., Oostrom, K. J., Lafeber, H. N., Brugman, A., & Oosteriaan, J. (2008). Sensory processing difficulties in school-age children born very preterm: An exploratory study. *Early Human Development, 117*, 22-31.

Broring, T., Oostrom, J., Van Dijk-Lokkart, E., Lafebar, H. N., Brugman, A. K., & Oosterlaan, J. (2018). Attention deficit hyperactivity disorder and autism spectrum disorder symptoms in school-age children born very preterm. *Researchn in Developmental Disabilities, 74*, 103-112.

Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., & Brian, J. (2008). The autism observation scale for infants: scale development and reliability data. *Journal of Autism and Developmental Disorders, 38*, 731-738.

Campbell, C., Horlin, C., Reid, C., McMichael, J., Forrest, L., Brydges, C., . . . Anderson, M. (2015). How do you think she feels? Vulnerability in empathy and the role of attention in school-aged children born extremely preterm. *British Journal of Developmental Psychology, 33*, 312-323.

CDC, C. f. (2014, March 27). *CDC estimates 1 in 68 children has been identified with autism spectrum disorder: Latest snapshot shows proportion of children with autism and higher IQ on the rise.* Retrieved from cdc.gov: https://www.cdc.gov/media/releases/2014/p0327-autism-spectrum-disorder.html

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., . . . Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders, 33*, 427-433.

Deeks, J. J. (2001). Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal, 323*, 157-162.

Doebler, P., & Holling, H. (2015). *Meta-analysis of diagnostic accuracy with mada.* Retrieved from https://Cran.Rproject.Org/Web/Packages/Mada/Vignettes/Mada.

Dudova, I., Kasparova, M., Markova, D., Zemankova, J., Beranova, S., Urbanek, T., & Hrdlicka, M. (2014). Screening for autism in preterm children with extremely low and very low birth weight. *Neuropsychiatric Disease and Treatment, 10*, 277-282.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463.

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. (2003). The diagnostic odds ration: A single indicator of test performance. *Journal of Clinical of Epidemiology, 56*(11), 1129-1135.

Glascoe, F. P. (2005). Screening for developmental and behavioural problems. *Mental Retardation and Developmental Disabilities Research Reviews, 11*, 173-179.

Gray, P. H., Edwards, D. M., O'Callaghan, M. J., & Gibbons, K. (2015). Screening for autism spectrum disorder in very preterm infants during early childhood. *Early Human Development, 91*(4), 271-276.

Guarini, A., Bonifacci, P., Tobia, V., Alessandroni, R., Faldella, G., & Sansavini, A. (2019). The profile of very preterm children on academic achievement. A cross-population comparison with children with specific learning disorders. *Research in Developmental Disabilities, 87*, 54-63.

Gurian, E. A., Kinnamon, D. D., Henry, J. J., & Waisbren, S. E. (2006). Expanded newborn screening for biochemical disorders: the effect of a false-positive result. *Pediatrics, 117*(6), 1915-1921.

Guy, A., Seaton, S. E., Boyle, E. M., Draper, E. S., Field, D. J., Manktelow, B. N., . . . Johnson, S. (2015). Infants born late/moderately preterm are at increased risk for a positive autism screen at 2 years of age. *The Journal of Pediatrics, 166*(2), 269-275.

Harbord, R. M., & Whiting, P. (2009). Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata Journal, 9*(2), 211-229.

Himmelmann, K., Hagberg, G., & Uvebrant, P. (2014). The panorama of cerebral palsy in Sweden. XI. Changing patterns in the birth-year period 2003–2006. *Acta Paediatrica , 103*(6), 618-624.

Hirvonen, M., Ojala, R., Korhonen, P., Haataja, P., Eriksson, K., Gissler, M., . . . Tammela, O. (2018). Visual and hearing impairments after preterm birth. *Pediatrics, 142*(2).

Hirvonen, M., Ojala, R., Korhonen, P., Haataja, P., Eriksson, K., Rantanen, K., . . . Tammela, O. (2017). Intellectual disability in children aged less than seven years born moderately and late preterm compared with very preterm and term-born children – a nationwide birth cohort study. *Journal of Intellectual Disability Research, 61*(11), 1034-1054.

Ibanez, L. V., Stone, W. L., & Coonrod, E. E. (2014). Screening for autism in young children. In F. R. Volkmar, *Handbook of Autism and Pervasive Developmental Disorders, Assessment, Interventions, Policy, the Future : Assessment, Interventions, and Policy* (pp. 585-608). New Jersey: John Wiley & Sons, Incorporated.

Johnson, S., & Marlow, N. (2011). Preterm birth and childhood psychiatric disorders. *Pediatric Research, 69*(8), 11-18.

Johnson, S., Hollis, C., Hennessy, E., Kochhar, P., Wolke, D., & Marlow, N. (2011). Screening for autism in preterm children: Diagnostic utility of the social communication questionnaire. *Archives of Disease in Childhood, 96*(1), 73-77.

Johnson, S., Marlow, N., Wolke, D. D., Davidson, L., Marston, L., O'Hare, A., . . . Schulte, J. (2004). Validation of a parent report measure of cognitive development in very preterm infants. *Developmental Medicine and Child Neurology, 46*(6), 389-397.

Jones, C. M., & Athanasiou, T. (2005). Summary Receiver Operating Characteristic Curve Analysis Techniques in the Evaluation of Diagnostic Tests. *The annals of thoratic surgery, 79*(1), 16-20.

Kim, S. H., Joseph, R. M., Frazier, J. A., O'Shea, T. M., Chawarska, K., Allred, E. N., . . . Kuban, K. K. (2016). Predictive validity of the modified checklist for autism in toddlers (M-CHAT) born very preterm. *Journal of Pediatrics, 178*, 101-107.

Leeflang, M. M., Rutjes, A. W., Reitsma, J. B., Hooft, L., & Bossuyt, P. M. (2013). Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ, 185*(11), 537-544.

Lejeune, F., Reveillon, M., Monnier, M., Huppi, P. S., Tolsa, C. B., & Barisnikov, K. (2016). Social reasoning abilities in preterm and full-term children aged 5–7 years. *Early Human Development, 103*, 49-54.

Luu, T. M., Vohr, B. R., Allan, W., Schneider, K. C., & Ment, L. R. (2011). Evidence for Catch-up in Cognition and Receptive Vocabulary Among Adolescents Born Very Preterm. *Pediatrics, 128*(2), 313-324.

Macaskill, P., Gatsonis, C., Deeks, J. J., Harbord, R. M., & Takwoingi, Y. (2010). Analysing the Presenting Results. In J. J. Deeks, P. M. Bossuyt, & C. Gatsonis, *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.* The Cochrane Collaboration.

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Manuck, T. A., Rice, M., Bailit, J. L., Grobman, W. A., Reddy, U. M., Wapner, R. J., . . . Tolosa, J. E. (2016). Preterm neonatal morbidity and mortality by gestational age: a contemporary cohort. *American Journal of Obstetrics and Gynecology, 215*(103-114).

Milne, S. L., McDonald, J. L., & Comino, E. J. (2013). Adaptive function in preschoolers in relation to developmental delay and diagnosis of autism spectrum disorders: Insights from a clinical sample. *Autism, 17*, 743-753.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *Annals of Internal Medicine, 151*(4), 264-269.

Padilla, N., Eklof, E., Martensson, G. E., Bolte, S., Lagercrantz, H., & Aden, U. (2017). Poor brain growth in extremely preterm neonates long before the onset of autism spectrum disorder symptoms. *Cerebral Cortex, 27*(2), 1245-1252.

Pascal, A., Govaert, P., Oostra, A., Naulaers, G., Ortibus, E., & Van den Broeck, C. (2018). Neurodevelopmental outcome in very preterm and very-low-birthweight infants born over the past decade: a meta-analytic review. *Developmental medicine and child neurology, 60*(4), 342-355.

Perra, O., McGowen, J. E., Grunau, R. E., Doran, J. B., Craig, S., Johnston, L., . . . Alderdice, A. (2015). Parent ratings of child cognition and language compared with Bayley-III in preterm 3-year-olds. *Early Human Development, 91*(3), 211-216.

Pritchard, M. A., De Dassel, T., Beller, E., Bogossian, F., Johnston, L., Paynter, J., . . . Scott, J. (2016). Autism in toddlers born very preterm. *Pediatrics, 137*(2).

Reitsma, J. B., Glas, A. S., Rutjes, A. W., Scholten, R. J., Bossuyt, P. M., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology, 58*(10), 982-990.

Reitsma, J. B., Rutjes, A. W., Whiting, P., Vlassov, V. V., Leefland, M. M., & Deeks, J. J. (2019). Assessing methodological quality. In J. Deeks, P. M. Bossuyt, & C. Gatsonis,

Cochrane Handbook for Systeatic Reviews of Diagnostic Test Accurcy version 1.0.0. The Cochrane Collaboration.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*, 131-144.

Rutter, C. M., & Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine, 20*(19), 2865-2884.

Sanchez-Garcia, A. B., Galindo-Villardon, P., Nieto-Librero, A. B., Martin-Rodero, H., & Robins, D. L. (2019). Toddler Screening for Autism Spectrum Disorder: A Meta-Analysis of Diagnostic Accuracy. *Journal of Autism and Developmental Disorders*.

Siegel, B. (2004). *Pervasive developmental disorder screening test-II (PDDST-II).* San Antonio, TX: Harcourt.

Simard, M. N., Luu, T. M., & Gosselin, J. (2012). Concurrent validity of ages and stages questionnaire in preterm infants. *Pediatrics, 130*(1), 108-114.

The Joanna Briggs Institute. (2015). *The Joanna Briggs Institute Reviewers' Manual.* Adelaide: The Joanna Briggs Institute.

Towle, P. O., & Patrick, P. A. (2016). Autism spectrum disorder screening instruments for very young children: A systematic review. *Autism research and treatment*, 1-29.

Tu, W. J., He, J., Chen, H., Shi, X. D., & Li, Y. (2012). Psychological effects of false-positive resullts in expanded newborn screening in Chine. *PLoS One, 7*(4), 1-6.

Van Den Heuvel, M., Borkhoff, C. M., Koroshegvi, C., Zabih, W., Reijneveld, S. A., & Maguire, J. (2016). Diagnostic accuracy of developmental screening in primary care

at the 18-month health supervision visit: a cross sectional study. *CMAJ Open, 4*(4), 636-640.

Van der Lee, J. H., Opmeer, B. C., Aarnoudse-Moens, C. S., Leenders, A. G., Mol, B. W., & De Haan, T. R. (2015). Predicting developmental outcomes in premature infants by term equivalent MRI: Systematic review and meta-analysis. *Systematic Reviews, 4*(1), 71-81.

Verhaeghe, L., Dereu, M., Warreyn, P., De Groote, I., Vanhaesebrouck, P., & Roeyers, H. (2016). Extremely preterm born children at very high risk for developing autism spectrum disorder. *Child Psychiatry and Human Development, 47*(5), 729-739.

Wetherby, A., & Prizant, B. (2002). *Communication and symbolic behavior scales developmental profile-First normed edition.* Baltimore: Brookes.

Whiting, P. F., Rutjes, W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., . . . Bossuyt, P. M. (2011). QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine, 155*, 529-536.

Wilson-Costello, D., Friedman, H., Minich, N., Fanaroff, A. A., & Hack, M. (2005). Improved survival rates with increased neurodevelopmental disability for extremely low birth weight infants in the 1990s. *Pediatrics, 115*(4), 997-1003.

Wong, H. S., Santhakumaran, S., Cowan, F. M., & Modi, N. (2016). Developmental assessments in preterm children: a meta-analysis. *Pediatrics, 138*(2), 1-13.

World Health Organization. (2010). *ICD-10: International statistical classification of diseases and related health problems, tenth edition.*

Yaari, M., Yitzhak, N., Harel, A., Friedlander, E., Bar Oz, B., Eventov Friedman, S., . . . Yirmiya, N. (2016). Stability of early risk assessment for autism spectrum disorder in preterm infants. *Autism, 20*(7), 856-867.

Yong-Hwee, N., Young, R. L., Brewer, N., & Berlingeri, G. (2014). Autism Detection in Early Childhood (ADEC): Reliability and Validity Data for a Level 2 Screening Tool for Autistic Disorder. *Psychological Assessment, 26*(1), 215-226.

Zwinderman, A. H., & Bossuyt, P. M. (2008). We should not pool diagnostic likelihood ratios in systematic reviews. *Statistics in medicine, 27*, 687-697.

CHAPTER TWO

A PILOT STUDY OF THE RELATIONSHIP BETWEEN SHARED INTENTIONALITY AND SOCIAL COMPETENCE IN TYPICALLY DEVELOPING PRESCHOOL CHILDREN

Abstract

Introduction. Early social-cognitive abilities are proposed to underpin atypical development of Theory of Mind evident in autistic people. However, tests of Theory of Mind are not widely used as part of autism diagnosis. Therefore, this pilot study begins to examine the links between a scalable measure of precursors of Theory of Mind (the Early Social Cognition Scale) and a widely used measure in diagnosing autism (the Autism Diagnostic Observation Schedule – Toddler Module).

Methods. Twelve typically developing children were enrolled into this pilot study. All children were assessed using the Mullen Scales of Early Learning, Autism Diagnostic Observation Schedule-Toddler module and the Early Social Cognition Scale. We estimated the effect size of the correlation between children's Autism Diagnostic Observation Schedule (Toddler Module) assessed social competence and shared intentionality measured by the Early Social Cognition Scale.

Results. Using spearman's rho, we found a very small negative correlation between children's performance on the Autism Diagnostic Observation Schedule (Toddler Module) and the Early Social Cognition Scale. A power analysis revealed that a large-scale study would require a sample size of 1716 children to demonstrate statistical significance of an effect of this size.

Discussion. The findings suggest that early developing shared intentionality may be related to social competence, however the relationship may be very small. Although further large-scale study is required to draw firmer conclusions regarding this relationship, recruiting 1716 children is not feasible.

Introduction

Autism is defined by difficulties in social communication and social interaction, and restricted or repetitive behaviours (American Psychiatric Association (APA), 2013). Multiple theories have been proposed to understand the underlying cognitive differences that underpin these diagnostic difficulties (Baron-Cohen, Leslie, & Frith, 1985; Frith, 2003; Hobson, 2002; Pennington & Ozonoff, 1996; Tager-Flusberg, 1993). Amongst the most widely researched theories have been differences in social cognition (the ability to make sense of social information; Leslie & Happe, 1989; Turner, 1999). Whilst there is good empirical evidence for social cognitive differences between autistic children and their typically developing peers (see Bora & Pantelis, 2016 for a review), tests of social cognition have remained surprisingly absent from diagnostic practice. For this reason, we actually know very little about how individual differences in social cognition relate to individual differences in real-world difficulties with social communication and social interaction. This study begins to bridge that gap. We present a pilot study comparing individual differences in social cognition with individual differences on the Autism Diagnostic Observation Schedule-2nd edition Toddler module (ADOS-2^{T}), in typically developing children aged 18-36 months. The pilot study will provide an effect size estimate, against which a fully powered study can be designed.

Social Cognition and Theory of Mind

Social cognition is a multi-dimensional cognitive process that underlies the ability to make sense of the sharing of social information (Brothers, 1990; Schonherr & Westra, 2019). Such cognitive processes include perceiving and interpreting social information such as emotional expressions, gestures, gaze and actions (Denham, McKinley, Couchoud, & Holt, 1990; Gobel, Kim, & Richardson, 2015; Straube, Green, Jansen, Chatterjee, & Kircher, 2010) and representing the mental states of others, such as beliefs, desires, intentions and goals

(Clements & Perner, 1994; Uithol & Paulus, 2014). Representing the mental states of others based on observable behaviour and responding accordingly is often described as Theory of Mind (ToM), mentalizing, or mindreading, which falls within the broader framework of social cognition. ToM and the developmental precursors of ToM have been a particular focus in autism research.

Scalable Measures of ToM

Traditional measures for assessing ToM involve children observing a protagonist, and making explicit responses about the protagonist's beliefs, desires, intentions, or behaviours (Baron-Cohen, Leslie, & Frith, 1985; Bennett, 1978; Wimmer & Perner, 1983). Only by understanding that other people's behaviour is driven by their own mental states, which might be different from the child's own, can children pass such tasks. Such measures enable researchers to explore the development of children's ToM.

It is now widely accepted that among typically developing children, a consistent developmental progression emerges, whereby ToM tasks are passed in a scalable fashion (Osterhaus, Koerber, & Sodian, 2016; Wellman & Liu, 2004; Wellman, Fang, & Peterson, 2011). This cumulative development can be captured by scalable measures of ToM that tap into ToM abilities that are conceptually similar but increase in complexity as the scale progresses. Using such a scale, Wellman and Liu (2004) found that understanding of desires precedes understanding of false beliefs, which in turn precedes understanding of hidden emotions (Wellman & Liu, 2004). They found that 80% of 3.5 – 6-year olds followed this stringent pattern of performance. Numerous studies have replicated these tasks with similar findings (Peterson, Wellman, & Liu, 2005; Peterson, Wellman, & Slaughter, 2012; Zhang, Shao, & Zhang, 2016). However, research has found that autistic children's progression of abilities in understanding others' mental states follows a different, though consistent, scalable

pattern of development; passing tasks in a different sequence to typically developing peers (Peterson et al., 2005), and at later stages of development (Happe & Siddons, 1994; Hoogenhout & Malcolm-Smith, 2014).

ToM precursors

More recently, researchers have developed measures that explore the early social cognitive abilities that may underpin ToM (precursors to ToM). Instead of requiring explicit verbal responses from children, these tasks measure infants' spontaneous behaviour such as helping and co-operation (Warneken & Tomasello, 2006). Such precursors to ToM have been demonstrated in children as young as 14-months old (Surian & Franchin, 2017), allowing researchers to measure the lower limits of ToM abilities (Hiller, Weber, & Young, 2014; Ellis et al., 2020). Around this age, children also begin to use mental state words to represent desires and beliefs (Bretherton & Beeghly, 1982; Moore, Furrow, Chiasson, & Patriquin, 1994). These early skills are hypothesised to be the foundations for later developing ToM abilities such as explicit understanding of others' mental states. Therefore, examining proficiency in these precursors to ToM is important in understanding the developmental trajectories of ToM in typical and atypical populations. Early identification of atypical trajectories can facilitate early monitoring or intervention where necessary (Webb & Jones, 2009) and help researchers to understand normative benchmarks and deviations in developmental trajectories associated with neurodevelopmental disorders. One such ToM precursor is shared intentionality. Shared intentionality refers to collaborative activities that require sharing psychological states with another person i.e., desires or intentions (Goodman & Tomasello, 2008), and is measured by Ellis et al's (2020) Early Social Cognition Scale (ESCogS).

The Early Social Cognition Scale (ESCogS)

Differentiating between intentional and accidental acts has been proposed to underlie more complex understandings of the intentions of others (Olineck & Poulin-Dubois, 2007). Based on Wellman and Lui's (2004) concept of developmental continuity from early emerging skills to more complex skills (cumulative development), Ellis et al. (2020) explored children's (aged = 14-34 months) development of shared intentionality through a series of tasks of increasing difficulty (ESCogS). The authors found a clear sequence of developing abilities in shared intentionality, with the majority of children passing tasks in a stringent order; a) helping, b) re-enactment of intended acts, c) understanding intention through communicative gestures, and lastly d) sharing an intention with an adult to achieve a joint goal. Indeed, they found that 88% of typically developing and 90% of autistic children passed tasks in this stringent order. However, autistic children demonstrated lower pass rates for the most difficult tasks, with 62% passing only the first two steps. These results indicate that autistic children followed the same pattern of development of shared intentionality. However, this development is delayed compared to their peers. Scaling tests in this manner allows researcher to precisely examine divergent or delayed trajectories of ToM development.

Theory of mind and Clinical measurement of Social Communication

Whilst ToM is demonstrably divergent and/or delayed in autism, and tools such as the ESCogS allow us to measure aspects of precursors of ToM, diagnostic tools rarely capture these abilities. One of the most widely used autism diagnostic tools in clinical settings is the Autism Diagnostic Observation Schedule, 2nd edition (ADOS-2; Akshoomoff, Corsello, & Schmidt, 2006; Ashwood, Buitelaar, Murphy, Spooren, & Charman, 2015; Lord et al., 2009; Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011). Considered the gold standard diagnostic tool, the ADOS-2 (Lord, Rutter, DiLavore, & Risi, 1999; Lord, et al., 2006) is a

standardised diagnostic instrument that evaluates social communication and interaction (defined by the Social Affect domain) through a series of structured and unstructured activities. In total there are five ADOS-2 modules. Modules are selected for assessment based on individuals' chronological age and verbal ability. The ADOS-2 is not designed to measure children's developing social cognition, despite an assumption that atypical social cognition may underlie the behavioural presentation in autism. Rather, it measures children's ability to demonstrate conventional social behaviour in response to structured behavioural presses. These responses are then coded and quantified into algorithms. Whilst differences in social communication and interaction (social competence) are defined by DSMV (American Psychiatric Association, 2013) as a core diagnostic component of autism, the links between social competence and social cognition are, as yet, unclear.

Theory of mind and social affect

It has been widely acknowledged that in order to successfully navigate social interactions, ToM abilities are required (Astington, 2003; Slaughter, Imuta, Peterson, & Henry, 2015; Tomasello, 2010; Weimer, Burleson, Stegall, & Eisenman, 2018). Indeed, a number of studies have found positive correlations between specific social interaction skills, such as making joint proposals with others and interpersonal communication skills such as fluency of speech, appropriate affect and level of engagement, with specific ToM abilities (Astington & Jenkins, 1995; Sasson, Nowlin, & Pinkham, 2013). However, it remains unclear how ToM relates to the more general domain of social competence. Due to the complexity of delineating the links between ToM abilities and social interaction abilities, the aetiology of atypical social competence development evident in autism is unclear. A small number of studies have attempted to explore the links between ToM abilities and social competence in typically developing and autistic populations

Shimoni, Weizman, Yoran, and Raviv (2012) explored correlations between a ToM measure (the Social Attribution Tasks, SAT) and the ADOS-G in a sample of typically developing (n=28) and autistic children (n=25) aged 7-18. The SAT consists of three indices; ToM affect (measuring affect talk), salience (measuring readiness to interpret ambiguous visual stimuli with social meaning) and pertinence (measuring the capacity to adhere to underlying social frames and impose meaning on visual displays). They found non-significant positive correlations between the ADOS-G Social Affect domain and pertinence (0.13) and salience (0.08) and a non-significant negative relationship with ToM affect (-0.06). In a sample of adults with high functioning autism (n=26), Rosenthal, Hutcherson, Adolphs, and Stanley (2019) found a significant correlation between the ToM ability to infer an agent's intentions from their actions and ADOS-G Social Affect domain scores.

Although findings from Rosenthal et al. (2019) and Shimoni et al. (2012) suggest that social competence as measured by the ADOS-2 Social Affect domain may be related to specific ToM domains such as shared intentionality, it is difficult to draw firm conclusions from such mixed findings. Both studies are limited in their use of tests that do not measure the trajectory of development of ToM (such as in Wellman & Liu, 2004). Instead they use single test paradigms that capture time-specific functioning. Furthermore, both studies (Shimoni et al, 2012) include small sample sizes, and so may be under-powered to detect true effect sizes. It is important that future studies are well powered in order to provide more conclusive evidence. Pilot studies that can provide vital data on putative effect sizes are required. Through this, appropriately powered studies can be designed and pre-registered, to increase the chances of a meaningful and replicable finding. Therefore, the use of a pilot study is even more imperative.

To our knowledge no study has explored the purported links between shared intentionality and social competence as measured by the ADOS-2 Social Affect Domain in

typically developing preschool children. Due to the ADOS-2's reliance on measuring social behaviour, it is possible that a null result may be achieved if an individual has appropriately learned social norms and expectations (i.e. when someone smiles, I should smile back). Indeed, Livingstone, Colvert, Bolton, & Happe (2019) found that high functioning autistic adults can achieve a null ADOS-2 outcome despite poor ToM. Therefore, early testing of ToM abilities and social competence may be more sensitive to atypical development in younger populations, compared to the older populations such as those recruited in Rosenthal et al's (2019) and Shimoni et al's (2012) studies.

Rationale

Due to the atypical development of ToM in children with autism, it is of clinical and scientific utility to delineate the links between ToM competence and social competence as assessed by the Social Affect subscale of the ADOS-2. Previous research exploring the links between ToM and ADOS-2 assessed social competence have used older atypical samples. Furthermore, they have not used scalable measures that are able to track the progression of ToM development. A more robust approach may be to compare ADOS-2 assessed social competence with a scalable measure of early social cognition (such as the ESCogS) in typically developing children. As typically developing children pass ESCogS tasks in the same order as their autistic peers, but with higher pass rates and at an early age, clearer association may be made between the development of shared intentionality and ADOS-2 assessed social communication in preschool populations. Furthermore, using a typically developing sample may provide insights into the developmental profile of children who do not score highly on the ADOS-2, but nevertheless have displayed developmental atypicalities significant enough to warrant an autism assessment. This pilot study aims to enhance the

current evidence base by exploring shared intentionality development and ADOS-2^{T} assessed competence in a small sample of typically developing children

Due to the limited amount of research exploring these links in preschool children, this project will be a pilot study. This will enable important parameters to be estimated such as required sample size for a suitably powered pre-registered future study.

Aim

This pilot study aimed to

1. Use a small sample of children aged 18-36 months to estimate the effect size for the correlation between shared intentionality, as measured by the ESCogS, and social competence measured by the ADOS-2^{T} Social Affect domain.
2. Use the obtained effect size to determine the sample size needed for an appropriately powered future study to test the significance of this relationship.

I hypothesise that scores on the ESCogS will be negatively correlated with scores on the ADOS Social Affect domain (social competence). The negative correlation reflects that a high score on the ESCogS reflects good performance, whilst a high score on the ADOS-2^{T} reflects cause for concern (poor performance).

Method

Recruitment

Children were recruited via social media and an electronic database held by the Birmingham Baby Lab at the University of Birmingham. Those recruited via the Birmingham Baby Lab (N = 9) previously consented to their data being stored as a method of recruitment for future projects within the University of Birmingham. Depending on the contact

information available, caregivers were contacted via email or telephone and asked if they wished to participate in the current project. Recruitment via social media involved posters being advertised on Twitter, Facebook and Instagram. Six participants contacted researchers directly, via email.

Participants

Children aged 18 – 36 months were invited to take part in a larger project (the Impact of Prematurity on Social Understanding (IPSU) project). Of these children, 12 completed the ADOS-2^{T}. Participants were excluded if they were diagnosed with autism. Participants were also excluded if they had a sibling with a diagnosis of autism, as first-degree relatives of those with autism have been evidenced to possess sub-threshold, genetically meaningful autistic traits (Losh et al., 2017; Shimoni et al., 2012).

Changes to recruitment strategies due to the COVID-19 pandemic.

Originally, the team made arrangements to hold assessment days at a local nursery. However, with increasing anxieties around COVID-19 and the growing risk of infection, parents attempted to minimise non-essential contact by withdrawing their children from nurseries. To prioritise participant safety, , the team made the difficult decision to cancel six assessments scheduled to take place throughout March. Shortly after this we were informed that in line with Government advice, all direct assessments with participants should cease immediately. These difficulties had a significant impact on the sample size of the empirical study. We anticipated a sample size of 25, however the total sample at the time of writing up the empirical chapter was 12. Recruiting from local nurseries may have resulted in a more representative sample. backgrounds who were familiar with socialising with adults outside of their family of origin, such as nursery staff. Fortunately, the sample size of 12 is the minimum recommendation for pilot studies (Moore, Carter, Nietert, & Stewart, 2011),

however a sample of 25 would have provided more confidence in the effect size obtained, and the opportunity to conduct significance testing.

Participant Characteristics

Sample demographic data on sex, age, gestational age and the Mullen Scales of Early Learning (MSEL; Mullen, 1995) scores are presented in Table 1.2.1. One child had cognitive ability within the very low range, one below average, three average, four above average, and two very high. Participant 4 did not complete the receptive language domain and so overall cognitive ability could not be determined. Participants were 12 children (seven females, five males; aged 18 – 28 months, M = 23.57, SD = 2.99). Eleven children were classified as little – no concern of autism, one was classified as mild-moderate concern of autism.

Table 1.2.1.

Individual child demographics and characteristics

				MSEL				
Participant	Child age	Sex	Gestational age	Visual reception	Fine motor	Receptive language	Expressive language	Composite score
1	18.43	M	38.86	59	69	67	48	122
2	20.79	F	41.14	54	52	56	40	101
3	21.36	M	38.86	56	58	42	60	108
4	21.49	M	38.43	41	41	N/A	36	N/A
5	22.07	F	41.14	52	49	65	68	117
6	22.23	F	37.43	70	45	51	56	111
7	24.30	M	40.29	59	64	71	54	124
8	24.56	M	40.14	70	64	56	60	125
9	26.13	M	36.71	37	23	30	30	63
10	26.33	M	40.14	68	62	69	68	133
11	26.46	M	36.71	30	35	37	42	73
12	28.76	F	39.71	80	64	67	73	141
Mean	23.57		39.13	56.33	52.17	55.55	52.92	110.73
Mean (SD)	2.99		1.57	14.80	14.03	14.05	13.79	24.00

Note: Child age is reported in fractions (months), gestational age is reported in fractions (weeks), MSEL = Mullens Scale of Early Learning, N/A Not Available

Procedure

The Science, Technology, Engineering and Mathematics (STEM) Ethical Review Committee at the University of Birmingham approved this study. A letter from STEM is documented in Appendix C. Written informed consent was obtained from the caregiver of each child, and each child was monitored for their willingness to participate, as a behavioural indicator of assent throughout the study (crying, clinging to caregiver, shaking their heads, refusal to participate).

Participation in the study required two research visits, each with an approximate duration of one-two hours. Children were assessed in a dedicated room at the University of

Birmingham. Caregivers remained present throughout the assessment, sitting on a chair behind the child. Each assessment was facilitated by two-three experimenters and was videotaped for further analysis. The child and Experimenter One sat together on a child size testing table. When children left the table, they were encouraged to return. If this caused distress, some tasks were conducted on the floor until the child was able to sit at the table again.

The same order of assessments was used for each participant. During visit one the ADOS-2^{T} was administered, followed by eye-tracking tasks (eye-tracking data are not reported here). During visit two the ESCogS and MSEL were administered.

As this study is part of a wider project not all data collected are reported within this paper. The two visits occurred within eight days of each other, and caregivers chose assessment slots during the morning or afternoon depending on their own preferences.

Measures

The current study reports a subset of measures used within the IPSU project to target social cognition. The MSEL was used to measure cognitive ability, the ESCogS measured early precursors to ToM, and the ADOS-2^{T} measured social competence.

Mullens Scales of Early Learning (MSEL). The MSEL is a measure of cognitive and developmental ability in children aged 0-68 months and was used to estimate developmental ability in the present study. Commonly used in research (Chawarska, Klin, Paul, & Volkmar, 2007; Dawson et al., 2009; Lord et al., 2006; Somer, Bishop, Coffing, & Lord, 2011). The MSEL measures aspects of verbal and non-verbal development across five domains: gross motor, visual reception, fine motor, receptive language, and expressive language. T-scores are tabulated across the latter four domains to provide a composite score of general cognitive development (M=100, SD=15). As the gross motor domain is not

included in the composite score, this score is not reported in the current study. The MSEL was chosen for use in the current study as it has demonstrated good psychometric properties in typically developing populations, and populations with atypical development (Mullen, 1995; Swineford, Guthrie, & Thurm, 2015).

Autism Diagnostic Observation Schedule-2 Toddler Module. Performance on the ADOS-2^{T} was examined as this is a widely used "gold standard" diagnostic tool in autism assessments. ADOS-2^{T} was created for the assessment of children aged 12-to-30 months. Within the ADOS-2^{T}, one of two algorithms are used depending on each child's age and phonological ability. Each algorithm can be divided into two subscales; Social Affect (measuring social communication and interaction), and Restricted and Repetitive Behaviour (RRB). This pilot study particularly focuses on the Social Affect subscale as a measure of social competence. It does not explore the RRB domain for a number of reasons. Firstly, within the age range of children included within this study (18-33 months) repetitive and restricted behaviours are relatively common (Harrop et al., 2014; Kim & Lord, 2010; Richler, Bishop, Kleinke, & Lord, 2007). Furthermore, a review of studies exploring early development of such behaviours concluded that it is unclear what particular features of early emerging RRB are considered typical and atypical (Symons, Sperry, Dropik, & Bodfish, 2005). Secondly, RRB have been shown to follow a specific developmental pathway, progressing from sensory and motor behaviours in early infancy, followed by rigidity in performing tasks, routine rigidity, and then restrictive interests in later infancy (Larkin, Meins, Centifanti, Fernyhough, & Leekam, 2017). This development is not captured by the ADOS^{T1} (Kim & Lord, 2010) making examination of restricted and repetitive behaviours limited. This pilot study focuses on social cognition and social competence, and RRB's are beyond its scope

Within ADOS-2^{T}, the total score from the algorithm is applied to standardised cut-off scores that classify the range of concern for autism as little-to-no concern, mild-to-moderate concern, and moderate-to-severe concern. Few studies have explored the psychometric properties of ADOS-2^{T}. However, the authors (Lord, et al., 2006) have demonstrated high sensitivity (0.91) and specificity (0.87) for nonverbal and verbal toddlers (sensitivity, 0.88, specificity 0.91).

Tests of early social cognition. Early social cognition was examined using the ESCogS (Ellis et al., 2020). Each test within the play-based battery was selected from a well-established body of work and involve helping and joint co-operation activities, whereby the researcher and the child must work together to achieve a shared goal (shared intentionality). Each stage of the scale included one or two specific tasks (forming four stages), with the child required to pass one task at each stage in order to be coded as "passed"(see Figure 1.2.1.). Together, these stages form a reliable scalable pattern of cognitive development, with success on more difficult stages being significantly related to success on easier stages in 88% of participants (co-efficient of reproducibility (0.96), index of consistency (.5). The ESCogS was used to examine children's ToM through the domain of shared intentionality.

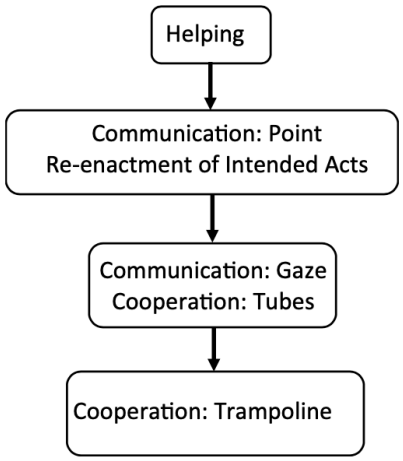


Figure 1.2.1. Four-stage scale of early social cognitive skills taken from Ellis et al. (2020)

The Helping task. Based on tasks developed by Warneken and Tomasello (2006), the helping task consists of two experimental trials. In the first trial the child observes the experimenter attempting to reach for objects to place in a box. In the second trial the child observes the researcher "accidentally" drop a pen on the floor. To pass the task the child is required to "help" the researcher by giving them the reached-for or dropped object in at least one of the trials. Typically, children demonstrate helping behaviour by 14-months old. In order to ensure that children were correctly identifying the experimenter's intentions in each trial (that the pen was dropped accidentally and that the researcher wished to obtain the out of reach objects), corresponding control trials were administered. In the first control the researcher deliberately threw the pen on the floor. In the second control trial the researcher is unable to obtain the objects without reaching.

The Re-enactment of intended acts. Based on a "demonstrate intentions" conditions used by Bellagamba and Tomasello, (1999), the re-enactment of intended acts task assesses infant's ability to infer another's intention from an unsuccessful goal-oriented action. In three experimental conditions the researcher "unsuccessfully" attempted to perform a target action; looping string on a peg, placing beads in a cup, and placing a ring on a peg. In order to pass this task, the child is required to perform the goal-oriented action successfully in two of three conditions. Typically, children pass this task by 18 months (Bellagamba & Tomasello, 1999).

The understanding others communicative cues task. Based on work by Behne, Carpenter, Tomasello, and Planck (2005) this task assesses children's ability to understand the communicative utility of pointing and gaze. Initially the child and experimenter one participate in a warm-up game. The experimenter places two identical boxes in front of the child, shows an attractive toy to the child and the hides it in one of the two identical boxes while saying "I'm going to hide it". The child is then invited to find the hidden toy. The task

consists of three warm up trials followed by four control and four experimental trials (alternating). Three identical pairs of boxes are alternated between trials to minimise perseveration errors (Behne et al., 2005). Perseveration errors would result in the child selecting the same box that they selected in the previous trial.

In each experimental condition, experimenter one hides a toy in one of two identical boxes for the child to find. Experimenter one then indicates to the child which box the toy is hidden in by saying "I can see it" and either alternating her gaze between the boxes or pointing. The task consists of four experimental trials (two communicative point and two communicative gaze). The child is required to locate the toy in both experimental trials in order to pass this task. Typically, children pass the communicative point task by 18 months and the communicative gaze task by 24 months.

In order to ensure that correct responses were not due to low level attentional cues, control trials were also administered. In control trials noncommunicative cues were used: 'distracted point' (finger extended towards the correct location of the toy with eye gaze directed downwards indicating that the experimenter was distracted), and 'control gaze' (gaze directed towards the correct location of the toy but with an absent minded facial expression).

The cooperation tasks. Two of Warneken et al's (2006) cooperations tasks were used to assess children's ability to collaborate with a partner to achieve a shared goal (shared intentionality). In each task, two experiemters first demonstrated the activity for the child. In the tube task two people are required to hold handles at either end of two overlapping tubes, and pull to retrieve the toy located inside. In the trampoine task two people are required to hold either end of a ring covered in cloth with a joint running down the middle, forming a small trampoline. The construction of the trampoline is such that two people are required to operate it sucessfully, otherwise it will collapse in the middle. In order to operate the

trampoline successfully, two people are required to hold it at either end and bounce an object place on the middle of the trampoline.

In both tasks, two demonstration trials are followed by four experiemtnal trials. In trials one and two the experimenter attempts to enagage the child in sucessfully cooperating to complete the task. In trials three and four, the experiment begins the task with the child but then ceases abruptly (letting go and looking at their hands on the floor) for 15 seconds before they resume their role. The child is scored on their attempts to reengage with the experiement (for example by making volcalisations) and their level of coordination (ability to sucessfully manipulate the tubes). A minimum medium score of three must be achieved across coorindation and reengagment conditions, and at least one attempt to re-engage with the experimenter must be made. Typically, children pass cooperation tasks from 24-months Warneken et al. (2006).

Scoring of the Early Social Cognition Battery. The ESCogS was analysed using the protocol published by Ellis et al. (2020). Scoring criteria for each stage can be found in Appendix D. At each stage of the ESCogS, the child was required to pass at least one of the tasks to be coded as "passed".

Statistical analysis

As the ADOS-2^{T} and the ESCogS are criterion-referenced measures, scores are not normally distributed in the same manner as "norm referenced" tests (Bradham & Houston, 2014), so non-parametric methods were used for statistical analysis. Correlations were conducted using SPSS Software, Version 26.0 (IBM Corp, 201). A spearman's rho was used to calculate an estimate of the correlation between ADOS-2^{T} assessed social competence and ESCogS. G*Power Version 3.1.9.6 (Erdfelder, Faul, & Buchner, 2005) was used to perform a

power analysis to determine the sample necessary for a full-scale study, given the effect size from the current sample.

Results

Task Performance on the ESCogS

To provide a comparison with Ellis et al. (2020), children's performance on the ESCogS is presented in Table 1.2.2. Table 1.2.2 presents tasks in order of difficulty (easiest – hardest) in the sequence reported by Ellis et al. (2020). Seven children (58%) conformed to this sequence, compared to 88% in Ellis et al's. (2020) sample. As expected, older children were more likely to pass more difficult tasks, as in Ellis, et al's study. However, there was some overlap of age between pattern 3 and 4 (see Table 1.2.2).

Table 1.2.2

Group scaling (individual performance at each stage of the ESCogS) and participant characteristics (age and Mullen composite score)

	Pattern					
	0	1	2	3	4	Other patterns
Participant number(s)	1	2	n/a	3,6,11	5,8	4,7,9,10,12
Helping	-	+	+	+	+	
REI or communication:						
Point	-	-	+	+	+	
Communication: Gaze						
or cooperation: Tubes	-	-	-	+	+	
Co-operation:						
Trampoline	-	-	-	-	+	
Number of cases	1	1	0	4	2	5
Age (mean months)	18.43	21.03	N/A	23.35	23.31	23.30
Age SD	N/A	N/A	N/A	2.73	1.76	4.20
Mullen composite score	122	101	N/A	97.33	121	115.25*
SD	N/A	N/A	N/A	21.13	5.66	35.52*

Note. + passed task, -failed task * Date for participant 12 in unavailable and so not included in the total score. N/A not applicable

Table 1.2.3 presents the observed patterns of performance in the five children whose performance diverged from this sequence.

Table 1.2.3

Other observed patterns on the ESCogS

Participant number	12	7,9,10	4
Helping	-	+	-
REI or communication: Point	+	-	+
Communication: Gaze or cooperation: Tubes	-	+	+
Co-operation: Trampoline	+	-	-
Number of cases	1	3	1
Age (months)	28.76	26.59	21.49
SD	NA	1.12	NA
Mullens	141	106.67	*
SD	NA	38.08	NA

Note.+ passed task, -failed task, * data not available, NA = not applicable

ADOS-2^{T} performance

All children completed the ADOS-2^{T}. Scores were calculated for the domains of Social Affect (M = 3, SD = 2.90) and RRB (M = 0.58, SD = 0.79). Total ADOS-2^{T} scores were calculated (M = 4.25, SD = 3.19) and used to classify children into concern for autism categories (little-no-concern, mild-moderate concern, and moderate-severe concern). Table 1.2.4 shows individual performance on the ADOS-2^{T} (Social affect, and RRB subscale and total scores).

Table 1.2.4

Individual performance on the ADOS-2^T

	ADOS-2^T			
Participant	Social affect	Restricted and repetitive behaviour	Total	Descriptive category
1	0	0	0	Little - no concern
2	2	0	2	Little - no concern
3	0	0	0	Little - no concern
4	9	1	10	Mild – moderate concern
5	3	1	4	Little - no concern
6	6	0	6	Little - no concern
7	6	1	7	Little - no concern
8	5	2	7	Little - no concern
9	1	0	1	Little - no concern
10	2	2	4	Little - no concern
11	3	0	3	Little - no concern
12	7	0	7	Little - no concern

The correlation between social competence and intentionality

Aim 1 of this paper was to produce an estimate of the correlation between children's shared intentionality and social competence measured by the ADOS-2^T Social Affect domain. It was hypothesised that scores on the ESCogS would be negatively correlated with ADOS-2^T assessed social competence. Spearman's rho yielded a very small negative correlation between social affect (ADOS-2^T) and shared intentionality (ESCogS) (ρ = -0.060).

A scatter plot of individual children's performance is shown in Figure 1.2.2. All participants who did not conform to the scale of progression evidenced by Ellis et al. (2020) scored 2 on the ESCogS. In contrast no child who followed the scalable pattern scored two.

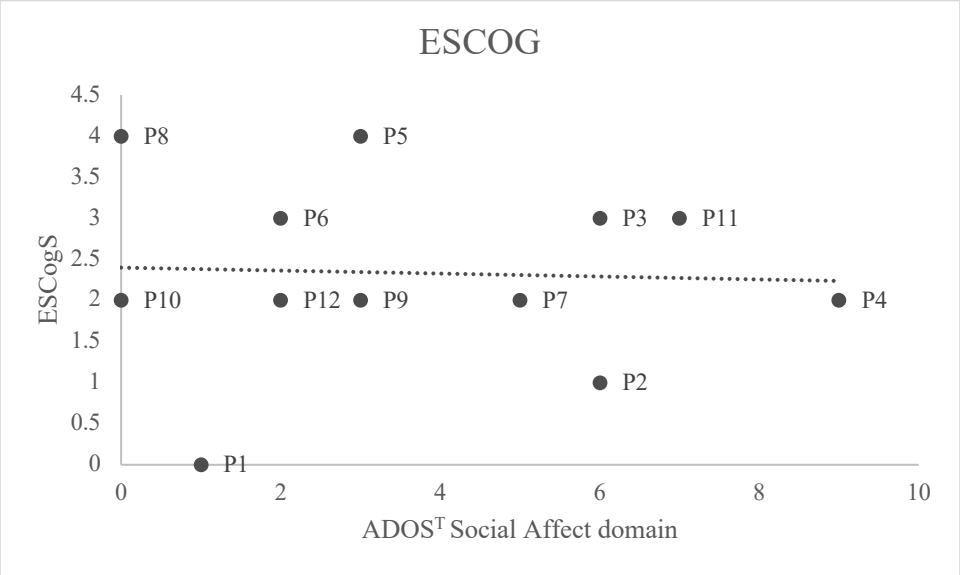


Figure 1.2.2. Individual participant performance on the Early Social Cognition Battery(ESCogS) and social competence assessed by the Autism Diagnostic Observation Schedule 2nd edition– Toddler module

Power analysis

In line with Aim 2, a power analysis was used to determine the minimum sample size required to produce a valid significance test for the Spearman correlation between ADOS-2^{T}-assessed social competence and ESCogS scores. The correlation in this study was -0.06, based on our one-tailed hypothesis, a sample size of 1716 would yield 80% power for detecting a correlation using $\alpha = 0.05$

Discussion

Previous studies have explored the relationship between aspects of ToM and clinical measures of social competence (such as the Social Affect subscale of the ADOS-2^T; Rosenthal et al., 2019; Shimoni et al., 2012). This was the first study to explore the relationship between shared intentionality and social competence in preschool children. This pilot study aimed to produce an estimate of the correlation between ADOS-2^T assessed social competence, and children's ability to engage in shared intentionality, as measured by the ESCogS, in a sample of typically developing preschool children. Within our small pilot sample, there was a very small negative correlation between scores on the ESCogS and scores on the ADOS-2^T Social Affect domain. A power analysis was conducted using this effect size. This returned a suggested sample of 1716 participants for an appropriately powered large-scale study.

Comparison with previous findings

In line with our predictions, our small pilot sample returned an estimate of a negative correlation between shared intentionality and social competence. However, the magnitude of this correlation was very small, resulting in an estimated sample size of 1716 required for a full study. If illustrative of the true relationship between these variables, this suggests that there is only a very small amount of shared variance in children's abilities in social competence and shared intentionality. The proposed sample suggested means a full-scale follow-up to this study in this form is unlikely to be feasible. Previous studies exploring the relationship between ToM abilities and social competence have found mixed results. Shimoni et al. (2012) found a non-significant positive relationship between social competence and indices of the SAT (salience; measuring readiness to interpret ambiguous visual stimuli with social meaning $\rho = 0.08$, and pertinence; measuring the capacity to adhere to underlying

social frames and impose meaning on visual displays, $\rho = 0.13$) and a non-significant negative relationship with the ToM affect index (measuring affect talk; $\rho = -0.06$). The effect size obtained between ToM affect and social competence is identical to the effect size obtained within this pilot ($\rho = -0.06$). Although the effect size of -0.06 is obtained from different ToM domains (ToM affect and shared intentionality), these correlations may be related to the same underlying abilities, that share a small degree of variance with the Social Affect domain of the ADOS-2. For example, understanding of others' affect and intentions requires sequential shifting from individual to shared perspectives that may be related to skills in joint attention or levels of social motivation captured by the ADOS-2. Indeed shared intentionality and mental state talk representing desires and beliefs emerge around the same chronological age, and so may be intrinsicaly linked (Bretherton & Beeghly, 1982; Moore et al., 1994; Surian & Franchin, 2017).

Rosenthal et al. (2019) found a significant negative correlation between understanding others' intent and social competence as measured by the ADOS-2 Social Affect domain ($\rho = -0.46$) in high functioning autistic adults. The effect size estimated in this study is much smaller than that observed by Rosenthal et al's (2019). However, the relationship between measures follows the same direction (negative), with higher scores on the ADOS-2 Social Affect domain more correlated with lower scores on understanding others intent.

Threats to validity

Pilot studies and small sample sizes

Though the direction of the effect from the small sample matched the initial prediction and the effect size was identical to an effect size found by Shimoni et al. (2012), there are reasons to be cautious. Whilst a pilot study aims to estimate a possible effect size, the small sample size by definition lacks power to do so accurately. Particular performance

patterns by individual participants can have a substantial impact on the results. A small sample is also less likely to be representative of a population from which it's drawn than a larger one. In this small sample, individual participant characteristics may have masked a larger effect between shared intentionality and social competence.

Furthermore, included children were aged 18- 28 months old. The significant amount of development that occurs within this age range may cause variability within data. The ESCogS is a useful tool in capturing the development of shared intentionality, with performance expected to improve with age. However, there is a body of evidence which has found that performance on the ADOS-2 may be confounded by age (Gotham, Risi, Pickles, & Lord, 2007; Gotham, Pickles, & Lord, 2009). It is recommended that a largescale study control for age to address this issue.

Novel assessment tools

The ESCogS was chosen as an appropriate assessment tool for the specific aims of this paper, with clear face validity in measuring the developmental maturation of shared intentionality abilities. However, it should be noted that although individual ESCogS tasks are taken from a body of established assessment measures, the complete battery is newly established (2020). Phillips, Leventhal and Burns (2017) highlight that a tool used to "operationalise a concept is not the same as the concept itself". Individual participant characteristics may alter the psychometric properties of tools, and so researchers should not overgeneralise such properties from one study to the next (Pedhazur & Schmelkin, 1991). Although the authors of the ESCogS demonstrated good reproducibility (0.96) and consistency (0.5), subtle participant differences may affect these properties. For example, a child with limited social exposure may underperform due to shyness rather than poor understanding of shared intentionality. Indeed, a lower percentage of children within this

study conformed to the sequence of shared intentionality development evidenced by Ellis et al. (2020), potentially affecting the scalability of the tool within this sample. . This disparity may be due to above expected atypicalities within this particular sample, or the comparably small sample size (N=12; Ellis et al.,2020, N = 86).

Taking these ideas into consideration we can consider an alternate explanation of the findings. Rather than reflecting a true effect size within the sample, the correlation co-efficient of -0.06 may be due to a diminished ability of the ESCogS in accurately capturing shared intentionality abilities within this small sample.

Strengths and Limitations

A number of strengths can be identified within the methodology of this pilot study. Previous studies comparing ToM and ADOS-2^{T} assessed social competence have focused on single ToM tasks. This pilot study however, used a scalable battery of assessments (ESCogS), as opposed to individual tasks, allowing consideration of the trajectory of development rather than static stages of development. Furthermore, this study extended the current literature by recruiting younger children (18-33 months) to explore the relationship between precursors of ToM (shared intentionality) and social competence.

As with all pilot studies, this study is limited by having a small sample. This means it serves to provide an estimate of a possible effect size, not a direct test of a hypothesis. However, such studies are necessary for high quality, reproducible science. This study measured social competence using the gold standard diagnostic tool for autism (ADOS-2^{T}), a tool commonly used in diagnostic practise, and so any findings can be directly related to diagnostic practise. It should be noted however, that the results of the pilot study are only applicable to the ADOS-2^{T}, as this study focused on early developing shared intentionality.

Clinical Considerations

The results of this pilot study demonstrate a lack of a large correlation between shared intentionality and ADOS-2^{T} assessed social competence. This indicates that the ADOS-2^{T} is not able to capture specific atypicalities in social cognition development that may assist in identifying those with and without autism. Administering autism-specific diagnostic tools that do not adequately capture the atypical development of autistic children may result in misleading information regarding the level of concern of autism. If young autistic children are misdiagnosed as typically developing, promising intervention may be delayed or denied. Research has found that the earlier interventions commence the more beneficial they will be, possibly due to the plasticity of young brains meaning there is more potential for remodelling (Dawson, 2008). Therefore, measures such as the ESCogS may be useful in diagnostic evaluation and planning specific interventions.

Research considerations

These finding are also of importance to researchers. The link between pre-school children's ability to engage in shared intentionality and social competence remains unclear, and so further investigation of this area is required. Drawing conclusions from the pilot study and previous research, further investigation is required into the relationship between ToM affect talk, shared intentionality and social competence to help delineate the links between ToM abilities and social interaction abilities.

An above expected number of children did not conform to the trajectory of shared intentionality development evidenced by Ellis et al. (2000), further investigation into the development of such children may shed light on the development of neurodevelopmental disorders or individual differences in other important areas of development such as social competence.

Recommendations for future study

The power analysis conducted following this pilot recommended that a sample of 1716 preschool aged 18-36 months to evidence a significant effect of this size. Due to the lengthy testing procedure per child (up to four hours), it is highly unlikely that a study on that scale will be conducted. Furthermore, a correlation of -.06 suggests minimal overlap of measures, regardless of whether they can be shown to share a statistically significant relationship. A more useful approach to take may be to estimate the required sample size to identify a statistically significant effect of small-medium size (-0.2 - -0.5). Our pilot study would predict that the null hypothesis would not be able to be rejected in such a case. With the support of Bayesian statistics, this would likely provide evidence in favour of largely independent variance in shared intentionality and social competence. In many ways such a finding could be more important to clinicians and researchers than one which showed a large overlap in variance. It would support the introduction of tests of shared intentionality to complement more traditional assessments, such as the ADOS-2^{T}.

Children who do not conform to the scalable pattern of performance identified by Ellis et al. (2020) may have specific atypicalities that require further investigation. Therefore, it is recommended that a large-scale study consider the impact of atypical ESCogS performance on the correlation found. Those with atypical shared intentionality development may represent specific differences in ToM development. Children showing atypical social development who do not meet specific diagnostic categories may still benefit from specialist support or interventions. Within this pilot study, ADOS-2^{T} assessment was limited to the toddler module. However, it is recommended that a large-scale study also uses ADOS-2^{T} module 1 and compares severity scores across modules, enabling older children to be assessed.

Summary

In an effort to understand the relationship between precursors of ToM and social competence, 12 children were enrolled in this pilot study. Children completed the MSEL, ESCogS and ADOS-2^T . A very small negative relationship was found between shared intentionality and social affect. Based on this very small effect size a full-scale study would require a sample size of 1716 to test the central hypothesis. Based on the data obtained in the pilot study, multiple recommendations for future study were provided.

References

Akshoomoff, N., Corsello, C., & Schmidt, H. (2006). The role of autism diagnostic observation schedule in the assessment of autism spectrum disorders in school and community settings. *California School of Psychology, 11*, 7-19.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed).* Retrieved from https://doi.org/10.1176/appi.books.9780890425596

Ashwood, K. L., Buitelaar, J., Murphy, D., Spooren, W., & Charman, T. (2015). European clinical network: Autism spectrum disorder assessments and patient characterisation. *European Child and Adolescent Psychiatry, 24*(8), 985-995.

Astington, J. W. (2003). Sometimes necessary, never sufficient: False-belief understanding and social competence. In B. Repacholi, & V. Slaughter, *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 13--38). New York: Psychology pRESS.

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition & Emotion, 9*(2), 151-165.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37-46.

Behne, T., Carpenter, M., Tomasello, M., & Planck, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Devlopmental Science, 8*(6), 492-499.

Bellagamba, F., & Tomasello, M. (1999). Re-enacting intended acts: Comparing 12 and 18-months-olds. *Infant Behavior and Development, 22*(2), 277-282.

Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences, 1*(4), 557-560.

Bora, E., & Pantelis, C. (2016). Meta-analysis of social cognition in attention-deficit/hyperactivity disorder (ADHD): Comparison with healthy controls and autistic spectrum disorder. *Psychological Medicine, 46*(4), 699-716.

Bradham, T. S., & Houston, K. T. (2014). *Assessing listening and spoken language in children with hearing loss.* San Diego, CA: Plural Publishing.

Bretherton, I., & Beeghly, M. (1982). Talking about internal states: The acquisition of an explicit theory of mind. *18*(6), 906-921.

Brothers, L. (1990). The social brain: A project for integrating primate behaviour and neuropsychology in a new domain. *Concepts in neuroscience, 1*, 27-51.

Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry, 48*, 128-138.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development, 9*, 377-395.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., . . . Varley, J. (2009). Randominzed, controlled trial of an intervention for toddlers with autism: The early start denver model. *Pediatrics, 125*(1), 17-23.

Denham, S. A., McKinley, M., Couchoud, E. A., & Holt, R. (1990). Emotional and behavioural predictors of preschool peer ratings. *Child Development, 61*, 1145--1152.

Ellis, K., Lewington, P., Powis, L., Oliver, C., Waite, J., Heald, M., . . . Crawford, H. (2020). Scaling of early social cognitive skills in typically developing infants and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 24*(2), 186-202.

Erdfelder, E., Faul, F., & Buchner, A. (2005). Power analysis for categorical methods. In B. S. Everitt, & D. C. Howell, *Encyclopedia of statistics in behavioural science* (pp. 1565-1570). Chichester: Wiley.

Frith, U. (1994). Autism and theory of mind in everyday life. *Social Development, 3*(2), 108-124.

Gobel, M., Kim, H., & Richardson, D. (2015). The dual function of social gaze. *Cognition, 136*, 359-364.

Goodman, M., & Tomasello, M. (2008). Baby steps on the road to society: Shared intentionality in the second year of life. *Zero to Three, 28*(5), 21-25.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 39*, 693-705.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule (ADOS): Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, 400-408.

Happe, F., & Siddons, F. (1994). Autism and theory of mind in everyday life. *Social Development, 3*, 109-125.

Harrop, C., McConachie, H., Emsley, R., Leadbitter, K., Green, J., & Pact Consortium. (2014). Restricted and repetitive behaviors in autism spectrum disorders and typical development: Cross-sectional and longitudinal comparisons. *Journal of autism and developmental disorders, 44*(5), 1207-1219.

Hiller, R. M., Weber, N., & Young, R. L. (2014). The validity and scalability of the theory of mind scale with toddlers and preschoolers. *Psychological Assessment, 26*(4), 1388-1393.

Hobson, R. P. (2002). *Autism and the development of mind.* London: Macmillan.

Hoogenhout, M., & Malcolm-Smith, S. (2014). Theory of mind in autism spectrum disorder: Does DSM classification predict development? *Research in Autism Spectrum Disorders, 8*(6), 597-607.

Kim, S. H., & Lord, C. (2010). Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the autism diagnostic observation schedule (ADOS). *Autism research, 3*(4), 162-173.

Larkin, F., Meins, E., Centifanti, L. C., Fernyhough, C., & Leekam, S. R. (2017). How does restricted and repetitive behavior relate to language and cognition in typical development? *Development and psychopathology, 29*(3), 863-874.

Leslie, A. M., & Happe, F. (1989). Autism and ostensive communication: The relevance of misrepresentatin. *Development and Psychopathology, 1*, 295-212.

Livingstone, L. A., Colvert, E., Bolton, P., & Happe, F. (2019). Good social skills despite poor theory of mind: Exploring compensation in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 60*(1), 102-110.

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

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

Lord, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., . . . Lord, C. (2009). The autism diagnostic observation schedule-Toddler module:A new module of a standardized diagnostic measure for autism spectrum disorder. *Journal of Autism and Developmental Disorders, 39*(9), 1305-1332.

Losh, M., Martin, G. E., Lee, M., Klusek, J., Sideris, J., Barron, S., & Wassink, T. (2017). Developmental markers of genetic liability to autism in parents: A longitudinal,

multigenerational study. *Journal of Autism and Developmental Disorders, 47*(3), 834-845.

Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T., & Manning-Courtney, P. (2011). Use of the autism Diagnostic Observation schedule (ADOS) in a clinical setting. *Autism, 15*(2), 143-162.

Moore, C., Furrow, D., Chiasson, L., & Patriquin, M. (1994). Developmental relationships between production and comprehension of mental terms. *First Language, 14*(42), 1-17.

Mullen, E. M. (1995). *Mullen scales of early learning learning manual.* MN: American Guidance Service, inc.

Olineck, K. M., & Poulin-Dubois, D. (2007). Imitation of intentional actions and internal state lanbguage in infancy predict preschool theory of mind skills. *European Journal of Developmental Psychology, 4*(1), 14-30.

Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory-of-mind tasks. *Child Development, 87*(6), 1971-1991.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach (Student ed.).* NJ: Lawrence Erlbaum Associates, Inc.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and develomental psychopathology. *Journal of Child Psychology and Psychiatry, 37*, 51-87.

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory-of-mind developmement for children with deafness or autism. *Child development, 76*(2), 502-517.

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or asperger syndrome. *Child Development, 83*(2), 469-485.

Phillips, A. L., Leventhal, H., & Burns, E. A. (2017). Choose (and use) your tools wisely: "Validated" measures and advanced analyses can provide invalid evidence for/against a theory. *Journal of Behavioral Medicine, 40*(2), 373-376.

Richler, J., Bishop, S. L., Kleinke, J. R., & Lord, C. (2007). Restricted and repetitive behaviors in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(1), 73-85.

Rosenthal, I. A., Hutcherson, C. A., Adolphs, R., & Stanley, D. A. (2019). *Deconstructing theory-of-mind impairment in high functioning adults with autism, 29*(3), 513-519.

Sasson, N. J., Nowlin, R. B., & Pinkham, A. E. (2013). Social cognition, social skill, and broad autism phenotype. *Autism, 17*(6), 655-667.

Schonherr, J., & Westra, E. (2019). Beyond 'interaction': How to understand social effects on social cognition. *The British Journal for the Philosophy of Science, 70*, 27-52.

Shimoni, H. N., Weizman, A., Yoran, R. H., & Raviv, A. (2012). Theory of mind, severity of autistic symptoms and parental correlates in children and adolescents with Asperger syndrome. *Psychiatry Research, 197*(1), 85-89.

Slaughter, V., Imuta, K., Peterson, C., & Henry, J. (2015). Meta-analysis of theory of mind and peer popularity in preschool and early school years. *Child Development*, 1159-1179.

Somer, L., Bishop, W. G., Coffing, M., & Lord, C. (2011). Convergent validity of the mullen scales of early learning and the differential ability scales in children with autism soectrum disorders. *American Journal on Intellectual and Developmental Disabilities, 116*(5), 331-343.

Straube, B., Green, A., Jansen, A., Chatterjee, A., & Kircher, T. (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia, 48*(2), 382-393.

Surian, L., & Franchin, L. (2017). Toddlers selectively help fair agents. *Frontiers in Psychology, 8*(944), 1-5.

Swineford, L., Guthrie, W., & Thurm, A. (2015). Convergent and divergent validity of themullen scales of early learning in young children with and without autism. *Psychological Assessment, 27*, 13641378.

Symons, F. J., Sperry, L. A., Dropik, P. L., & Bodfish, J. W. (2005). The early development of stereotypy and self-injury: a review of research methods. *Journal of Intellectual Disability Research, 49*, 144-158.

Tager-Flusberg, H. (1993). Evaluating the theory-of-mind hypothesis in autism. *Current Directions in Psychological Science, 16*(6), 311-315.

Tomasello, M. (2010). *Origins of human communication.* MIT Press.

Turner, M. (1999). Repetitive behaviour in autism: A review of psychological research. *Journal of Child Psychology and Psychiatry, 40*(6), 839-849.

Uithol, S., & Paulus, M. (2014). What do infants understand of others' action? a theoretical account of early social cognition. *Psychological Research, 78*(5), 609-523.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*(5765), 1301-1303.

Webb, S. J., & Jones, E. J. (2009). Early identification of autism: Early characteristics, onset of symptoms, and diagnostic stability. *Infants and Young Children, 22*(2), 67-74.

Weimer, A. A., Burleson, C., Stegall, S. E., & Eisenman, R. (2018). Theory of mind and social competence among school-age latino children. *Early Child Development and Care*, 1-9.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Chil Development, 75*, 523-524.

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development, 82*(3), 780-792.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Zhang, T., Shao, Z., & Zhang, Y. (2016). Chinese children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 83*(2), 210-220.

CHAPTER THREE

PRESS RELEASES

Autism screening tools are not as accurate in detecting autism in children born preterm.

Preterm births are births that occur prior to 37 weeks gestational age. Prevalence rates of autism are approximately five times higher in children born preterm than the general population (Agrawal, Rao, Bulsara, & Patole, 2018; Centre for Disease Control and Prevention [CDC], 2014). Therefore, it is important that clinicians are able to detect autism accurately in children born preterm as well as in the general population. It is common practice for clinicians to use brief screening tools in order to determine if a child shows developmental patterns characteristic of autism. If a high score is obtained on a screening tool, further diagnostic assessment is indicated. There are a number of screening tools available that have demonstrated good accuracy in identifying children who require further assessment (Sanchez-Garcia et al., 2019). This accuracy has only been established in the general population. However, emerging evidence has found that test accuracy is often poorer when applied to preterm populations (Wong, Santhakumaran, Cowan, & Modi, 2016). Until now our understanding of how screening tools perform in preterm populations has been unclear.

Lead researcher Kaysha Luces says “we cannot confidently administer screening tools to preterm children without first understanding how the accuracy of such tools differ when applied to children from the general population compared to children born preterm. Using screening tools with poor accuracy may ultimately result in misdiagnosis, and children being denied promising interventions”.

This study combined the findings from research exploring the diagnostic accuracy of autism-specific screening tools when applied to a preterm population and compared these findings to a similar study exploring the diagnostic accuracy of autism specific screening tools in the general population. The findings highlighted that overall, screening tools are less

accurate in identifying children that most need further diagnostic assessment when administered to a child born preterm. Specific difficulties were found with an increased rate of false positive results. False positive screen results may result in unnecessary diagnostic assessments being administered and increased parental distress (Tu, He, Chen, Shi, & Li, 2012; Gurian, Kinnamon, Henry, & Waisbren, 2006). It is imperative that clinicians understand the limitation of screening tools used in their practice and share this information with parents in order to decrease their distress.

References

Agrawal, S., Rao, S. C., Bulsara, M. K., & Patole, S. K. (2018). Prevalence of Autism Spectrum Disorder in Preterm Infants: A Meta-analysis. *Pediatrics, 142*(3), 1-15.

CDC, C. f. (2014, March 27). *CDC estimates 1 in 68 children has been identified with autism spectrum disorder: Latest snapshot shows proportion of children with autism and higher IQ on the rise.* Retrieved from cdc.gov: https://www.cdc.gov/media/releases/2014/p0327-autism-spectrum-disorder.html

Gurian, E. A., Kinnamon, D. D., Henry, J. J., & Waisbren, S. E. (2006). Expanded newborn screening for biochemical disorders: the effect of a false-positive result. *Pediatrics, 117*(6), 1915-1921.

Sanchez-Garcia, A. B., Galindo-Villardon, P., Nieto-Librero, A. B., Martin-Rodero, H., & Robins, D. L. (2019). Toddler Screening for Autism Spectrum Disorder: A Meta-Analysis of Diagnostic Accuracy. *Journal of Autism and Developmental Disorders*.

Tu, W. J., He, J., Chen, H., Shi, X. D., & Li, Y. (2012). Psychological effects of false-positive resullts in expanded newborn screening in Chine. *PLoS One, 7*(4), 1-6.

Wong, H. S., Santhakumaran, S., Cowan, F. M., & Modi, N. (2016). Developmental assessments in preterm children: a meta-analysis. *Pediatrics, 138*(2), 1-13.

Gold standard autism diagnostic assessment does not capture important autistic traits

In England, the Autism Diagnostic Observation schedule-2nd edition (ADOS-2) is one of the most commonly used tools to aid clinicians in diagnosing autism (Akshoomoff, Corsello, & Schmidt, 2006; Ashwood, Buitelaar, Murphy, Spooren, & Charman, 2015; Lord et al., 2009; Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011).

It is widely acknowledged that autistic children display specific differences in their understanding of the mental states of others (Happe & Siddons, 1994; Hoogenhout & Malcolm-Smith, 2014). For example, autistic children's ability to understand the intentions of others (shared intentionality) develops at a much later age than children without autism (Ellis, et al., 2020). Such abilities are suggested to be linked to social skills that autistic children tend to find difficult (Astington, 2003; Slaughter, Imuta, Peterson, & Henry, 2015; Tomasello, 2010; Weimer, Burleson, Stegall, & Eisenman, 2018). Contrary to expectations, shared intentionality development is not commonly explored when children are being assessed for autism. Furthermore, there has been very little research exploring the links between shared intentionality abilities and ADOS-2 scores. In an attempt to bridge this gap a research team at the University of Birmingham began to explore these links in typically developing preschool children. In collaboration with the Baby Lab, a novel pilot study examined the links between a measure of shared intentionality (ESCogS; Ellis et al., 2020) and scores on the ADOS-2T (Toddler Module).

Their findings identified little overlap between the two measures. If the same results are confirmed in a larger study this would demonstrate that the ADOS-2T does not capture atypical shared intentionality development evident in autism, and so may not identify the

children who are most in need of further support. It is recommended that clinicians use measures such as the ESCogS in clinical practice to provide a more adequate developmental profile of children who access their services.

Kaysha Luces, a member of the research team, says “Our findings have important implication for clinicians. The ADOS-2^{T} is a valuable tool in supporting clinicians in decided whether to diagnosis a child with autism or not. However, if this is the only assessment used clinicians risk missing important information that could provide a clearer picture of the child’s strengths and difficulties”.

Appendices

Appendix A

Database search terms

Psychinfo		
#1 Autism	"Autism Spectrum Disorders" (exp/TI/A) OR "Autism" (/TI/A) OR "Asperger syndrome" (TI/A) OR "autism spectrum" (TI/A) OR "child developmental disorders" (TI/A) OR "developmental disabilities" (TI/A)	56322
#2 Prematurity	"Premature birth" (exp/TI/A) OR "birth weight (exp/TI/A) "preterm infant" (TI/A) OR "pre-mature infant" (TI/A) OR "low birth weight" (TI/A) OR "Extremely low birth weight" (TI/A) OR "Very low birth weight" (TI/A) OR LBW (TI/A) OR ELBW (TI/A) OR VLBW (TI/A) OR "Extremely preterm" (TI/A) OR gestation (TI/A)	7818
#3 Diagnosis	Diagnosis (exp/TI/A) OR "Early diagnosis" (TI/A) OR Detection (TI/A) OR "Early detection" (TI/A) OR "Early Identification" (TI/A) OR "delayed identification" (TI/A) OR "Early Prediction" (TI/A) OR "predictive validity" " (TI/A) OR "false positive" " (TI/A) OR "false negative" " (TI/A)	369073
#4 Screening	Screening (TI/A) OR "Early Screening" (TI/A) OR "Mass Screening" (TI/A) OR "Screening Tool" (TI/A) OR "Screening Tools" (TI/A) OR "Screening Test" (TI/A) OR "Screening Tests" (TI/A) OR "Screening Instrument' (TI/A) OR "Screening Instruments" (TI/A) OR Checklist (TI/A) OR Checklists (TI/A) OR Follow-up (TI/A)	194387
# 5	#1 AND # 2 AND # 3 AND # 4	64

Note: TI title search, AB abstract search, *truncation

Embase		
#1 Autism	"Autistic disorder (SH/TI/A) OR "autism spectrum disorder" (SH/TI/A) OR "autism" (SH/TI/A) OR "asperger*" (SH/TI/A) OR " "autism" (SH/TI/A) or "autistic" (SH/TI/A) OR "developmental disorder*" (SH/TI/A) OR "developmental disabiliti*" (SH/TI/A)	110183
#2 Prematurity	"Premature birth" (SH/TI/A) OR "birth weight (exp/TI/A) "infant, premature (exp) "premature birth" (SH/TI/A) OR "premature infant" (SH/TI/A) OR "infant, low birth weight" (exp) OR "Extremely low birth weight" (SH/TI/A) OR "Very low birth weight" (SH/TI/A) OR LBW (SH/TI/A) OR ELBW (SH/TI/A) OR VLBW (SH/TI/A) OR "Extremely preterm" (SH/TI/A)	196994
#3 Diagnosis	Diagnosis (SH/TI/A) OR detection (SH/TI/A) OR identification" (SH/TI/A) OR "prediction" (SH/TI/A) OR "predictive validity" " (SH/TI/A) OR "false positive" " (SH/TI/A) OR "false negative" (SH/TI/A)	4716363
#4 Screening	"mass screening (exp/TI/A) OR "screening" (SH/TI/A) OR "screen" (SH/TI/A) OR "screening tool*" (TI/A) OR "Screening Test" (TI/A) OR "Screening Tests" (TI/A) OR "Screening Instrument' (TI/A) OR "Screening Instruments" (TI/A) OR Checklist* (SH/TI/A) OR "follow-up" (SH/TI/A)	1403602
# 5	#1 AND # 2 AND # 3 AND # 4	163

Note: SH subject heading search, TI title search, AB abstract search, *truncation

Medline		
#1 Autism	“autistic disorder” (exp/TI/A) OR "autism spectrum disorder" (SH/TI/A) “autism” (SH/TI/A) OR “asperger*” (SH/TI/A) OR “autistic” (TI/A) OR “child developmental disorders” (TI/A) OR “developmental disabilities” (TI/A) OR “developmental disorder*” (SH/TI/A) OR “developmental disabiliti*” (SH/TI/A)	71685
#2 Prematurity	“Premature birth” (exp/SH/TI/A) OR “birth weight (/SH/TI/A) OR "infant, low birth weight" (exp) “infant, premature (exp) OR “pre-mature infant” (SH/TI/A) OR "preterm infant" (SH/TI/A) OR “low birth weight” (SH/TI/A) OR “Extremely low birth weight” (SH/TI/A) OR “Very low birth weight” (SH/TI/A) OR LBW (SH/TI/A) OR ELBW (TI/A) OR VLBW (SH/TI/A) OR “Extremely preterm” (SH/TI/A) OR gestation (SH/TI/A)	140021
#3 Diagnosis	Diagnosis (SH/TI/A) OR detection (SH/TI/A) OR "identification” (SH/TI/A) OR “prediction” (SH/TI/A) OR “predictive validity” ” (SH/TI/A) OR “false positive” ” (SH/TI/A) OR “false negative” (SH/TI/A)	2851101
#4 Screening	"mass screening » (exp) OR “screen” (SH/TI/A) OR checklist (TI/A) OR follow-up (SH/TI/A) OR "positive screen" (SH/TI/A) OR "false positive" (SH/TI/A)	842438
# 5	#1 AND # 2 AND # 3 AND # 4	16

Note: SH subject heading search, TI title search, AB abstract search, *truncation

CINAHL		
#1 Autism	“autistic disorder” (MM) OR “asperger syndrome” (MM) OR autistic (TI,AB) asperger* (TI,AB) “developmental disabilities” (MH) “developmental disability*” (TI,AB)	33902
#2 Prematurity	“infant premature” (MH) OR, “premature birth” (TI,AB) OR “preterm infant” (TI,AB) OR “birth weight (MH) OR “infant low birth weight” (MH) OR “low birth weight” (TI,AB) OR LBW (TI,AB) OR VLBW (TI,AB) OR ELBW (TI,AB)	37888
#3 Diagnosis	Diagnosis (MH,TI,AB) detection (TI,AB) identification (TI,AB) prediction (TI,AB) “predictive validity”	414444
#4 Screening	Screening (TI,AB) OR checklist (TI,AB) OR follow-up (TI,AB) OR “positive screen” (TI,AB) OR “false positive” (TI,AB)	323222
# 5	#1 AND # 2 AND # 3 AND # 4	47

Note: MM exploded and focussed/major concept search, MH subject heading search, TI title search, AB abstract search, *truncation

Appendix B

Rates of true positives and negatives and false positive and negatives extracted from individual studies

Study	FN	FP	TP	TN	N	Prevalence
Beranova et al (2017)	0	18	13	100	131	10%
Boon et al (2018)	6	174	16	332	528	4%
Dudova et al (2014)	4	9	9	112	134	10%
Gray et al. (2015)	-	13	1	-	14	-
Johnson et al. (2010)	2	19	9	143	173	6%
Kim et al (2016)	28	123	34	646	831	7%
Padilla et al (2017)	3	13	7	61	84	12%
Pritchard et al (2016)	-	12	3	-	15	-
Verhaeghe et al (2016)	2	10	19	16	47	45%
Verhaeghe et al (2016)	10	5	10	20	45	44%
Yaari et al (2016)	4	15	4	76	99	8%

Appendix C.

Letter from Ethics Committee

From: Susan Cottam (Research Support Group)
Sent: 28 May 2019 15:45
To: Caroline Richards (School of Psychology); Andrew Surtees (School of Psychology)
Cc: Catherine Laverty (PhD Psychology Lab FT); Kaysha Luces (ClinPsyD Clinical Psychol FT)
Subject: Ethics Application ERN_09-048AP9
Dear Dr Richards and Dr Surtees

Re: "The Impact of Prematurity on Social Understanding"
Ethics Application ERN_09-048AP9

Thank you for the above application to use Programme of Work ERN_09-048P. This has now been considered by the Science, Technology, Engineering and Mathematics Ethical Review Committee.

On behalf of the Committee, I can confirm a favourable ethical opinion for this application.

I would like to remind you that any substantive changes to the nature of the study as described in the Application for Ethical Review, and/or any adverse events occurring during the study should be promptly brought to the Committee's attention by the Principal Investigator and may necessitate further ethical review.

Please be aware that whilst Health and Safety (H&S) issues may be considered during the ethical review process, you are still required to follow the University's guidance on H&S and to ensure that H&S risk assessments have been carried out as appropriate. For further information about this, please contact your School H&S representative or the University's H&S Unit at healthandsafety@contacts.bham.ac.uk.

If you require a hard copy of this correspondence, please let me know.

Kind regards

Susan Cottam
Research Ethics Officer
Research Support Group
[redacted]
University of Birmingham
Edgbaston B15 2TT
[redacted]
Web: https://intranet.birmingham.ac.uk/finance/RSS/Research-Support-Group/Research-Ethics/index.aspx

Please remember to submit a new Self-Assessment Form for each new project.

Click Research Governance for further details regarding the University's Research Governance and Clinical Trials Insurance processes, or email researchgovernance@contacts.bham.ac.uk with any queries relating to research governance.

Appendix D

ESCogS scoring criteria

Stage	Condition	Target Behaviour	Pass Criteria
1	Helping	Passes the researcher their pen Passes or pushes a bell to the researcher	Complete one of two target behaviours
2	Communicative point	Selects the box with the toy on the first attempt	Correct box chosen on all control trials (2 of 2)
	Re-enactment of intended acts	Successful completes the act attempted buy the researcher	Successfully completes 2 of 3 acts
3	Communicative gaze	Selects the box with the toy on the first attempt	Correct box chosen on all control trials (2 of 2)
	Co-operation tubes	Level of co-ordination	0 = tubes not opened 1 = success after more than 5 seconds of inappropriate action 2 = success but some inappropriate actions for no more than 5 seconds 3 = success after immediate understanding of their role
		Level of engagement	0 = child disengages with the tack 1 = attempts to retrieve the hidden toy independently 2 = child waits on their side of the apparatus ready to engage in task 3 = attempts to reengage the researcher
4	Co-operation trampoline	Level of co-ordination	0 = child does not hold and lift trampoline 1 = Joint play but lots of stopping and not too excited. Infant needs a lot of persuasion. 2 = Some stopping or not too excited. 3 = Continuous play and rather excited (placing block on trampoline; initiating play; active shaking)
		Level of engagement	0 = child disengages with the tack 1 = attempts to retrieve operate the trampoline independently 2 = child waits on their side of the apparatus ready to engage in task 3 = attempts to reengage the researcher