Title

Sensitivity and Expected Change of Commonly Used Social Communication Measures in Longitudinal Research of Children with Autism

Author
Sterrett, Kyle

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Sensitivity and Expected Change of Commonly Used Social Communication Measures in Longitudinal Research of Children with Autism

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Education

by

Kyle Thomas Sterrett

2021

ABSTRACT OF THE DISSERTATION

Sensitivity and Expected Change of Commonly Used Social Communication Measures in Longitudinal Research of Children with Autism

by

Kyle Thomas Sterrett

Doctor of Philosophy in Education

University of California, Los Angeles, 2021

Professor Connie L. Kasari, Chair

Most social communication tests used to measure change in young children with autism spectrum disorder have not undergone rigorous psychometric evaluation. Notably, and most relevant to the early intervention literature, there is little information on the sensitivity of these measures to change over time, despite their frequent use. Further, most syntheses of intervention studies combine data based on constructs (e.g. language) without accounting for the potential error introduced when different measures are combined together. While this may be appropriate, more information is needed on whether there is substantial heterogeneity across commonly used social communication outcomes. The aims of this study were to examine the sensitivity to change and expected change over time of social communication measures in ASD clinical trial and longitudinal research studies. A systematic review and meta analyses was conducted to generate pooled effect sizes within each identified outcome measure, rather than pooling multiple measures together. Meta regression was used to determine whether the length of the measurement period was related to the magnitude of change over time and whether this relationship was influenced by factors such as cognitive ability, children's age and year of publication. The average length of the included studies ranged from 3 months to 20 months. Overall, the expected change over time, measured using standardized mean differences, was small to medium, although there was considerable variability. For example, ADOS Severity scores had an expected change of 0.114 standardized mean units and an average measurement period length of 16 months whereas Vineland socialization scores had an expected change of 0.3581 standardized mean units over a shorter average measurement period of 10 months. Most outcomes were not sensitive to change over time; the expected change over time was independent of the length of the study in 9 of the 42 measures. Further, change in some measures was influenced by factors like cognitive ability, chronological age, whether children were receiving behavioral interventions and how the outcome was reported (e.g. age equivalent versus standard score). These data suggest that careful consideration is needed when selecting an appropriate outcome measure and tests that measure similar constructs can vary considerably in their expected change over time. Some recommendations based on these data include: use caution when reporting standard scores to measure change over time, use both parent report measures and clinician administered measures to track progress in behavioral interventions, and use caution when combining different outcome measures to synthesize intervention data.

The dissertation of Kyle Thomas Sterrett is approved.

Mark P Hansen

Alexandra Sturm

Catherine Lord Morrison

Connie L Kasari, Committee Chair

University of California Los Angeles

2021

Dedication

This dissertation is dedicated to my grandfather, Colonel Robert J. Latina. Your legacy lives on through works like this; they would not be possible without the lifelong commitment to family, hard work and learning that you modeled for our family.

Table of Contents

List of Figures

List of Tables

List: Appendix A

List: Appendix B

Curriculum Vitae

Education:

PhD Education, University of California Los Angeles. In progress.
Master of Arts, Social Research Methodology, University of California Los Angeles,
B.A. Psychology, University of California Los Angeles
B.A. History, University of California Los Angeles

Publications

Harrop, C., **Sterrett, K**., Shih, W., Landa, R., Kaiser, A., Kasari, C. (2021). Short-Term Trajectories of Restricted and Repetitive Behaviors in Minimally Verbal Children with Autism Spectrum Disorder. *Autism Research.*

Dimachkie, A., **Sterrett, K**., Gulsrud A., Kasari, C. (2021). What are the Odds?: Predicting Likelihood of a Negative Episode in a Sample of Toddlers with ASD. *Autism.*

Kim, J., Freeman, S., Paparella, T., **Sterrett, K**. (2020). Pre-Communication Social Engagement Skills: A Curriculum for Children with Autism Spectrum Disorder and Global Delays. Different Roads to Learning Inc.

Byrne, K., **Sterrett, K**., Elias, R., Hus Bal, V., McCauley, J., Lord, C. (Revise and Resubmit). Trajectories of Major Health Issues Throughout Adulthood in Autistic Individuals. *Autism in Adulthood.*

Sterrett, K, Hayashida, K., Kim, J., Paparella, T., Freeman, S. (Revise and Resubmit). Evaluation of a Novel Curriculum Targeting Early Social Communication Skills in Minimally Verbal, Preschool Aged Children with Autism Spectrum Disorder (ASD): A Pilot Study. *Young Exceptional Children.*

Clarke, E., **Sterrett, K.,** Lord, C. (Under-Review). Work & Well-Being: Vocational Trajectories in Young Adults with Autism Spectrum Disorder. *Autism Research.*

Sterrett, K., Holbrook, A., Kaiser, A., Landa, R., Kasari, C. (Under-Review). The Role of Speech Generating Devices in Expressive Language Growth of Minimally Verbal Children with Autism Spectrum Disorder. *Augmentative and Alternative Communication.*

Dimachkie Nunnally, A., Baczewski, L., **Sterrett, K.,** Holbrook, A., Kaiser, A., Kasari, C. (Revise and Resubmit). Profiles and Trajectories of Executive Functioning in Young Children with Down Syndrome. *Journal of Intellectual Disability Research.*

Sterrett, K., Pizzano, M. (2019) Language Interventions for School Aged Children with Autism Spectrum Disorder. Autism Spectrum News.

Sterrett, K., Gulsrud, A., (2017) Developmental Approaches to Intervention for Children with Autism Spectrum Disorder. Encyclopedia of Autism Spectrum Disorders.

Kasari, C., **Sterrett, K.** (In Press) The JASPER Intervention Model. In Prelock and McCauley (Ed.) *Treatment of Autism Spectrum Disorders,* MD: Brookes Publishing Co.

Sterrett, K., Kasari, C., (2018) Review of Naturalistic Interventions for Children with ASD with Communication Delays. In J. Ganz and R.L. Simpson (Ed.) *Interventions for Individuals with Autism Spectrum Disorders with Significant Communication Challenge.* Baltimore, MD: Brookes Publishing Co.

Sterrett, K., Shire, S., Kasari, C. (2017) Peer relationships among children with ASD: Social acceptance, friendships and peer networks. In M. Fisher (Ed.) I*nternational Review of Research in Developmental Disabilities (IRRDD).*

Research and Teaching Experience:

2021
California State University Northridge
Adjunct Lecturer, Department of Child and Adolescent Development
Autism and Neurodevelopmental Disorders

Symposiums/ Paper Talks:

Sterrett, K., Libster, L., Shire, S., Kasari, C. (2020). Ratings of Social Functioning and Improvement: Perspectives from Specialists and Paraprofessionals. Accepted at American Educational Research Association Annual Conference.

Dickinson, A., Baczewski, L., **Sterrett, K**., Gulsrud A., Kasari, C., Jeste, S. (2020). Measuring Neural Changes with Early Intervention through Electroencephalography. Accepted at International Society for Autism Research Annual Conference.

Baczewski, L., **Sterrett, K.,** (2020; Co-Chairs). Family Stress & Well-being: Caring for individuals with ASD across the lifespan. Accepted at Gatlinburg Conference.

Gulsrud, A., **Sterrett, K.,** Kasari, C (2020). Initial Efficacy of a Short Term Parent-Mediated Intervention for Core Social Communication Skills: Baby Jasper. Accepted at American Academy of Child & Adolescent Psychiatry Annual Conference.

Select Poster Presentations:

Sterrett, K., Pizzano, M., Chiang, L., Kuhfeld, M., Sturm, A. (2019). An Item Response Theory Evaluation of the Autism Diagnostic Interview-Revised (ADI-R). Poster presented at International Society for Autism Research, Montreal, Canada.

Sterrett, K., Carr, T., Shire., S., Gulsrud, A., Kasari, C. (2018). Predictors of Reliable Change in a Parent Mediated Intervention for Children with Autism Spectrum Disorder. Poster presented at Gatlinburg Conference, San Diego California.

Introduction

Defining Autism Spectrum Disorder

Autism spectrum disorders (ASD) are a set of neurodevelopmental conditions both defined and diagnosed based on the presence of a specific combination of behaviors that present early in development, persist through life and interfere significantly with individuals' daily living (American Psychiatric Association, 2013). Specifically, the behavioral phenotype that characterizes ASD is marked by deficits and abnormal developmental trajectories of social communication skills and the presence of restricted and repetitive behaviors and interests. This phenotypic classification has remained remarkably stable since it was first recognized and conceived. Leo Kanner's thorough observations of a group of individuals in 1943 described a distinct pattern of behaviors; he noted social deficits such as their general difficulty in relating to other people as well as repetitive behaviors and interests such as spinning the wheels of a toy car or insistence on sameness in daily routines (Kanner, 1943). Although the field's more mature conceptualization of social communication delays in ASD is described below these core components remain central to our understanding of individuals with autism.

Social communication delays

Specific social communication deficits are well documented in terms of their prevalence, consistency, and stability. In general, social communication delays begin early in life with very young children showing differences in foundational social communication skills such as requesting, social orienting and joint attention (Dawson et al., 2004; Mundy et al., 1992; Charman et al., 1997; Mundy et al., 1994; Osterling et al., 2002; Paparella et al., 2011). Specific behaviors that have been shown to be delayed across these domains include use of requesting gestures such as pointing to request, joint attention gestures such as showing and giving to share, orienting to one's name, and sustaining eye contact (Charman et al., 1997; Mundy et al., 1994; Osterling et al., 2002; Paparella et al., 2011). These very early delays in social communication, specifically joint attention, are related to the development of more complex social communication skills such as language (Kasari et al., 2008; Mundy et al., 1990). Early concerns among parents are most often driven by social communication delays and are also the most common target of early intervention programs (Anagostou et al., 2014; De Giacomo & Fombonne, 1998).

Etiology of ASD and Social Communication Delays

While the phenotypic expression and trajectories of early social communication delays are well documented, what is less well understood, and underemphasized, is that each of the behaviors operationalized and described above (and in diagnostic manuals such as the DSM-5; American Psychiatric Association, 2013) are very likely manifestations of underlying biological and genetic differences in individuals with ASD. The role of genetics in ASD is an active area of research but overall the evidence suggests risk for ASD is genetically conferred (Geschwind, 2008; Robinson et al., 2016; Veenstra-VanderWeele et al., 2014) and that specific combinations of genes are likely linked to differences in biological and neural mechanisms in individuals with ASD when compared to typically developing children (Ruzzo et al., 2019). This means that although we understand the behavioral phenotypes that distinguish individuals with ASD, and that these differences are biologically based, the specific mechanism by which social communication delays manifest, progress and interact with the complex environmental influences present in early development remains unclear. There is no single unified theory of the underlying cause of the social communication differences that we observe.

Longitudinal Trajectories

Most theories attempting to describe social communication delays in young children with ASD have categorized behaviors based on the timing of onset and severity of observed behavioral delays and differences. As an example, how does the frequency or quality of a specific behavior of a child with ASD, such as joint attention, compare with the frequency or quality of a matched typically developing or developmentally delayed child without ASD. Relying only on cross-sectional comparisons can limit the practical relevance of the comparisons being made (Lord et al., 2015). The expected quality and frequency of specific skills can change rapidly throughout development and this change can be missed when looking at a single point in time. As an example, a typically developing child may be expected to have a burst of language from 12 to 18 months of age. If you compared those children with children with ASD at 12 months and not at 15 and 18 months of age valuable information would be lost and different conclusions could be drawn about the relationship between diagnostic status and language as those trajectories diverge.

Another important dimension to consider, beyond onset and severity of specific behaviors, is the evaluation of children's behavioral trajectories (i.e. the stability and change of specific behaviors across time). Trajectories can be used both as an outcome to determine what factors such as nonverbal IQ (NVIQ), gender and language predict specific trajectories (e.g. rapid growth, slow growth or no change) and can themselves be used as a predictor of specific outcomes. The amount of change in NVIQ between 18-24 months could be used as a predictor of social communication at 36 months of age for example. There have been a number of studies that have utilized the concept of trajectories as predictors or as outcomes themselves. Trajectories of restricted and repetitive behaviors (Richler et al., 2010), autism severity (Gotham, Pickles & Lord, 2012), adaptive functioning (Szatmari et al., 2015) and language (Anderson et al., 2007) have been analyzed. Each of these studies provides valuable information on the patterns of behaviors across time, have moved the fields understanding of autism forward and point to the value of increased attention on issues surrounding trajectories for young children with ASD.

Another body of research where the issues of trajectories becomes important is in early intervention and clinical trial literature for children with ASD. Both the variety of early interventions available for young children with ASD and the quality of the clinical trials testing the efficacy and effectiveness of those interventions has increased dramatically since the first clinical trials for children with ASD (Rogers & Vismara, 2008; Smith & Iadarola, 2015; French & Kennedy, 2018; Sandbank et al., 2020). While the field's understanding of what constitutes effective intervention continues to grow, a number of clear limitations to this body of literature still remain. One such issue, affecting both intervention science specifically and longitudinal research broadly, is the accuracy, reliability and validity of current measurement practices used to measure change over time.

Measurement Issues in ASD Research

Accurate and sensitive measures are critical in order to better understand the trajectories and behavioral intervention outcomes of young children with ASD in longitudinal research. One of the most consistent critiques over the last 20 years of ASD research has been the lack of standardized, ASD specific, developmental and behavioral measures that are sensitive to change, reliable and valid. In recent reviews, a total of 131 unique measures (excluding observational and study specific measures) were identified, but only twelve were found to have adequate evidence of validity (Mokkink et al., 2010; McConnachie et al., 2015). The large number of measures identified likely reflects the lack of consensus on the mechanisms of early social communication delays and as a result what and how those constructs should be measured.

Further, McConachie and colleagues (2015) concluded that while there is some evidence of validity in that small subset of tests, there is no evidence that they are reliable and sensitive to change over short periods of time (e.g. 3 months), the typical length of an early intervention trial for young children with ASD. This lack of evidence of the sensitivity to change of these measures also reflects a lack of consensus as to what constructs should be measured in clinical trials and what tests would be useful to do so (Bolte & Diehl, 2013). There have been attempts in recent years to develop new, higher quality and psychometrically valid measures such as the Eliciting Language Samples for Analysis (ELSA; Barokova et al., 2021) to measure language progress, the Autism Impact Measure (AIM; Mazurek et al., 2020) and the Brief Observation of Social Communication Change (BOSCC; Grzadzinski et al., 2016; Kim et al., 2019) to measure changes in autism symptomology.

The development of these measures represents a positive step forward. However, given the lack of consensus as to what appropriate outcomes should be, it is unlikely that there will be full, consistent and timely adoption of any single test purported to measure a particular construct of interest across the research community. Institutional memory. strong opinions about the appropriateness of specific measures and the need for consistency and comparability across past studies often drive selection of outcomes. It is thus important to undertake additional systematic efforts to better understand, evaluate, and provide data on the performance of the assessments that are currently available to measure change in the social communication skills to supplement ongoing measure development efforts.

Gap in the Literature

Understanding and being able to measure the trajectories of young children with ASD is critically important, both to our understanding of the etiology of social communication delays as well as our attempts to intervene on those delays. While the sensitivity of the various tests used to measure social communication outcomes has been called into question (McConnachie et al., 2015; Tager-Flusberg & Kasari, 2013), there have been no attempts to systematically measure the sensitivity to change over time or the expected change over time of commonly used tests of social communication for young children. Here, sensitivity to change is operationalized as the amount of change on an outcome being related to the length of the measurement period, this definition tracks whether more progress made in longer studies. Having a benchmark for expected change is important both because it establishes a metric by which to compare the results of new trials and places the results of old studies in a more appropriate context. No such benchmark exists in the early intervention literature for children with ASD. It is also important to understand how participant characteristics (e.g. cognitive ability) and study level characteristics (i.e. length of the study period and year of publication) may influence the magnitude of change and sensitivity to change over time across different outcomes.

To address the aims of this study a systematic review and meta analysis were conducted to generate point estimates of the expected change over time (effect size) of measures of social communication in ASD longitudinal research. Rather than being clustered at the study level in the meta analysis, the data were clustered across measures (i.e. a pooled estimate for each measure was generated). In previous meta-analyses of ASD intervention trials it is common to cluster across an outcome such as non-verbal IQ, adaptive behavior or language (Eldevik et al., 2009; Virués-Ortega, 2010; Sandbank et al., 2020) but in those analyses different tests are being used to represent the same outcome. There is no evidence that these measures should be considered psychometrically comparable and as such the focus of this meta analysis will not be on the outcomes (e.g. Expressive Language) but on the specific tests meant to measure those outcomes (e.g. Mullen Scales of Early Learning-Expressive Language Domain).

Aims

Primary aim

To determine the sensitivity to change and expected change over time of social communication measures in ASD clinical trial and longitudinal research. It is hypothesized that the expected change for a majority of social communication measures will be characterized by small to medium effects and that the size of the effect will be positively correlated with the length of the measurement period.

Secondary aim 1

To determine the influence of moderating variables including the presence of treatment, age of children, cognitive ability and type of measure (i.e. standardized, age equivalent, raw or developmental quotient) on change in social communication scores across measures. It is hypothesized that the expected change for social communication measures will be associated with or moderated by, (a) the presence of treatment, with enrollment in the intervention arm of a trial being associated with larger gains over time, (b) the age of children, with younger ages being associated with larger gains, (c) study year, with more recent studies being associated with larger gains over time (d) children's cognitive ability, with higher cognitive ability being associated with larger gains and (e) standard scores, with less change over time than age equivalent, raw or developmental quotient scores.

Method

Search Strategy

A systematic search of three large databases, PsychINFO, Web of Science and PubMed was conducted to cull relevant articles. These databases were chosen after consultation with a university librarian to allow for a wide breadth of journals to be accessed across different disciplines and to capture all relevant articles based on the inclusion criteria of the study. The full list of terms used for the search of the databases is included in Appendix A. The first layer of search terms consisted of various diagnostic labels related to ASD (e.g. autism, ASD, PDD), the next layer consisted of age labels (e.g. preschool, child, toddler) and the last layer consisted of names of specific social communication measures (e.g. PCX, Preschool Language Scales, Social Responsiveness Scale). The list of social communication measures was generated by reviewing previously published reviews and meta analyses (Bolte & Diehl 2013; McConachie et al., 2015).

Study Screening Process

A multi-stage screening and review process was used to determine whether each identified study met inclusion criteria for the review and meta-analysis. Organization and review of the identified studies took place using the Zotero software.

Inclusion Criteria

There were five inclusion criteria that a study must have met: (a) the study must have an experimental or quasi-experimental design, specifically it must have been randomized controlled trial or longitudinal study, (b) all the included children must have a confirmed diagnosis of autism or related disorders. No restrictions were placed on the method of diagnosis. Children could be diagnosed with a clinical measure such as the Autism Diagnostic Interview-revised (ADI-R; Rutter, Couteur & Lord, 2003) or Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2012), a review of records by the study team or from school personnel, (c) the mean age of the participants in the sample must fall between 24 months of age and 96 months of age during the entire length of the study. This age range was selected because two years of age is the age at which diagnosis of autism tend to stabilize (Klienman et al., 2008; Guthrie et al., 2013) and eight years of age typically represents the end of the early intervention period, (d) the study must have a measure of social communication that was administered at two time points or more and separated by at least two weeks— a list of social communication measures that were included in the search terms are listed in Appendix A – and (e) the studies must have been published between 1990 and 2020.

Database Search

The database search for this study took place over three stages but all three stages used identical search strings. Searches of the three databases took place in August 2017, June 2018 and May 2020.

Study Selection

Due to the amount of studies culled through the database searches, the study selection procedure took place over two stages. The first stage consisted of screening based on the title and abstracts; the goal of this screening step was to identify clear violations of the inclusion criteria, specifically, whether the age of the sample fell out of the required range, whether the study was longitudinal and whether the sample included children with ASD. The second stage consisted of a full review of each article, subjecting each study to the full set of inclusion criteria. A full accounting of the reasons why studies were excluded is provided in Figure 1. Any duplicate studies (some studies were identified through multiple databases) were deleted at this point so that only a list of unique articles remained. Interrater reliability data was collected; two raters screened a subset of the articles at both stages. Reliability statistics and procedures are reported in further detail below.

Hand Search

Following the completion of the two stage screening and study selection process a hand search was conducted to identify any additional articles that may not have been found through the database search. The hand search consisted of reviewing the reference lists of the included studies as well as the reference lists of other reviews and meta-analyses that had cited the included studies. After saturation was reached, and no new articles were found, the articles identified in the hand search were subjected to the inclusion criteria through the same procedures described above.

Data Extraction

Key dependent and independent variables were extracted from each individual study and combined into a single dataset for further analysis. These variables were all selected a-priori and as with the study screening step. A full list of the raw variables and data that were extracted from the included studies along with operationalization's are provided in Table 1.

Coding

While the raw data were used when possible, modifications were needed for some variables to be used in the final analyses. Descriptions of the coding procedure for those modifications is presented below.

Cognitive Ability. The data for cognitive ability across the samples was often not directly comparable, there were often different tests used to measure cognitive ability and they were often on different scales. For example, some studies reported age equivalent scores for specific domains, others reported standard scores, developmental quotients or overall mental ages. As a result of these differences the scores were transformed to a binary variable indicating whether the sample was on average more or less cognitively able. Age equivalent scores were converted to developmental quotients by dividing the average developmental age by average chronological age and multiplying by 100. For both developmental quotient scores and standardized scores, a cutoff of 70 was used where samples with average scores less than 70 were considered less cognitively able and samples with average scores greater than 70 were considered more cognitively able. Lastly, for domain scores (such as the visual receptive scores on the Mullen Scales of Early Learning) a cutoff of two standard deviations below the mean was used as the cutoff for more or less cognitively able.

Intervention Received. Due to extensive variability in the types of services received during the measurement period the decision was made to code the intervention data based on whether the sample was receiving treatment as usual (TAU), an identifiable behavioral intervention (Behavioral) or was receiving medication as a part of a pharmacological trial

(Medication). While the treatment as usual group could have been receiving outside services during the measurement period there was unreliable reporting of the amount and intensity of these services. Because of this, the categorical variable should not be viewed as intervention versus no intervention but rather more apt characterization would be: identifiable research based intervention, community based services or the medication arm of a pharmacological trial.

Length of the Measurement Period. This variable represents the total time from baseline to follow up or from one follow up period to another. For example, in an intervention trial with a two-month intervention phase and a four month follow up there would be two lengths, the two-month period from entry to exit and the four-month period from exit to follow up. In order to have a consistent scale for this variable the lengths of all measurement periods were converted to months.

Type of Outcome. In the cases of standardized measures, where the outcome could be reported in a number of ways, the scores were classified based on whether they were raw scores, age equivalent scores, developmental quotients or standard scores.

Study Quality. In meta analyses, data from different studies is collapsed together and so it becomes important to rate them based on their methodological quality. Evidence from high quality studies should be trusted more than evidence from medium quality studies and evidence from poor quality studies should be interpreted with caution, if at all. To that end, studies that met all of the inclusion criteria were evaluated for methodological quality and rigor.

The scale that was used to evaluate the methodological quality of the included studies was the Quality Assessment Tool for Observational Cohort and Cross Sectional Studies. This scale is recommended by the National Institute of Health (NIH) and was particularly suited to this study because it allows for both controlled and observational studies to be rated on the same scale. Although many of the studies included were randomized controlled trials, because the arms within the randomized trials are not being compared, they can be seen as observational cohorts. As a result, traditional quality rating scales of randomized trials that are often used in meta analyses would not have been appropriate. There were 14 items in the original measure however one item related to the participation rate of contacted participants was excluded as this information is rarely provided or available. This left 13 items which were rated as, “Yes”, “No” or “NA” (CD/NA/NR; cannot determine, not applicable or non reported). The information from the individual items is used to generate a holistic rating of study quality that rates studies as, “Poor Quality”, “Fair Quality” or “Good Quality.” The individual items are not summed or counted to come to the quality ratings, rather they are used in conjunction with substantive information gleaned by the rater after reading the manuscript to come to a final rating. The full measure is provided in Appendix A.

Reliability

At each stage of the study selection process approximately 20% of decisions were double coded and interclass correlations were run using Kappa statistics. In total, approximately 20% of the abstracts and titles reviewed for inclusion and 20% of the articles in the full review stage were reviewed by two independent coders.

Analytic Plan

The procedures for the meta analysis portion of the analysis follow the guidelines of Harrer and colleagues (2019). This procedure is broken down into five primary steps; (a) calculating the effect sizes for the study groups, (b) pooling the effect sizes, (c) identifying influential cases, (d) subgroup and meta-regression analyses and (e) assessing publication bias. The “Metafor” package (Viechtbauer, 2010) in R (version 4.0.2; R Core Team, 2020) was used to estimate the pooled effects across all models and generate forest plots and other graphics. Outliers and influential statistics and supplementary analyses were run using the "Dmetar" package (Harrer et al., 2019).

Effect size calculations

Hedges g (Hedges & Olkin, 2014) was used to generate a standardized mean difference (SMD) score for each group. Hedges g is generally more appropriate than Cohen's d (Cohen, 2013) for samples that have less than 20 participants which was often the case in these analyses. A group was defined not as a study but any group for which the outcome of interest was reported across two time periods (e.g. a two arm RCT would have two effect sizes calculated). In the case where there were more than two measurement time points within a group, separate standardized mean difference estimates were calculated for the difference between the first and second time points and the second and third time points.

Pooling the Effect Sizes

There are two primary methods that can be used to pool effect sizes in meta analyses. The first is using a fixed-effects model; fixed effects models assume that there is complete homogeneity across the sample of included studies (Borenstein et al., 2011). The other method is referred to as a random-effects model. Random-effects models do not have the same assumption of homogeneity across the sample and instead assume that there is variance between studies that indicate that they are not representative of a single population (Borenstein et al., 2021). Random-effects models are more commonly used in psychological research because of the lack of homogeneity in clinical populations (Higgins et al., 2003).

A random effects model using the Hartung-Knapp-Sidik-Jonkman estimator (Sidik & Jonkman, 2005) was used to estimate the pooled effect size and to generate confidence intervals for the pooled effect sizes. This estimator provides a more conservative estimate of effect sizes than other methods (e.g. DerSimonian & Laird 1986) and has been shown to be less biased in cases of high levels of heterogeneity between studies (Sidik & Jonkman, 2007).

Lastly, the standard errors for the pooled effects were estimated with Knapp-Hartung adjustments (Knapp & Hartung, 2003) which again has been advised in cases where there is expected to be substantial heterogeneity between studies.

The last consideration in the pooling of the effect sizes across the included measures was the potential multi-level structure of the data. It was often the case that single studies contributed multiple effect sizes to the pooled effect size within a measure because separate effects were estimated if there was more than one group in a single study (e.g. two groups within a randomized controlled trial) and also if there were more than two timepoints for a group within a study. The inclusion of effect size estimates from the same study could lead to issues with independence in the data. As is the case with many statistical tests, lack of independence in the data when pooling effect sizes can lead to an increased type-1 error rate. To address this potential concern, a third model was fit for each outcome with sufficient data, to account for the multi-level structure of the data and whether "study" accounted for a significant portion of the variance in the pooled effect size estimate.

The structure of this model was as follows: Participant at level 1, Group at level 2 and Study at level 3. These models were fit using the restricted maximum-likelihood (REML; Viechtbauer, 2005) estimator. First, the overall model was fit to obtain information on the amount of variance that each level of the model explained. Next, the fit of the overall model was compared to two nested models; the overall model without level 2 and the overall model without level 3. If level 3 did not explain a significant portion of the variance and did not improve model fit it was dropped and the simple random effects model described above was used.

Outliers and Influential Cases

Identifying outliers and influential cases is a critical step when conducting a meta-analysis (Viechtbauer & Cheung, 2010). It is important that the pooled estimates are not being influenced (pulled up or down) by any influential studies or outliers that have extremely large or small effects. A study was considered to be an outlier if the 95% confidence interval for the effect size of an individual study did not cross the 95% confidence interval of the pooled estimate. Simultaneously, influential cases were identified using the leave-one-out method which re-estimates the pooled effect size and confidence interval and I^2 statistic with each study removed one at a time (Viechtbauer & Cheung 2010). This evaluates the impact of removing each case on the overall pooled effect and heterogeneity. A visual inspection of the plot of the leave-one-out analysis along with information from the outlier analysis was used to identify cases that may have been unduly affecting the pooled effect size estimate. The final decisions about whether or not a case should be used was based on those data along with substantive knowledge of the studies in questions. However, a conservative approach was taken and in questionable cases analyses were run twice, with and with the dropped cases to determine if any observed effects were driven by those cases.

Sub-Group Analyses

Subgroup analyses allow for the comparison of effect sizes across two or more groups within the dataset. Subgroup analysis is only appropriate when a variable within the dataset is categorical and so there were be three separate subgroup analyses conducted within the proposed study. The three categorical variables of interest for the subgroup analyses were the binary cognitive ability variable, type intervention received and, when applicable, the type of measure (e.g. raw scores, age equivalent scores).

There are two steps to a subgroup analysis, the first step is to individually estimate the pooled effect size for each of the subgroups (i.e. those who received intervention versus those who did not) and the second step is to statistically compare the difference between the two pooled effect size (Borenstein & Higgins, 2013). The same estimation method, a random-effects model with the Hartung-Knapp-Sidik-Jonkman estimator (Sidik & Jonkman, 2005) and Knapp-Hartung adjustment (Knapp and Hartung, 2003) that was used for the overall pooled effect size estimate was used for the sub-group analyses. This model estimated the magnitude of the difference between the subgroup effect size estimates, the standard error of the difference and the confidence intervals for those standard errors. The null hypothesis is that there is no difference between the magnitude of the effect sizes between subgroups and so non-overlapping confidence intervals would indicate that we would reject the null hypothesis that there is no difference between the subgroups.

Meta-Regression Procedure

A multi-step process was used to evaluate the relationship between each continuous predictor (i.e. length of the measurement period, chronological age, and study year) and the overall pooled effect size as well as whether that relationship was moderated or influenced by the categorical variables listed above (i.e. cognitive ability, type of intervention received and type of measure). First, separate simple meta-regression models were run for each continuous predictor (i.e. a single independent variable and dependent variable in each model), next, because the length of measurement period variable was related to the central research question, each categorical variable was added in as an interaction term with length of the measurement period. If that interaction term was non-significant the interaction term was removed and length of measurement period and the categorical variable were included as simple main effects.

In meta-regression as well as the subgroup models the unit of analysis is at the study level (as opposed to the participant level in typical regression), as a result issues concerning sample size in the models were important to consider. It is recommended that there be at least ten studies per covariate (Borenstein & Higgins, 2013). In order to avoid post-hoc decision making about the appropriateness of whether there was sufficient data to include covariates, an a-priori rule was established that pooled effects would be estimated only when there were more than five effect sizes for a measure, simple meta-regression models would only be fit in cases with ten or more effect sizes for a measure and interaction terms were only included in cases with 15 or more effect sizes for a measure. Separate models were fit for each variable and interaction term in order to avoid the issue of over-specified models. No correction or adjustment of the significance level across the models for multiple comparisons was used.

Simple Meta-Regression. Meta regression is conceptually similar to subgroup analyses; the primary difference is that meta regression can handle data that are continuous. The remaining four variables of interest that were evaluated via meta-regression were the length of the measurement period (in months), children's age (in months), the study year and study quality.

To measure the fit of all of the models and the strength of the relationship between the independent and dependent variables in the meta-regression both the R^2 value, which measures the amount of variance explained by the predictors (R^2 here is comparable to that of an R^2 in linear regression) and the t-statistic for each predictor, which measure the strength of the relationship to the dependent variable, were used.

Meta-Regression with Interaction Terms. The restricted maximum likelihood estimator was used to generate the effect size estimates and meta-regression models that included interaction terms. In the case that an interaction terms was non-significant the simple main effect model was generated using the same estimation procedure.

All continuous predictors included in the meta-regression models were mean centered to allow for easier interpretation of the slope coefficients.

Publication bias

One issue in meta-analyses is that studies with larger effect sizes are more likely to be published, especially in higher quality journals, so it becomes important to determine the extent to which the studies that were culled are representative of the expected effect size of the population (Sterne et al., 2006). In meta analytic terms this effect is called publication bias. Publication bias was assessed using a visual inspection of a funnel plot along with Egger's test (Egger et al., 1997).

Given that the unit of analysis is the groups within studies, and not the comparison of change between the groups, the central marker of the funnel plot was not set at "0" but at the average pooled effect size for the measure. Because the variability of the estimated pooled effect sizes based on the covariates was the central question of interest. rather than the overall pooled effects, potential "publication bias" could also be reasonably attributed to variation in measured characteristics of the studies rather than bias. A funnel plot could identify a deviant estimate –an effect size of 1.0 for example – with no comparable estimate of -1.0 in the plot. However, an effect size of -1.0 is likely not in the data because we would not expect a child to decrease substantially on the measure, not because an effect size of -1.0 is unlikely to be published. Thus, a conservative re-estimation based on the Duval and Tweedie trim and fill procedure (Duval & Tweedie, 2000) was used only in cases of extreme or unexpected deviance based on substantive expectations.

Reporting of Results

When reporting contrasts from the subgroup analyses and slope coefficients from the meta-regression analyses, all coefficients (interaction terms and main effects) for intervention received, cognitive ability, chronological age, study quality and study year are non-significant unless specifically noted; only significant effects are reported below. The only exception is the effect of length of measurement period which is reported for each included measure.

Results

Study Screening

Across the three databases, 7098 articles were culled for review. During the abstract screening phase 6073 studies were excluded and in the full text review an additional 820 articles were excluded. Refer to Figure 1 for a flow diagram of the screening process, including the reasons for studies were not included. The hand search identified an additional 82 articles. A total of 203 articles met full inclusion criteria and contributed data to the meta analysis.

Characteristics of Included Studies

Of the 203 included articles, 13 were dissertations and 190 were peer reviewed articles. Descriptive information for the sample as a whole is provided in Table 2.

From the included studies, 119 unique measures of social communication were identified. A full list of identified measures is included in Appendix A. Many of the identified measures were used in only one study (n=55), and 42 measures included enough data (based on the a-priori guideline of needing greater than 5 effect sizes) to estimate a pooled effect size. In total, there were a total of 1232 effect sizes pooled together across the included studies

A list of measures for which effect sizes were pooled is provided in Table 3 along with the total number of effect sizes across each measure and average length of time of the measurement period across each measure. Additionally, for descriptive purposes the overall pooled effect for the random effect models across each measure (with and without outliers and influential cases removed) is provided and also visualized in a separate figure for each measure with potential outliers included (where applicable).

Quality of Included Studies

From the included studies, 58 were rated as "Good", 103 were rated as "Fair" and 42 were rated as "Poor". Complete ratings for each study are available upon request. Study quality was used as a predictor in the meta-analytic models and the effect of study quality is reported for each individual measure when there was sufficient data based on the a-priori criteria established above and when study quality was significantly associated with the pooled effect size within a measure.

Reliability

242 articles were double coded during the screening phase, the percent agreement on whether or not to include an article across the raters was 83.4% and the Kappa was 0.522. During the study inclusion phase 181 articles were double coded, the percent agreement across raters was 93.9% and the Kappa was 0.712.

Individual Measures

The addition of an author level random effect did not improve the fit of any of the estimated models nor did it explain a significant portion of the variance across any of the estimated models. As a result, all models discussed below were estimated without including a random effect for study author. The only exception was the model for ADOS Social Affect, a significant portion of the variance in the overall pooled effect was explained by the additional random effect. However, this effect was driven by a single outlier and when these data were dropped the author level random effect no longer explained a significant portion of the variance in the model. The results below are organized based on the amount of data present within each measure, with the more frequently reported measures discussed first.

Common Social Communication Measures

Autism Diagnostic Observation Schedule- Severity (ADOS; Gotham et al., 2009). There were 47 effect sizes included in the final analysis for ADOS Severity scores. The mean length of the measurement period (length) was 16.79 months (SD=13.48) and the median length was 12 months (IQR=18.72) across those effect size estimates. Overall, the average SMD was -0.1140, 95% CI [-0.2203; -0.0078]. Two influential cases were identified and while they had little effect on the estimate of expected change over time (SMD= -0.1146, 95% CI [-0.2179; -0.0113]), the exclusion of these cases reduced the overall I^2 from 48.8% to 42%. Based on the conventional rule of thumb the SMD would be considered small. A negative effect size here represents a reduction in symptoms.

The effect of length verged on significance (b=0.0068, se=0.0036, p=.066). For every additional month there is an expected increase in the effect size of approximately .01 units and because the SMD is negative this means that there is less expected change as length increases. However, the interaction between length and type of intervention received was significant and indicated that those who were receiving a behavioral intervention experienced more change as length of the measurement period increased compared to those receiving TAU (b=0.02, SE=0.0071, p=.0392).

The Egger's test was non-significant indicating that publication bias was not likely to be present for this measure (intercept=0.866, t=1.37, p=.18). Bubble plots and forest plots for the subgroup analyses for ADOS-Severity scores are provided in Figure 2.

ADOS Social Affect (Lord et al., 2000; Lord et al., 2012). There were 38 cases that were included in the analyses of ADOS Social Affect scores. The mean length of the measurement period was 10.68 months (SD=8.37) and the median was 7 months (IQR=6.02). The average SMD was -0.447, 95% CI [-.5947, -.2993] but there were two studies (5 effect sizes) that were identified as being influential and were excluded, the SMD with these cases removed was -0.388, 95% CI [-.4848, -.2912] and the I^2 was reduced from 81.3% to 23%. Based on the conventional rule of thumb the SMD would be considered a small to medium effect size. A negative effect size here represents a reduction in symptoms.

The effect of the length of the measurement period was non-significant (b=-.066, SE=.0055, p=.2381), indicating that larger effect sizes are not expected as the length of the measurement period increases.

No significant results emerged from the subgroup or meta-regression models, however there were two studies with extremely long measurement lengths (40 months) and so the subgroup and meta-regression models were re-run without these studies. With these cases removed there was a main effect of type of intervention received with the medication group having a larger effect overall compared to behavioral intervention group that approached significance (b=-.4655, SE=.25, p=.073) and a significant main effect of length of the measurement period (b=-.2075, se=.01, p=.011) when controlling for type of intervention received. For every additional month the effect size is expected to decrease by approximately .02 units. In the simple meta-regression model, length of the measurement period remained non-significant (b=.0194, SE=.0114, p=.10). There were only two effect sizes included in the pooled estimate for Medication so this result should be interpreted with caution.

Eggers test was significant (intercept= -1.437, t=4.049, p=.0003). However, on inspection the pattern observed in the funnel plot did not seem to be driven by publication bias but rather the imbalance seems to be driven by the fact that there were no studies that reported extreme increases in ASD symptomatology. It is therefore unlikely that publication bias is present in the reporting of this measure. Bubble plots and forest plots for the subgroup analyses for ADOS-Social Affect scores are provided in Figure 3.

ADOS Language and Communication (Lord et al., 2000; Lord et al., 2012). There were 20 effect sizes of ADOS Language and Communication scores present in the included studies. The mean length of the measurement period for those effect sizes was 7.21 months (SD=4.95) and the median length was 6 months (IQR=9.76). The average SMD was -0.6155, 95% CI [-0.8565, -0.3746], there was one potential influential case identified and with this case removed the SMD was -0.6624, 95% CI [-0.8995, -0.4254] and the final I^2 was

51.2%. Based on the conventional rule of thumb the SMD would be considered a medium effect size. A negative effect size here represents a reduction in symptoms.

In the simple meta-regression models the effect of the length of the measurement period was significant (b=-0.055, SE=0.019, p=.01), for every additional month there was an expected decrease in the average effect size of approximately .05 units. Additionally, there was a significant effect of chronological age (b=0.017, SE=0.004, p=.001) where for every month increase in the average age of the sample the effect size is expected to increase by approximately .02 units. When chronological age and length were included together in the model age continued to be significantly related to the pooled effect size (b=0.016, SE=.008, p=.05) but length was no longer significant (b=-0.003, SE=0.028, p=.92).

Those receiving a medication based treatment (k=4, SMD=-0.2979, 95% CI [-0.9447; 0.3489]) had the lowest expected change (p=.059) compared to the TAU (k=5, SMD= -0.8766, 95% CI [-1.2505; -0.5028]) and behavioral intervention groups (k=10, SMD= -0.6998, 95% CI [-1.0872; -0.3124]). Further, those in the higher cognitive group had nearly twice as much change on average than the lower cognitive group (k=6, SMD= -0.4179, 95% CI [-0.7166; -0.1192] compared to k=8, SMD=-0.7795, 95% CI [-1.4420; -0.1170]) but this difference was not significant due to the large confidence intervals.

The Eggers test was non-significant (intercept=-0.179, t=-0.183, p=.86) indicating that there was likely no publication bias present in this outcome. Bubble plots and forest plots for the subgroup analyses for ADOS-Language and Communication scores are provided in Figure 4.

ADOS Reciprocal Social Interaction (Lord et al., 2000; Lord et al., 2012). There were a total of 18 effect sizes included in the pooled effect size estimates for ADOS Reciprocal Social Interaction scores. The mean length of the measurement period was 6.57 months (SD= 4.8) and the median was 6 months (IQR= 9.87). The average SMD was -

0.4904, 95% CI [-0.7404; -0.2404], however 2 effect sizes were identified as potential influential cases. With these two effect sizes remove the SMD was -0.6262 [-0.8381; -0.4142] and the I^2 was 12%. Based on the conventional rule of thumb the SMD would be considered a medium effect size. A negative effect size here represents a reduction in symptoms.

There was a significant effect of length of the measurement period (b=-0.0416, SE=.0187, p=.04) where for ever additional month the effect size is expected to decrease by approximately .04 units. There was also a significant effect of age (b= .015, SE=.004, p=.002) where for every month increase in age the effect size is expected to increase (get closer to 0) by approximately .02 units. There was also a significant difference in the expected change between the lower and higher cognitive groups (p=.04) where the lower cognitive group (k=5) had an expected SMD change of -0.2904, 95% CI [-0.6064, 0.0256] units and the higher cognitive group (k=6) had an expected SMD change of -0.6575, 95% CI [-1.0358; -0.2793] units.

Lastly, there was a significant interaction between type of intervention received and length of measurement period where those receiving medication made more change over time than those receiving behavioral interventions (b=-0.5848, SE=.163, p=005). However, there were only three effect sizes in the Medication group so this contrast should be treated with caution.

The Eggers' test was non-significant (intercept=0.433, t=.316, p=.76) indicating that there is likely no publication bias present within this outcome. Bubble plots and forest plots for the subgroup analyses for ADOS-Reciprocal Social Interaction scores are provided in Figure 5.

Childhood Autism Rating Scale (CARS; Schopler et al., 1980). There were a total of 33 effect sizes pooled together for the analyses of the CARS, the average length of the measurement period was 9.98 months (SD=9.95) and the median length was 10 months (IQR=9). The estimated expected change was a SMD of -0.4932, 95% CI [-0.6853; -0.3010]. There was one effect size that was identified as an influential case and its removal had little effect on the overall expected change (SMD=-0.4656 [-0.6581; -0.2731]). The I^2 for this model was 69.1% and based on the conventional rule of thumb the SMD would be considered a medium effect size. A negative effect size here represents a reduction in symptoms.

The effect of length was non-significant (b=0.001, SE=.0096, p=.9132). Behavioral interventions (k=10, SMD= -0.7012 [-1.1316; -0.2708]) had larger expected change on average than TAU groups (k= 16 -0.3433 [-0.5683; -0.1182]) or Medication groups (k=6 -0.4185 [-1.1133; 0.2763]), but this difference was not significant due to the large standard errors (p=.26). There was not enough data to estimate the effect of cognitive ability on expected change and no other estimated effects were statistically significant for the CARS measure.

The Eggers' test was not significant (intercept=1.059, t=1.094, p=.28) indicating that there is not evidence of publication bias. Bubble plots and forest plots for the subgroup analyses for CARS scores are provided in Figure 6.

Early Social Communication Scales- Initiations of Joint Attention (ESCS IJA; Seibert et al., 1982). In total there were 42 effect sizes included in the meta analysis of ESCS IJA frequency, the mean length of the measurement period was 5.77 months (SD=4.29) and the median length was 3 months (IQR=8.74). The overall pooled effect estimate over this time was a SMD of 0.1936, 95% CI [0.0866; 0.3005] with an I^2 of 19.1%. There were no outliers or influential cases identified. Based on the conventional rule of thumb the SMD would be considered small.

The effect of the length of the measurement period approached significance (b=0.0224, SE=0.124, p=.08) where for every additional month one would expect the effect size for ESCS IJA to increase by 0.02 units. The effect of publication year also approached significance (b=-0.189, SE=.01, p=.06) where the expected effect decreases in studies published more recently. There was also a significant interaction between cognitive ability and length of the measurement period where those with higher cognitive ability had larger effects as length increased (b=0.57, SE=0.2259, p=.0173).

The Eggers' Test was non-significant (intercept=-0.36, t=-0.579, p=.57) indicating that there was likely no publication bias present across ESCS IJA outcomes. Bubble plots and forest plots for the subgroup analyses for ESCS IJA frequency are provided in Figure 7.

Early Social Communication Scales- Percent Response to Joint Attention (ESCS RJA; Seibert et al., 1982). There were 19 effect sizes included in the analyses of ESCS RJA outcomes, the average length of the measurement period was 6.92 months (SD=4.67) and the median length was 6 months (IQR=9). The expected change based on the overall pooled effect size estimate was a SMD of 0.1608, 95% CI [0.0121; 0.3095] however one case was identified as an influential case and dropped from the analyses. With this case dropped the expected change was a SMD of 0.1209, 95% CI [-0.0052; 0.2471] and the I^2 was approximately 0%. Based on the conventional rule of thumb the SMD would be considered very small.

Length of the measurement period was not significantly related to the size of the effect (b=0.004, SE=0.0146, p=.79). However, both study year and chronological age were significant predictors of effect size (b=-0.444, SE=0.0164, p=.02 and b= -0.157, SE=0.0069, p=.04 respectively) where more recent studies had less expected change and those who were older had less expected change respectively.

The Eggers' Test non-significant indicating that there was likely no publication bias present across ESCS RJA outcomes (intercept=-0.429, t=-0.511, p=.62). Bubble plots and forest plots for the subgroup analyses for ESCS RJA outcomes are provided in Figure 8.

Early Social Communication Scales- Frequency of Requests (Seibert et al., 1982). There were 18 effect size estimates included in the analysis of ESCS requesting frequency. The average length of the measurement period for this outcome was 6.9 months (SD=4.27) and the median length was 6 months (IQR=8.73). The overall expected change over this time was a SMD of 0.3154, 95% CI [0.0618; 0.5690], however one study was identified as an influential case and removed. With this case removed the expected change was a SMD of 0.2439, 95% CI [0.0133; 0.4745] and the I^2 of this final model was 36.9%. Based on the conventional rule of thumb the SMD would be considered small.

The effect of length of the measurement period was not significant (b=0.003, SE=0.028, p=.92), however, the interaction between age and length was significant. As chronological age increased the expected change over the measurement period is expected to decrease (b=-0.0224, SE=0.0089, p=.02). Further, the effect of study year verged on significance where more recent studies were expected to have smaller effects (b=-.0262, SE=0.0130, p=.06).

The Eggers' Test non-significant indicating that there was likely no publication bias present across ESCS Requesting outcomes (intercept=1.375, t=1.253, p=.22). Bubble plots and forest plots for the subgroup analyses for ESCS requesting frequency are provided in Figure 9.

MacArthur-Bates Communicative Development Inventories- Words Produced (MCDI; Fenson, 2007). There were a total of 33 effect sizes that were pooled in these analyses, the average length of the measurement period for this measure was 7.55 months (SD=5.01) and the median length was 6 months (IQR=7.23). The pooled effect size for the 33 effect sizes, which represents the average expected change over the measurement period, was a SMD of 0.4112, 95% CI [0.3102; 0.5121] with an I^2 of 32.8%. There were no outliers or influential cases identified and so all effect sizes were retained in the meta-regression and sub-group analyses. Based on the conventional rule of thumb the overall SMD would be considered small to medium.

There was an effect of study quality where those studies rated as “Poor” had the largest effect sizes (k=12, SMD=0.6888, 95% CI [0.4468; 0.9309]) compared to those rated as “Fair” (k= 22, SMD= 0.3699, 95% CI [0.2082; 0.5316]) and those rated as “Good” (k= 14, SMD=0.3714, 95% CI [0.1976; 0.5453]. Study quality was therefore included as a control variable in the rest of the models that were estimated.

The effect of length of the measurement period was non-significant (b=0.0114, SE=0.0103, p=.27). The effect of chronological age was significant where for every additional month there was an expected decrease of the SMD by .02 units (b=-0.232, SE=0.0044, p<.0001), further the interaction between chronological age and length of the measurement period was significant (b=-0.033, SE=0.0009, p=.0003). This indicates that the effect of age at entry depends on the length of the measurement period; younger children on average saw more change in shorter study periods than older children, this effect was not present in longer studies.

The Eggers’ test was non-significant indicating that the risk of publication bias was low for this measure (intercept=0.049, t=0.097, p=.92). Bubble plots and forest plots for the subgroup analyses for MCDI- words produced are provided in Figure 10.

MacArthur-Bates Communicative Development Inventories- Words Understood (MCDI; Fenson, 2007). Thirty-three effect sizes were included in the analyses for MCDI- words understood; the average length of the measurement period was 7.55 months (SD=5.01) and the median length was 6 months (IQR=9). The expected change across this time period was a SMD of 0.4027, 95% CI [0.2842; 0.5213], however there was one influential case identified and with this case removed the SMD was 0.3721, 95% CI [0.2580; 0.4862] and the I^2 was approximately 0%. Based on the conventional rule of thumb the SMD would be considered small to medium.

There was a significant effect of the length of the measurement period on the pooled effect size where the effect size is expected to increase by .02 SMD units every month (b=0.022, SE=0.0112, p=.05). There was also a significant effect of chronological age where every additional month is expected to lead to a decrease in the SMD by approximately .02 units (b=-0.0181, SE=0.0056, p=.003).

The Eggers' Test was non-significant indicating that there is likely not publication bias in the reported effect sizes for MCDI- Words Understood (intercept=-0.731, t=-1.384, p=.18). Bubble plots and forest plots for the subgroup analyses for MCDI- words understood are provided in Figure 11.

Mullen Scales of Early Learning Expressive Language Domain (Mullen, 1985). For the analyses of the Mullen Expressive Language domain there were 59 effect sizes included. However, there were 4 cases that were identified as potential outliers and dropped, leaving 55 effect sizes. The average length of the measurement period was 10.7 months (SD=7.03) and the median length was 11 months (IQR=6.47). The expected change with the potential influential cases included was a SMD of 0.5608, 95% CI [0.4101; 0.7115] and without the potential influential cases this effect was reduced to a SMD of 0.4385, 95% CI [0.3264; 0.5505]. The I^2 was also reduced from 88.5% to 58.4%. Based on the conventional rule of thumb the SMD would be considered a medium effect size.

There was a significant effect of the length of the measurement period on the pooled effect where an increase in length of one month leads to an expected increase in the effect size by approximately .02 units (b=0.0189, SE=0.0077, p=.017).

As a result of too few cases, those who were receiving Medication based interventions were excluded from the sub-group analyses. In the uncontrolled sub-group analysis, the estimated pooled effect for the TAU group was larger than that of those receiving behavioral interventions (k= 28, SMD=0.5407, 95% CI [0.3371; 0.7444] and k= 26, SMD= 0.3317, 95% CI [0.2201; 0.4434], p=.06). However, this difference was accounted for by the length of the measurement period and was no longer significant when controlling for length (p=.49). There was a main effect of Type of Measure where age equivalent scores had the largest effects (k= 23, SMD= 0.5784, 95% CI [0.3986; 0.7582] compared to standard scores (k= 15, SMD= 0.2984, 95% CI [0.0981; 0.4987] and developmental quotients (k= 16, SMD= 0.3032, 95% CI [0.1643; 0.4422]. The interaction between Type of Measure and length approached significance where the effect of length was reduced for developmental quotients (b=-0.0280, SE=0.0154, p=.07) compared to age equivalent scores. The comparison of the interaction terms for standard scores and age equivalent scores was of comparable magnitude but the standard error was much larger (b=-0.269, SE=0.0434, p=.54).

The Eggers' test was not significant (intercept= -0.913, t=-0.948, p=.35) indicating that there was little indication of publication bias in Mullen Expressive Language domain. Bubble plots and forest plots for the subgroup analyses for the Mullen Expressive Language domain are provided in Figure 12.

Mullen Scales of Early Learning- Receptive Language Domain (Mullen, 1985). For the analyses of the Mullen Receptive Language domain there were 40 effect sizes included. However, there were three cases that were identified as potential outliers and dropped, leaving 37 effect sizes. The average length of the measurement period was 10.61 months (SD=6.54) and the median length was 11 months (IQR=6.47). The expected change with the potential influential cases included was a SMD of 0.6494, 95% CI [0.4631; 0.8358] and without the potential influential cases this effect was reduced to a SMD of 0.5183, 95%

CI [0.3812; 0.6554]. The I^2 was also reduced from 90.4% to 71.1%. Based on the conventional rule of thumb the SMD would be considered a medium effect size.

The effect of length of the measurement period was not significant indicating that the pooled effect was not expected to increase or decrease as length increases (b=.01, SE=.009, p=.19). Looking next at the type of outcome, standard scores had the lowest expected change over time (k= 7, SMD= 0.1940, 95% CI [-0.0184; 0.4063]) compared with both age equivalent scores (k= 13, SMD=0.6981, 95% CI [0.4809; 0.9154]) and developmental quotient scores (k=16, SMD= 0.5006, 95% CI [0.2805; 0.7207], p=.0005). There was also an effect of cognitive ability where those in the less cognitively able group had greater expected change over time (k= 24, SMD= 0.6064, 95% CI [0.4178; 0.7950] than the more cognitively able group (k= 6, SMD= 0.2375, 95% CI [0.1025; 0.3725], p=.0005). This affect approached significance when controlling for the length of the measurement period (b=-.3467, SE=.1802, p=.06).

The interaction between Type of Measure and length was significant where the effect of length was reduced for standard scores (b=-0.4762, SE=0.1720, p=.009) compared to age equivalent scores. The comparison of the interaction terms for standard scores and developmental quotient scores was not significant (b=-0.203, SE=0.1420, p=.1639).

The Eggers' test for this outcome provided no indication that there is publication bias the Mullen Receptive Language domain are provided in Figure 13.

Preschool Language Scales- Expressive Language (PLS; Zimmerman et al., 2011). Seventeen effect sizes were included in the analyses for PLS- Expressive Language domain scores; the average length of the measurement period was 7.43 months (SD=8.35) and the median length was 5.52 months (IQR=3.24). The expected change across this time period was a SMD of 0.2875, 95% CI [0.1503; 0.4247], however there was one influential case identified and with this case removed the SMD was 0.2394, 95% CI [0.1464; 0.3325] and the I^2 was reduced from 49.3% to approximately 0%. Based on the conventional rule of thumb the SMD would be considered a small effect.

There was an additional case that was removed from the analyses involving the length of the measurement period. This case had a length of approximately 48 months, 36 months longer than the next closest study. With this case dropped there was a significant effect of the length of the measurement period on the pooled effect size where the effect size was expected to increase by .04 SMD units for every additional month (b=0.04, SE=0.0067, p<.0001).

The Eggers' Test was significant indicating that there may be publication bias in the reported effect sizes for PLS- Expressive domain scores (intercept=-2.044, t=-2.886, p=.01). However, the funnel plot showed that the skew in the plot was driven by a larger number of studies with smaller effect rather than larger effects, it is not likely therefore that this skew can be attributed to publication bias. Bubble plots and forest plots for the subgroup analyses for the PLS- Expressive Language domain are provided in Figure 14.

Preschool Language Scales- Receptive Language Domain (PLS; Zimmerman et al., 2011). Thirteen effect sizes were included in the analyses of PLS- Receptive Language domain scores; the average length of the measurement period was 8.86 months (SD=9.13) and the median length was 6 months (IQR=8). The expected change across this time period was a SMD of 0.3748, 95% CI [0.1644; 0.5852], however there were two influential case identified and with these case removed the SMD was 0.2571, 95% CI [0.1329; 0.3812] and the I^2 was reduced from 74% to approximately 0%. Based on the conventional rule of thumb the SMD would be considered small.

There was a significant effect of the length of the measurement period on the pooled effect size where the effect size was expected to increase by approximately .04 SMD units for every additional month (b=0.036, SE=0.0094, p=.004).

The Eggers' Test was significant indicating that there may be publication bias in the reported effect sizes for PLS- Receptive Language domain scores (intercept=-2.261, t=-2.591, p=.03). However, as with the PLS- Expressive Language domain scores the funnel plot showed that the skew in the plot was driven by a single study with smaller effect rather than studies with larger effects, it is not likely therefore that this skew can be attributed to publication bias. Bubble plots and forest plots for the subgroup analyses for the PLS-Receptive Language domain are provided in Figure 15.

Peabody Picture Vocabulary Test (Dunn & Dunn, 1997). Sixteen effect sizes were included in the analyses of the PPVT; the average length of the measurement period was 9.1 months (SD=8.57) and the median length was 6.90 months (IQR=8.20). The expected change across this time period was a SMD of 0.1289, 95% CI [-0.0422; 0.3001] and the I^2 was approximately 0%. There were no influential cases or outliers identified. Based on the conventional rule of thumb the SMD would be considered very small.

There was a significant effect of length of the measurement period on the pooled effect size where for every one-month increase in length, the effect size is expected to change by approximately .04 units (b=0.036, SE=0.0094, p=.004). The effect size for those receiving behavioral interventions (k= 5, SMD= 0.3117, 95% CI [-0.0461; 0.6694] was larger than both the effect size for those receiving TAU (k= 6, SMD= 0.0464, 95% CI [-0.2646; 0.3574]) and Medication (k= 5, SMD= -0.0886, 95% CI [-0.5235; 0.3464]. However, this difference was non-significant do to the large confidence intervals of the estimates (p=.11). Bubble plots and forest plots for the subgroup analyses for the PPVT are provided in Figure 16.

An inspection of the funnel plot did not provide any evidence of publication bias.

Reynell Developmental Language Scales- Expressive Language (Reynell; Reynell & Gruber, 1997). For the analyses of the Reynell- Expressive Language domain there were 29 effect sizes included. However, there was one case that was identified as potential outliers and dropped, leaving 28 effect sizes. The average length of the measurement period was 8.11 months (SD=4.56) and the median length was 6 months (IQR=6). The expected change with the potential influential case included was a SMD of 0.5520, 95% CI [0.3789; 0.7251] and without the potential influential cases this effect was reduced to a SMD of 0.5107, 95% CI [0.3543; 0.6671] The I^2 was also reduced from 55.7% to 44.5%. Based on the conventional rule of thumb, a medium effect.

There was an additional case that was removed from the analyses involving the length of the measurement period. This case had a length of approximately 23 months, 11 months longer than the next closest study. With this case dropped there was a significant effect of the length of the measurement period on the pooled effect size where the effect size was expected to decrease by .05 SMD units for every additional month (b=-0.0527, SE=0.0189, p<.009). There was only one effect size in the high cognitive ability group and so it was not able to be modeled via the subgroup analyses or through meta-regression.

The Eggers' Test for this measure was non-significant (intercept= .786, t=.648, p=.52) indicating that publication bias is likely not present. Bubble plots and forest plots for the subgroup analyses for the Reynell- Expressive Language domain are provided in Figure 17.

Reynell Developmental Language Scales- Receptive Language (Reynell; Reynell & Gruber, 1997). For the analyses of the Reynell- Receptive Language domain there were 30 effect sizes included. However, there was one case that was identified as potential outliers and dropped, leaving 29 effect sizes. The average length of the measurement period was 8.04 months (SD=4.49) and the median length was 6 months (IQR=6). The expected change with the potential influential case included was a SMD of 0.4778, 95% CI [0.3415; 0.6141] and without the potential influential cases this effect was reduced to a SMD of 0.4319, 95% CI [0.3155; 0.5483]. The I^2 was also reduced from 36.5% to 10.1%. Based on the conventional rule of thumb the SMD would be considered small to medium.

There was an additional case that was removed for the analyses involving the length of the measurement period. This case had a length of approximately 23 months, 11 months longer than the next closest study. With this case dropped the effect of the length of the measurement period on the pooled effect size verged on significance (b=-0.026, SE=0.0148, p=.09). When controlling for type of intervention received the effect of length was significant (b=-0.348, SE=0.0144, p=.02).

There was only one effect size that was in the high cognitive ability group and so it was not able to be modeled via the subgroup analyses or through meta-regression. The Eggers' Test for this measure was non-significant (intercept= -.35, t=-.361, p=.72) indicating that publication bias was likely not present. Bubble plots and forest plots for the subgroup analyses for the Reynell- Receptive Language domain are provided in Figure 18.

Structured Play Assessment (Ungerer & Sigman, 1981). There were a total of 41 effect sizes pooled together for the analyses of the SPA, the average length of the measurement period was 3.31 months (SD=1.86) and the median length was 2.76 months (IQR=4.16). The estimated expected change over time was a SMD of 0.1940, 95% CI [0.1102; 0.2777]. The I^2 for this model was 7.9%. There were no effect sizes that were identified as an influential cases or outliers. Based on the conventional rule of thumb this is a small effect.

There was a significant difference in the estimated effect size based on study quality where those studies rated as "Fair" had larger effects (k= 21, SMD= 0.3046, 95% CI [0.1796; 0.4296]) than those rated as "Good" (k= 18, SMD= 0.1314, 95% CI [0.0140; 0.2489]). Study quality was included as a covariate in the remaining models for the SPA.

The effect of length of the measurement period was not-significant (b=-0.0085, SE=0.0232, p=.72). However, there was a significant interaction between cognitive ability and length of the measurement period where the effect of time was stronger in those who had higher cognitive ability compared to those who did not (b=0.7155, SE=0.2962, p=.02).

There was no difference in the size of the effect based on the type of play outcome coded from the SPA when comparing level of play, symbolic play and functional play (overall model significance p=.70).

The Eggers' Test approached significance (intercept=1.04, t=1.99, p=.054) however on inspection of the funnel plot the skew was driven by two cases with smaller effects. It is not likely that publication bias is present in SPA outcomes. Bubble plots and forest plots for the subgroup analyses for SPA scores are provided in Figure 19.

Social Responsiveness Scale- Total Score (SRS; Constantino & Gruber, 2012). There were a total of 77 effect sizes included in the analysis of SRS Total scores. The average length of the measurement period across these outcomes was 4.67 months (SD=3.71) and the median length was 3.68 months (IQR=3.24). The expected change over this period of time was a SMD of - 0.2623, 95% CI [-0.3418; -0.1829], however two influential cases were identified which, when removed, reduced the SMD to -0.2318, 95% CI [-0.2933; -0.1703] and the I^2 from 17.1% to approximately 0%. Based on the conventional rule of thumb the SMD would be considered small. For SRS Total scores and all domain scores a negative effect size represents a reduction in symptoms.

The effect of the length of the measurement period was not significant (b=0.00313, SE= 0.077, p=.87). TAU groups had a smaller effect (k=27, SMD= -0.1472, 95% CI [-0.2471; -0.0472]) compared to those receiving behavioral interventions (k= 37, SMD= -0.2553, 95% CI [-0.3494; -0.1612]) and those receiving Medication (k=11, SMD= -0.3484 [-0.4783; -0.2186]) and this difference was significant (p=.03). There was some signal that there was an interaction between type of intervention received and length of the measurement period where those in receiving behavioral interventions saw more gains as length increased but the contrasts with TAU (p=.13) and Medication (p=.09) were both not significant.

Eggers' test was not significant (intercept=-0.438, t=-1.804, p=.08) indicating that there was likely no publication bias in the sample. Bubble plots and forest plots for the subgroup analyses of SRS Total scores are provided in Figure 20.

Social Responsiveness Scale- Social Cognition (SRS; Constantino & Gruber, 2012). There were a total of 36 effect sizes included in the analysis of SRS Social Cognition scores. The average length of the measurement period across these outcomes was 4.71 months (SD=3.17) and the median length was 3 months (IQR=3.24). The expected change over this period of time was a SMD of -0.2055, 95% CI [-0.3259; -0.0851], however there was one influential case identified which, when removed, reduced the SMD to - 0.2321 [-0.3244; -0.1398] and the I^2 from 23.8% to approximately 0%. Based on the conventional rule of thumb the SMD would be considered small.

The effect of the length of the measurement period was not significant (b=-0.0036, SE=0.146, p=.80. However, there was a significant interaction between length and study quality where those studies rated as "Poor" had a weaker relationship between length and effect size than both studies rated as "Fair" (b=.1335, SE=.0529, p=.02) and studies rated as "Good" (b=0.1928, SE=0.0551, p=.0015).

Further, the interaction between type of intervention received and length of the measurement period was significant; those in the behavioral intervention group saw larger effect sizes as length increased (b=-0.1331, SE=0.0577, p=.03). Lastly, there was a main effect of cognitive ability where those with higher cognitive ability (k= 13, SMD= -0.1714, 95% CI [-0.2658; -0.0770] had less expected change than those with lower cognitive ability (k= 5, SMD= -0.3852, 95% CI [-0.4957; -0.2748]). This effect was consistent even when controlling for study quality.

Eggers' test was non-significant indicating that there was likely no publication bias present in SRS Social Cognition scores (intercept=-.507, t=-1.218, p=.23). Bubble plots and forest plots for the subgroup analyses for of SRS Social Cognition scores are provided in Figure 21.

Social Responsiveness Scale- Social Communication (SRS; Constantino & Gruber, 2012). There were a total of 38 effect sizes included in the analysis of SRS Social Communication scores. The average length of the measurement period across these outcomes was 5.58 months (SD=4.24) and the median length was 3.23 months (IQR=3.24). The expected change over this period of time was a SMD of -0.2523, 95% CI [-0.3214; -0.1831], and the I^2 of the model was approximately 0%. There were no outliers or influential cases identified. Based on the conventional rule of thumb a small effect.

The effect of the length of the measurement period was non-significant (b=.0010, SE=.0072, p=.89). However, the interaction between length and type of intervention received was significant; the medication group saw a greater reduction in SRS Social Communication scores as the length of the measurement period increased (b=-.02, SE=.01233, p=.005).

Lastly, the Egger's test was not significant (intercept= -0.544, t=-1.651, p=.11) indicating that there is likely no publication bias across this measure. Bubble plots and forest plots for the subgroup analyses of SRS Social Communication scores are provided in Figure 22.

Social Responsiveness Scale- Social Affect (SRS; Constantino & Gruber, 2012). There were a total of 38 effect sizes included in the analysis of the SRS Social Affect domain. The average length of the measurement period across these outcomes was 5.58 months (SD=4.24) and the median length was 3 months (IQR=3.24). The expected change over this period of time was a SMD of - 0.1648, 95% CI [-0.2578; -0.0718] and the I^2 of the model was approximately 0%. There were no outliers or influential cases identified. Based on the conventional rule of thumb the SMD would be considered small.

The effect of the length of the measurement period was not significant (b=0.0019, SE=0.0135, p=.89) however there was a significant main effect of type of intervention received (p= .02). Those receiving TAU had the smallest SMD (k=11, SMD= -0.0161, 95% CI [-0.1318; 0.0996] when compared to those receiving a behavioral intervention (k= 11, SMD= -0.2639, 95% CI [-0.4318; -0.0959] and those receiving Medication (k= 8, SMD= -0.1808, 95% CI [-0.4039; 0.0424].

The Eggers' test was not significant indicating that publication bias was not likely present in this measure (intercept= -0.048, t=-0.106, p=.92). Bubble plots and forest plots for the subgroup analyses of the SRS Social Affect domain are provided in Figure 23.

Social Responsiveness Scale- Social Motivation (SRS; Constantino & Gruber, 2012). There were a total of 38 effect sizes included in the analysis of the SRS Social Motivation domain. The average length of the measurement period across these outcomes was 5.12 months (SD=3.48) and the median length was 3 months (IQR=3.24). The expected change over this period of time was a SMD of -0.3123, 95% CI [-0.4195; -0.2050], however there was one influential case identified which, when removed, reduced the SMD to -0.2830, 95% CI [-0.3754; -0.1906] and the I^2 from 21.7% to 3.4%. Based on the conventional rule of thumb the SMD would be considered small.

The effect of length of the measurement period was non-significant (b=-.0057, SE=0.137, p=.68), however there was a significant interaction between length and study quality where those studies rated as "Poor" saw less change in effect size than studies rated as "Good" (b=.1510, SE=.0593, p=.02) and the difference between "Poor" and "Fair studies verged on significance in the same direction (b=.0992, SE=.0581, p=.09). There was also a significant interaction between length and and type of intervention received where those receiving Medication saw greater reductions in SRS Social Motivation scores as length increased compared to those receiving behavioral interventions (b=-.0660, SE=0.0311, p=.04). This interaction was no longer significant when controlling for study quality however (p=.59).

The Eggers' Test was significant indicating that there may be publication bias in this sample (intercept=-1.222, t=-2.967, p=.005). On inspection of the forest plot the skew was driven by two effect sizes that were extremely negative that had no comparably large positive effect sizes. This is to be expected given that it is unlikely for parents to report sharp increases in ASD symptomology over such a short time period. Bubble plots and forest plots for the subgroup analyses of the SRS Social Motivation domain are provided in Figure 24.

Vineland Adaptive Behaviors Scales- Communication (VABS; Sparrow et al., 1984). For the analyses of the VABS- Communication domain there were 110 effect sizes included. However, there was one case that was identified as potential outliers and dropped, leaving 109 effect sizes. The average length of the measurement period was 10.33 months (SD=8.81) and the median length was 9 months (IQR=6.48). The expected change with the potential influential case included was a SMD of 0.3670, 95% CI [0.2769; 0.4571] and without the potential influential case this effect was reduced to a SMD of 0.3445, 95% CI [0.2627; 0.4263]. The I^2 was also reduced from 68.6% to 56.1%. Based on the conventional rule of thumb the SMD would be considered small.

The effect of the length of time was not significant (b=.0005, SE=.0044, p=.901). The subgroup analysis comparing the types of intervention received approached significance (p=.069) where those receiving a behavioral intervention (k= 56, SMD= 0.4353, 95% CI [0.3451; 0.5255] had the largest effects, compared to both the TAU groups (k= 41, SMD=0.2472, 95% CI [0.1027; 0.3918] and Mediation groups (k=12, SMD= 0.2712, 95% CI [-0.1147; 0.6570]. However, this was followed up by a significant interaction between type of intervention received and length where those in the Medication groups saw more change as the length of the measurement period increased compared to the behavioral groups (b=.1528, SE=.0454, p=.0011). The contrast between the behavioral and TAU groups was not significant (b=-0.0026, SE=.0086, p=.77).

There was also a significant effect of type of measure where the effects sizes derived from standard scores had the smallest effect (k= 88, SMD= 0.2928, 95% CI [0.1990; 0.3866]) compared to both the age equivalent scores (k=14, SMD 0.5507, 95% CI [0.4171; 0.6843]) and raw scores (k= 5, SMD= 0.7120, 95% CI [0.2424; 1.1817]).

The Eggers' test was not significant indicating that there is likely no publication bias across this measure (intercept=0.253, t=.571, p=.57). Bubble plots and forest plots for the subgroup analyses of the VABS- Communication domain are provided in Figure 25.

Vineland Adaptive Behaviors Scales- Socialization (VABS; Sparrow et al., 1984). For the analyses of the VABS Socialization domain there were 94 effect sizes included. However, there was one case that was identified as a potential outlier and dropped, leaving 93 effect sizes. The average length of the measurement period was 9.96 months (SD=8.4) and the median length was 8.5 months (IQR=6.86). The expected change with the potential influential cases included was a SMD of 0.3581, 95% CI [0.2709; 0.4453] and without the potential influential cases this effect was increased to a SMD of 0.3635, 95% CI [0.2771; 0.4500]. There was little change in the I^2, decreasing from 60.7% to 60.5%. Based on the conventional rule of thumb the SMD would be considered a small effect.

The effect of the length of the study period was not significant (b=-0.0054, SE=0.0047, p=.26). The subgroup analysis comparing the types of intervention received was significant (p=.04) where those receiving a behavior intervention (k=45, SMD 0.4827, 95% CI [0.3641; 0.6013] had the largest effects, compared to both the TAU groups (k=33, SMD= 0.2645, 95% CI [0.1071; 0.4220]) and the Mediation groups (k=15, SMD 0.2707, 95% CI

[0.0723; 0.4691]). There was a trend in the interaction between type of treatment received and length of the measurement period where those receiving behavioral interventions saw more change as the length increased compared to those receiving TAU (b=-0.0167, SE=0.0094, p=.07).

The effect of type of measure approached significance where the effects sizes derived from standard scores had the smallest effect (k=69, SMD= 0.3122, 95% CI [0.2146; 0.4099]) compared to both the age equivalent scores (k= 13 0.6228 [0.3343; 0.9113]) and raw scores (k= 11, SMD= 0.4173, 95% CI [0.1684; 0.6663]). However, there was also a significant interaction between type of measure and length where raw scores increased more as the length of the measurement period increased compared to age equivalent scores (b=.0542, SE=0.0269, p=.04).

Lastly, there was a significant interaction between length and cognitive ability where those in the higher cognitive ability group increased less as the length of the measurement period increased compared to the lower cognitive ability group (b=-0.0284, SE=0.0119, p=.02).

The Egger's test was not significant (intercept=0.639, t=1.249, p=.22) indicating that there is likely no publication bias across this measure. Bubble plots and forest plots for the subgroup analyses of the VABS Socialization domain are provided in Figure 26.

Less Common Social Communication Measures

Bubble plots for the outcomes described below are provided in Appendix B.

Brief Observation of Social Communication Change (BOSCC; Grzadzinski et al., 2016; Kim et al., 2019). While sufficient data was present to estimate a pooled effect size for both the Social Communication domain and overall BOSCC scores because published trials have used multiple versions of the scoring system and active refinement of the measure is ongoing the decision was made to not analyze data for this outcome.

Communication and Symbolic Behavior Scales- Social Communication (CSBS; Wetherby & Prizant, 2002). There were eight effect sizes included in the analyses of the CSBS Social Communication domain. The mean length of the measurement period was 4.25 months (SD=2.43) and the median was 3.5 months (IQR= 2.25). The average expected change for the CSBS Social Communication was a SMD of 0.3681, 95% CI [0.1449; 0.5912], there were no influential cases that were identified and the I^2 for this model was 1.3%. An Eggers' Test was not run due to the small number of cases however a visual inspection of the funnel plot showed little evidence of publication bias. Based on the conventional rule of thumb the SMD would be considered small. Bubble plots for the CSBS- Social Communication scores are provided in Figure 27.

Expressive One Word Picture Vocabulary Test (EOWPVT; Gardner, 1990). There were ten effect sizes included in the analyses and the average length of the measurement period was 4.43 months (SD=1.62) with a median length of 4.14 months (IQR=1.86). The expected change over this period was a SMD of 0.1159, 95% CI [0.0627; 0.1691] and the I^2 of the final model was approximately 0%. There were no outliers or influential cases identified so all data was retained. Based on the conventional rule of thumb the SMD would be considered very small.

The effect of length of the measurement period on the pooled effect was non-significant (b=.0071, SE=0.0165, p=.6801), meaning that on average the effect size is not expected to increase as the length of the measurement period increases. An Eggers' Test was not run due to the small number of cases however a visual inspection of the funnel plot showed little evidence of publication bias. Bubble plots for the CSBS- Social Communication scores are provided in Figure 27.

Functional Emotional Assessment Scale (FEAS; Greenspan et al., 2001). There were six effect sizes included in the pooled effect size estimate for the FEAS. The average length of the measurement period was 7.49 months (SD=4.93) and the median length was 7.5 months (IQR=9). The expected change over this time was a SMD of 0.6613, 95% CI [0.2337; 1.0890] and the I^2 was 57.1%. One of the effect sizes was identified as a potential influential case and without this estimate the expected change went up to a SMD of 0.7970, 95% CI [0.3666; 1.2273] and an I^2 of 11%. This outlier that was pulling down the SMD was a TAU arm of a randomized controlled trial. Based on the conventional rule of thumb the SMD would be considered large. There were not enough cases to use an Eggers' test however a visual inspection of the funnel plot did not indicate the presence of publication bias.

Griffiths Mental Development Scales- Hearing and Language (GMDS; Griffiths, 1970). There were seven effect sizes included in the pooled effect size estimate for the GMDS- Hearing and Language domain. The average length of the measurement period was 20.56 months (SD=18.17), the median length was 6 months (IQR=34.0) and over this period the expected change was a SMD of 0.1774, 95% CI [-0.3104; 0.6652]. There were two potential outliers identified and with these cases the SMD was 0.1388, 95% CI [-0.1856; 0.4632] but the I^2 was reduced from 73.8% to 35.1%. Based on the conventional rule of thumb the SMD would be considered very small. A visual inspection of the funnel plot did not indicate the presence of publication bias.

Griffiths Mental Development Scales- Personal Social (GMDS; Griffiths, 1970). There were seven effect sizes included in the pooled effect size estimate for the GMDS- Personal Social domain. The average length of the measurement period was 20.56 months (SD=18.17), the median length was 6 months (IQR=34.0) and the expected change was a SMD of -0.0161, 95% CI [-0.3600; 0.3278]. The I^2 was 58.2% and no outliers or influential cases were identified. Based on the conventional rule of thumb the SMD would be considered negligible. A visual inspection of the funnel plot did not indicate the presence of publication bias.

MacArthur-Bates Communicative Development Inventories- Gestures Produced (MCDI; Fenson, 2007). There were six effect sizes pooled for the analyses of the MCDI-Gestures Produced scores. The average length of the measurement period was 9.3 months (SD=6.63) and the median length was 12 months (IQR=10.56). There were no clear outliers or influential effect sizes. The expected change over the measurement period was a SMD of 0.5181, 95% CI [0.3157; 0.7205] and the I^2 was approximately 0%. Based on the conventional rule of thumb the SMD would be considered a medium effect. A visual inspection of the funnel plot showed there was no indication that publication bias was present in this measure.

Mullen Scales of Early Learning- Receptive and Expressive Combined (Mullen, 1985). There were six effect sizes pooled together for the analysis of the combined receptive and expressive domains of the Mullen Scales of Early Learning. The average length of the measurement period was 16.66 months (SD=5.89) and the median length was 14 months (IQR=10). The expected change over this period was a SMD of 0.5903, 95% CI [0.1115; 1.0692]. One influential case was identified and without this effect size the SMD became 0.7362, 95% CI [0.2972; 1.1752] with an I^2 of 48.5%. Based on the conventional rule of thumb the size of the SMD would be considered medium. There was no indication of publication bias based on a visual inspection of the funnel plot.

Psychoeducational Profile-Revised-Verbal Cognitive Domain (PEP-R; Schopler et al., 1990). There were five effect sizes available for the analysis of the PEP-R; the average length of the measurement period was 8.4 months (SD=3.29) and the median length was 6 months (IQR=6). There were no outliers or influential cases identified, the expected change over the measurement period was a SMD of 0.3860, 95% CI [-0.0728; 0.8448] and the I^2 for the model was 17.2%. Based on the conventional rule of thumb the SMD would be considered a small to medium effect. There was no indication of publication bias based on an inspection of the funnel plot.

Preschool Language Scales- Receptive and Expressive Combined (PLS; Zimmerman et al., 2011). There were nine effect sizes included in the analyses of this measure with an average length of the measurement period of 8.49 months (SD=5.27) and a median length of 6 months (IQR=5). The average expected change over this time period was a SMD of 0.5793, 95% CI [-0.0805; 1.2391], however there were three influential cases identified and with these cases removed the expected change was reduced to 0.2998, 95% CI [0.1751; 0.4245] and the I^2 was approximately 0%. Based on the conventional rule of thumb the SMD would be considered a small effect. There was no indication of publication bias based on an inspection of the funnel plot.

Parent Interview for Autism Social Reciprocity Domain (PIA; Stone & Hogan, 1993). There were six effect sizes included in the final analysis for the PIA- Social Reciprocity domain and one was identified as potential influential case. The average length of the measurement period was 2.83 months (SD=2.84) and the median length was one month (IQR=4.125). The overall expected change for this measure was a SMD of 0.1522, 95% CI [-0.2387; 0.5431] but with the one extremely negative effect size that was identified as potential influential case removed the expected change was increased to a SMD of 0.2717, 95% CI [-0.0744; 0.6179] and the I^2 for this final model was 39.3%. Based on the conventional rule of thumb the SMD would be considered small to medium. There was no indication of publication bias based on an inspection of the funnel plot.

Parent Interview for Autism Non-verbal Communication Domain (PIA; Stone & Hogan, 1993). There were six effect sizes included in the final analysis for the PIA-NV Communication domain and one was identified as a potential influential case. The average length of the measurement period was 2.83 months (SD=2.84) and the median length was 1 month (IQR=4.125). The overall expected change for this measure was a SMD of 0.2921, 95% CI [-0.1139; 0.6982] however with the one extremely negative effect size that was identified as potential influential case removed the expected change was increased to a SMD of 0.4188, 95% CI [0.1195; 0.7180] and the I^2 was 13.8%. Based on the conventional rule of thumb the SMD would be considered a small to medium effect. There was no indication of publication bias based on an inspection of the funnel plot.

Social Communication Questionnaire (SCQ; Rutter et al., 2003). There were six effect sizes included in the final analysis for the SCQ and one was identified as a potential influential case. The average length of the measurement period for this measure was 14.13 months (SD=11.84) and the median length was 12 months (IQR=1.71). The expected change with all effect sizes included was a SMD of -0.6043, 95% CI [-0.9618; -0.2469] and with the one potential influential case removed the expected change was reduced to a SMD of -0.5164, 95%, [-0.7491; -0.2837]. Based on the conventional rule of thumb the SMD would be considered a medium effect. The I^2 for the model was reduced from 37% to approximately 0% with the influential case removed. There was no indication of publication bias based on an inspection of the funnel plot.

Social Skills Improvement System (SSIS; Gresham & Elliot, 2007). There were 8 effect sizes included in the final analysis for the SSIS and one was identified as a potential influential case. The average length of the measurement period for this measure was 3.53 months (SD=1.24) and the median length was 2.64 months (IQR=2.18). The expected change with all effect sizes included was a SMD of 0.6354, 95% CI [-0.0591; 1.3300] and with the one potential influential case removed the expected change was reduced to a SMD of 0.4015, 95% CI [0.1416; 0.6615]. Based on the conventional rule of thumb the SMD would be considered small to medium. The I^2 for the model was reduced from 51% to approximately

0% with the influential case removed. There was no indication of publication bias based on an inspection of the funnel plot.

Vineland Adaptive Behaviors Scales- Expressive Language (VABS; Sparrow et al., 1984). Ten effect sizes were included in the analyses for VABS- Expressive Language domain; the average length of the measurement period was 9.2 months (SD=9.55) and the median length was 6 months (IQR=7.74). The expected change across this time period was a SMD of 0.4003, 95% CI [0.1879; 0.6127], however there was one influential case identified and with this case removed the SMD was 0.3383, 95% CI [0.1937; 0.4830] and the I^2 was reduced from 15% to approximately 0%. Based on the conventional rule of thumb the SMD would be considered small.

The effect of the length of the measurement period on the pooled effect size was non-significant (b=-0.0018, SE=0.0106, p=.87). There were not enough cases to estimate the sub-group effects based on cognitive ability. Lastly, there was no indication of publication bias based on an inspection of the funnel plot for the VABS- Expressive Language domain.

Vineland Adaptive Behaviors Scales- Receptive Language (VABS; Sparrow et al., 1984). Ten effect sizes were included in the analyses for VABS- Receptive Language domain; the average length of the measurement period was 9.2 months (SD=9.55) and the median length was 6 months (IQR=7.74). The expected change across this time period was a SMD 0.3401, 95% CI [0.2282; 0.4520] and the I^2 was approximately 0%. Based on the conventional rule of thumb a small effect. There were no outliers or influential cases identified.

There was a difference in the size of the pooled effect based on study quality where “Fair” quality had the largest effect (SMD=0.79), “Poor” quality had the lowest (SMD=.09) and “Good” quality was in the middle (SMD=.34). However, the “Fair” group was based off of a single case and the "Poor" group was based off only two cases so these differences should be interpreted with caution.

The effect of study length was non-significant (b=.0038, SE=.0052, p=.49) and the effect of the Eggers' Test was non-significant, indicating that publication bias is likely not present in this outcome (intercept= -.497, t=.931, p=.38).

Discussion

Issues surrounding outcome measurement have been central to the progress and agenda of autism research since it began to burgeon into maturity nearly 20 years ago. Multiple research and working groups have identified the importance of issues surrounding outcome measurement including: the need for valid and reliable tests, the need for careful selection of tests by researchers based on the construct that one expects to change, whether general measures of development are appropriate to track progress across complex behavioral phenomena and whether change within specific tests are influenced by factors like cognitive ability, age or gender that are not the construct of interest (Charman et al., 2003; Kasari, 2002; Lord et al., 2005).

Despite the early recognition of these measurement issues, the nature of the development and rapid proliferation of interventions for young children with autism lead simultaneously to the rapid proliferation of tests to evaluate those interventions. Recent reviews estimated that from 131 to 289 different outcome measures have been used in evaluation studies of young children with ASD (Bolte & Diehl, 2013; McConachie et al., 2015). This, corroborated by the current study which identified 119 unique measures of social communication outcomes alone, points to a serious issue in the approach taken to measure progress and the reliability of the current evidence base.

Focusing on social communication outcomes, despite the large variety of available outcome measures very few of those measures have undergone rigorous psychometric evaluation (McConachie et al., 2015) and fewer than 10 have been deemed appropriate for use as a measure of treatment response (Anagostou et al., 2015). Further, as noted by others, many of those measures that have been deemed appropriate were not intended or created to measure or track autism specific behaviors or development in individuals with autism (Grzadzinski et al., 2020). The overarching aim of this article was to provide a systematic appraisal of two specific psychometric properties of social communication measures used in autism research: the magnitude of expected change and sensitivity to change over time.

Magnitude of Expected Change

These analyses examined the expected change over time of 40 social communication outcome measures with sufficient data to estimate a pooled effect size and was inclusive of 1099 effect sizes across those measures. The overall pooled effect within each measure represents the expected change over time over the length of the measurement period. Using conventional metrics, five measures had negligible to very small effect sizes, seventeen of the measures had a small effect size, eight had a small to medium effect size, nine had medium effect sizes and one had a large effect size (See Table 3.).

As was hypothesized, these data indicate the magnitude of the expected change is variable with the majority of the standardized mean differences across the measures ranging from small to medium. While one of the strengths of using standardized metrics of change is the ability to compare and contrast effects across measures on different scales, one drawback is the loss of substantive information about what a specific effect size means within each measure. It is important to consider the substantive meaning of observed effect sizes. Small change in one measure may be more impactful on a child or family's quality of life compared to large change on another. Whether the observed changes across measures is related to substantive changes in the quality of life of individuals with ASD remains an important empirical question that deserves more explicit attention (Tavernor et al., 2013). One avenue for this research may be linking progress in objective outcomes, like those included in this study, to more holistic, subjective or qualitative measures of progress. This would allow for the study of the impact of developmental progress on the everyday functioning of the child or family.

Sensitivity to Change

The inclusion of the length of the measurement period as a predictor of effect size provided a metric to examine the degree to which the included measures were sensitive to developmental change over time. In general, one would expect scores from measures that are more sensitive to change to increase more as the length of the measurement period increased. While it was hypothesized that most measures would be sensitive to change over time, there were only nine measures whose effect sizes were related to the length of the measurement period: ADOS Severity, ADOS Language Domain, ADOS SI Domain, MCDI Words Understood, Mullen Expressive Language, PLS Expressive and Receptive Domains, PPVT, Reynell Expressive Language. The magnitude of the effect of length varied from any increase of .006 SMD units to an increase of .06 SMD for each additional month. Interestingly for both ADOS Severity and Reynell Expressive Language scores the direction of the effect was the opposite of what was expected with less change expected as time increased. For standard scores this effect could be explained by children failing to keep up with norms rather than indicating that children were not making gains over time.

Overall, these data indicate that most of the commonly used social communication measures, on average, may not be sensitive to incremental developmental progress over time or that they are measuring stable developmental constructs such as severity of autism symptomology.

Rather than disqualifying these tests as appropriate outcomes to measure change, these analyses should serve as an impetus for further detailed psychometric evaluations of participant level data. One framework for such evaluations in health outcomes research is through the Item Response Theory framework (IRT; Edelen & Reeve, 2007). Recent studies have found success in refining or evaluating commonly used measures including the SRS (Sturm et al., 2017), the SCQ (Wei et al., 2015), ADOS (Kuhfield & Sturm, 2018) and SSIS (Anthony et al., 2016) through this framework. Of particular note to the question of the sensitivity and expected change of these measures may be issues concerning item difficulty. There is an assumption that progress at all points of the tests is equally easy (e.g. making progress from item 5 to 8 on the Mullen Expressive Language domain is equally as easy as making progress from item 15-18) although this assumption remains largely untested and may prove to be an important direction for future research. The implication of having inadequate information about the discrimination and difficulty of items at different points of a test is that comparing information (specifically change) of two individuals starting at different points on the scale becomes problematic. As an example, are children with lower cognitive ability actually making less progress in intensive interventions or is it more difficult to progress through early items on the test being used to track progress? Further rigorous psychometric work is needed to answer this and other clinically important questions.

Intervention Effects

The overall effect size estimates in prior meta-analyses have been primarily based on the contrasts between treatment and control groups at the end of the study period, though some have used gain scores within groups. When evaluating social communication outcomes studies have reported small to medium effect sizes: g=.355 (Fuller & Kaiser, 2020), SMD=.14 to .90 (Oono et al., 2013), standardized gain score=.32 (Nahmias et al., 2019), SMD=.27 (Sandbank et al., 2020). Comparable information in this study was obtained through the sub-group analyses based on the type of intervention received during the measurement period. Overall, eight of the included measures showed a clear indication of a differential effect based on the type of intervention received. Four of these measures were parent reports (SRS- Total Score, SRS- Social Cognition, SRS- Social Awareness and the CARS), two were parent interviews (Vineland Communication and Socialization domains), one was ADOS severity scores and the last the PPVT. Those receiving identifiable behavioral interventions on average saw larger effects across these measures and so it may be that the aforementioned measures are particularly sensitive to detecting progress in behavioral interventions. Given the majority were parent reports, it could also be that parents are in tune with subtle changes in their child over time or, as some have suggested, are biased reporters based on their proximity to the typical behavioral intervention context or their involvement in the intervention itself in the case of parent mediated interventions (Crank et al., 2021). With these potential limitations in mind, it is likely beneficial to supplement the use of parent report measures with clinician administered measures to gain a more holistic view of progress.

Overall, however these results indicate that these outcomes (the VABS in particular) should be considered as a candidate for additional psychometric validation as an outcome in clinical trials (Anagnostou et al., 2015). Though future evaluations of the VABS are still warranted to evaluate the influence of additional characteristics such as language level, whether the sample includes minimally verbal children and comorbidity. It should also be noted that it likely is tracking more global changes rather than subtle changes in social communication that are the target of many early interventions.

Chronological Age

A recent meta-analysis found no relationship between chronological age and treatment effects (Sandbank et al., 2020; Crank et al., 2021). In the current study, older age was related to a decrease in the expected change of ADOS Language and Reciprocal Social Interaction scores, ESCS RJA and requesting scores and MCDI Words Produced and

Understood but in the majority of measures there was no relationship between age and the size of the effect. In line with previous suggestions, earlier age at intervention onset appears to lead to more change but only on some social communication measures (Zwaigenbaum et al., 2015).

Cognitive Ability

Overall, there was inconsistent reporting of participants' cognitive ability across the included studies and overall cognitive ability, as operationalized in this study did not influence the magnitude of expected change over time across most measures. There were a few exceptions. However, as with many of the other variables, the direction of the effects was not consistent. Higher cognitive ability was associated with larger effect sizes in the ADOS Language and SI domain scores and lower cognitive ability was associated with larger effect sizes on the Mullen Receptive Language domain, SRS Social Cognition Domain and Vineland Socialization domain scores. For ESCS IJA and SPA scores the effect of length of the measurement period was moderated by cognitive ability where higher cognitive ability was associated with more change over time.

These results are consistent with past studies showing that higher cognitive ability has been associated with more progress in receptive language (Ben-Itzchak et al., 2014), that verbal IQ is related to stability of ADOS scores over time (Gotham et al., 2012) and that higher IQ is associated with decreasing autism severity (Pender et al., 2020). The operationalization and crude coding of cognitive ability at the study level (i.e. the coding was capturing average cognitive ability across the sample rather than at the participant level) is likely one reason for the lack of relationship observed across the measures. Further, participant level, analyses of cognitive ability on these outcomes would provide more nuanced information and would be a useful direction for future research.

Types of Measures

These data corroborate previous investigations, which have identified that standard scores are generally less sensitive to change over time than raw or age equivalent scores (Carter et al., 1998; Willliams et al., 2006). Across both the Vineland Communication and Socialization domains and the Mullen Receptive and Expressive domains the smallest effect sizes were seen across standard scores and there was some indication that raw and age equivalent scores change the most as the length of the measurement period increased (a potential indication of sensitivity to change). The use of standard scores may attenuate observed change when used as an outcome measure and so it may be helpful to report both standard and age equivalent scores when possible.

Bias and Measurement Context

While they were not included as variables in these analyses a particular focus in recent synthesizes has been placed on the ideas of the "proximity" and "boundedness" of outcome measures. Proximity is related to the intervention received and is characterized by how similar the construct being measured is to the behavior or skill being taught (the ESCS measuring progress in a joint attention intervention would be considered highly proximal; Sandbank et al., 2020). Boundedness refers to whether the assessment context is different from the learning context (a play based intervention assessed via a play based interaction would be considered highly context bound; Sandbank et al., 2020). A nuanced consideration of these characteristics, an understanding of the mechanism of the intervention and a thorough understanding of the developmental construct being targeted are all important in the selection of outcomes in clinical trials. What construct is the intervention explicitly targeting? How does that construct manifest in children's behavior? Is the construct being targeted the end goal of the intervention or are downstream changes expected? These, among many other questions, are important and have been explicitly discussed in only a few recent clinical trials (Pickles et al., 2015; Shih et al., 2021).

In relation to proximity and boundedness, there is an assumption that more proximal and bounded measures will lead to larger effects, however based on these analyses and other recently published papers the data do not fully support this assumption (Crank et al., 2021) Equally important seems to be the characteristics of the tests themselves, in addition to the relationship of the tests to the measurement context (i.e. their proximity and boundedness). As an example many of the standardized measures such as the Mullen Expressive and Receptive Language domain scores and Reynell Expressive Language domain scores, which are unlikely to be highly bound or subject to biases such as "trainer related correlated measurement error" (Crank et al., 2021), had relatively large expected change over time (SMD of .4385, .5181 and .5107 respectively). Further work appears needed to evaluate the constructs of proximity and boundedness to reconcile these inconsistencies.

Quality of Included Studies

The relationship between the methodological quality of studies and the magnitude of expected change was inconsistent. In five of the outcomes (MCDI Words Understood, SPA, SRS Social Cognition, SRS Social Motivation and Vineland Receptive Language) quality was related to the expected change. However, the effect of quality both inflated and deflated estimated effects across the measures; studies rated as "Fair" and "Poor" had larger effects on the MCDI Words Understood and SPA but had smaller effects within the SRS Social Cognition and Social Motivation domains and Vineland Receptive Language domain.

The influence of the methodological quality on outcomes and the conclusions drawn from those outcome has been a theoretical concern in autism research for many years (Smith et al., 2006). Quality indicators have been included in recent large scale meta analyses of parent mediated interventions (Oono et al., 2013) as well as interventions more broadly (Sandbank et al., 2020; Crank et al., 2021). In some cases, quality indicators have been unrelated to effect size estimates (Fuller et al., 2020) and in others they have been (Crank et al., 2021). It is important to note that the meta-regression analyses of quality indicators across each of these studies, including this project, are correlational in nature. While it is important to review and report on study quality, care should be taken in not drawing firm causal conclusions based on these analyses, especially when looking at specific quality indicators (e.g. selection bias) as opposed to holistic quality ratings.

Further, many of the quality rating scales used in meta-analyses themselves are not validated or reliable (Conn & Rantz, 2003). Therefore, the inconsistencies noted above seem to reinforce the use of quality ratings as a metric for progress in the methodology of clinical trials rather than as a moderator of specific effects.

Limitations and Future Directions

One of the major limitations of these data, as is true of any meta analysis, is the variable methodological quality of the studies from which the data were extracted. While study quality was included as a covariate and small sample studies are weighted in the pooled estimates, concern of undue influence based on study quality remain.

Second, complete and adequate data was not accessible from a number of studies (e.g. means and standard deviations for the outcomes of interest) and so they could not be included in this meta analysis. It is possible that the observed relationships or patterns of findings could shift with the inclusion of these studies.

Third, while the analyses of effects within measures (rather than pooled across different measures) is a strength of the study it did limit the power to analyze measures that were less frequently present in the included studies. As a result, less definitive conclusions could be drawn about some measures.

Lastly, meta-regression and sub-group analyses using the average of participant level characteristics (i.e. average age of sample) is a crude metric, especially if there is variability within the sample within that characteristic (i.e. a wide age range). The results of these analyses should be considered exploratory and used as a guide for more in depth participant level analyses in the future across these measures.

Conclusions

One of the strengths of this manuscript is the focus on specific measures as the unit of analysis rather than pooling effect sizes across measures that may or may not be measuring comparable constructs. Increased clarity and providing an explicit rationale for the selection of specific tests has been recommended as an important step in elucidating potential mechanisms or active ingredients of early interventions (Grzadzinski et al., 2020). These analyses are an initial step towards this goal.

The sheer amount of tests available to measure social communication outcomes in young children with ASD can be overwhelming. These data seem to point to a siloed approach to the choice of appropriate outcome measures, with research groups rarely providing an explicit rationale for their selection of measures. A number of measures including the Mullen language subscales, the Vineland Adaptive Behavior Scales, Social Responsiveness Scales, the MCDI and ESCS were used fairly frequently.

Although the effect sizes were small, from the most commonly used measures, the VABS Socialization and Communication domain scores and SRS Total in particular seemed to differentiate between behavioral and TAU groups well and the expected change for Mullen domain scores was quite large, particularly when using age equivalent rather than standard scores. The increased use of these measures is likely driven by submission requirements for large data repositories like the National Database for Autism Research and Simon Simplex Complex and seems to represent a positive step forward for the field.

This study extends on previous reviews (Bolte & Diehl, 2013; McConnachie et al., 2015) in an effort to better understand the breadth of social communication outcomes but also to specifically evaluate and provide a descriptive overview of two important psychometric characteristics of those measures; their expected change and sensitivity to change over time. Efforts to validate and refine recently developed measures including the Brief Observation of Social Communication Change (BOSCC; Grzadzinski et al., 2016), the Eliciting Language Samples for Analysis (ELSA; Barkova et al., 2021) and the Autism Impact Measure (AIM; Mazurek et al., 2020) remain important goals for the autism research agenda.

Tables and Figures

Table 1 *Extracted Variables and Operationalization*

Variable	Definition
Citation	The full APA citation for the included study.
Type of Study	Whether the study was a dissertation or peer reviewed article.
Name of Intervention Groups	If applicable, the name of the intervention package or type of intervention that the child is receiving (if any).
Intervention Dose	The number of hours per week that the children are receiving the interventions described (if any).
Sample Size	The total sample size of each of the groups.
Diagnostic Procedure	The method of diagnosis for the children enrolled in the study.
Sample Age	The mean age of the children in the sample.
Gender	The percentage of male subjects in the sample.
Ethnicity	The percentage of each ethnic group in the sample.
Cognitive Ability	The name of the measure used to measure cognitive ability in the sample (if any). This could included verbal IQ, non-verbal IQ or subdomains of cognitive measures (e.g. Visual Reception domain of the Mullen Scales of Early Learning).
Cognitive Ability Scores	The mean and standard deviation of the cognitive ability score provided.
SC Measure Name	The name of the specific measure that was used to measure social communication in the trial.
SC Measure Score	The reported score for each of the measures that was used.
Length of Study	The total amount of time between each measurement period.
Study Quality	Whether the study met the criteria for each of the 13 study quality questions.

Table 2 *Descriptives Information of Included Studies*

	Continuous Outcomes			
	Mean (SD)	Median	Range	
Sample Size	34.26(29.20)	26	4-421	
Chronological Age	49.97 (10.58)	49 months	21- 99.6	
Study Year	2013 (5.31)	2015	1990-2020	
Length of Measurement Period	7.82(8.94)	6 months	0.23- 48 months	
	Categorical Outcomes			
Type of Article	192 Peer Reviewed	13 Dissertations		
Diagnostic Instrument	61 Record Review	14 Only ADI-R	65 only ADOS	65 Both ADI-R and ADOS

Table 3 *Pooled Effect Size Across Included Studies*

Measure Name	Number of Effect Sizes	Average Length of Measurement Period	SD Length of Measurement Period	Overall Random Effect	Lower CI	Upper CI	Adjusted Random Effect	Adjusted Lower CI	Adjusted Upper CI
ADOS Severity	47	16.79	13.48	0.114	0.2203	0.0078	0.1146	0.2179	0.0113
ADOS Social Affect	38	10.68	8.37	0.3243	0.4441	0.2044	0.3772	0.4725	0.2819
ADOS Language	20	7.21	4.95	0.6155	0.8565	0.3746	0.6624	0.8995	0.4254
ADOS SI	18	6.57	4.8	0.4904	0.7404	0.2404	0.6262	0.8381	0.4142
ADOS Total	20								
ATEC Language	4								
ATEC Social	4								
BASC Social	2								
BASC FC	2								
BOSCC SA	9								
BOSCC Total	12								
CARS	33	9.98	9.95	0.4932	0.6853	0.301	0.4519	0.6022	0.3017
CBRS Initiations	3								
CBRS Attention	3								
CSBS SC	8	4.25	2.43	0.3681	0.14449	0.5912	0.3681	0.14449	0.5912
CSBS	4	2.41	2.42						
EOWPVT	10	4.43	1.62	0.1159	0.0627	0.1691	0.1159	0.0627	0.1691
ESCS Gestures	18	1.61	0.34						
ESCS IJA	42	5.77	4.29	0.1936	0.0866	0.3005	0.1936	0.0866	0.3005
ESCS RJA	19	6.92	4.67	0.1608	0.0121	0.3095	0.1209	-0.0052	0.2471
ESCS Requesting	18	6.9	4.27	0.3154	0.0618	0.569	0.2439	0.0133	0.4745
FEAS	6	7.49	4.93	0.6613	0.2337	1.089	0.797	0.3666	1.2273
FEAS Questionnaire	3								

GARS	3								
GMDS Language	7	20.56	18.17	0.1774	-0.31	0.66	0.1388	-0.1856	0.4632
Griffiths Social	7	20.56	18.17	0.0161	-0.36	0.3278	-0.0161	-0.36	0.3278
MCDI Comprehension	33	7.55	5.012	0.4027	0.2842	0.5213	0.3721	0.258	0.4862
MCDI Expressive	49	6.72	4.68	0.4112	0.3102	0.5121	0.4153	0.3129	0.5177
MCDI Gestures	6	9.3	6.63	0.5181	0.3157	0.7205	0.5181	0.3157	0.7205
MCDI MLU	4								
Mullen Expressive	59	10.7	7.03	0.5608	0.4101	0.7115	0.4385	0.3264	0.5505
Mullen Receptive	40	10.61	6.54	0.6494	0.463	0.8358	0.5183	0.3812	0.6554
Mullen Verbal Combined	6	16.66	5.89	0.5903	0.1115	1.0692	0.7362	0.2972	1.175
PEP-R Expressive	3								
PEP-R Receptive	3								
PEP-R Verbal Cognitive	5	8.4	3.29	0.387	-0.0728	0.8448	0.387	-0.0728	0.8448
PJAM IJA	6								
PJAM RJA	6								
PJAM Turn Taking	6								
PLS Expressive	17	7.43	8.35	0.2875	0.1503	0.4247	0.2394	0.1464	0.3325
PLS Receptive	13	8.86	9.13	0.3748	0.1644	0.5822	0.2571	0.1329	0.3812
PLS Total	9	8.49	5.27	0.5793	-0.0805	1.2391	0.50974	0.1064	0.9085
PIA Social Reciprocity	6	2.83	2.84	0.1522	-0.2387	0.5431	0.1522	-0.2387	0.5431
PIA NV Communication	6	2.83	2.84	0.2921	-0.1139	0.6982	0.2921	-0.1139	0.6982
PPVT	16	9.1	8.57	0.1289	-0.0422	0.3001	0.1289	-0.0422	0.3001
Reynell Total	3								
Reynell Expressive	29	8.11	4.56	0.552	0.3789	0.7251	0.5107	0.3543	0.6671

Reynell Receptive	30	8.04	4.49	0.4778	0.3415	0.6141	0.4319	0.3155	0.5483
SCQ	6	14.13	11.84	0.6043	0.9618	0.2469	0.5164	0.7491	0.2837
SPA	41	3.31	1.86	0.194	0.1102	0.2777	0.194	0.1102	0.2777
SPACE	11								
SRS Total	77	4.67	3.71	0.2623	0.3418	0.1829	0.2318	0.2933	0.1703
SRS Cognition	36	4.71	3.17	0.2055	0.3259	0.0851	0.2321	0.3244	0.1398
SRS SA	30	4.76	3.34	0.1648	0.2578	0.0718	0.1648	0.2578	0.0718
SRS SC	38	5.58	4.24	0.2523	0.3214	0.1831	0.2523	0.3214	0.1831
SRS SM	38	5.12	3.48	0.3123	0.4195	0.205	0.283	0.3754	0.1906
SSIS	8	3.53	1.24	0.6354	-0.0591	1.33	0.4015	0.1416	0.6615
Symbolic Play Test	4								
Vineland Communication	110	10.33	8.81	0.367	0.2769	0.4571	0.3445	0.2627	0.4263
Vineland Expressive	10	9.2	9.55	0.4003	0.1869	0.6127	0.3383	0.1937	0.483
Vineland Receptive	10	9.2	9.55	0.3401	0.2282	0.452	0.3401	0.2282	0.452
Vineland Socialization	94	9.96	8.4	0.3581	0.2709	0.4453	0.3635	0.2771	0.45

Note. ABAS= Adaptive Behavior Assessment System, ADOS= Autism Diagnostic Observation Schedule, ATEC= Autism Treatment Evaluation Checklist, BASC= Behavior Assessment System for Children, BOSCC= Brief Observation of Social Communication Change, CARS= Childhood Autism Rating Scales, CBRS= Conner's Comprehensive Behavior Rating Scales, CSBS= Communication and Symbolic Behavior Scale, EOWPVT= Expressive One Word Picture Vocabulary Test, ESCS= Early Social Communication Scales, FC= Functional Communication, FEAS= Functional Emotional Assessment Scale, GARS= Gilliam Autism Rating Scale, GMDS= Griffiths Mental Development Scale, IJA= Initiations of Joint Attention, PEP-R= Psychoeducational Profile-Revised, PJAM= Precursors of Joint Attention Measure, PLS= Preschool Language Scales, MCDI= Macarthur Communicative Development Inventories, MLU= Mean Length of Utterance, PIA= Parent Interview for Autism, PPVT=Peabody Picture Vocabulary Test, RJA=Response to Joint Attention , SA=Social Awareness, SM= Social Motivation, SC= Social Communication, SCQ= Social Communication Questionnaire, SI= Social Interaction, SPA= Structured Play Assessment, SRS= Social Responsiveness Scale, SSIS= Social Skills Improvement System.

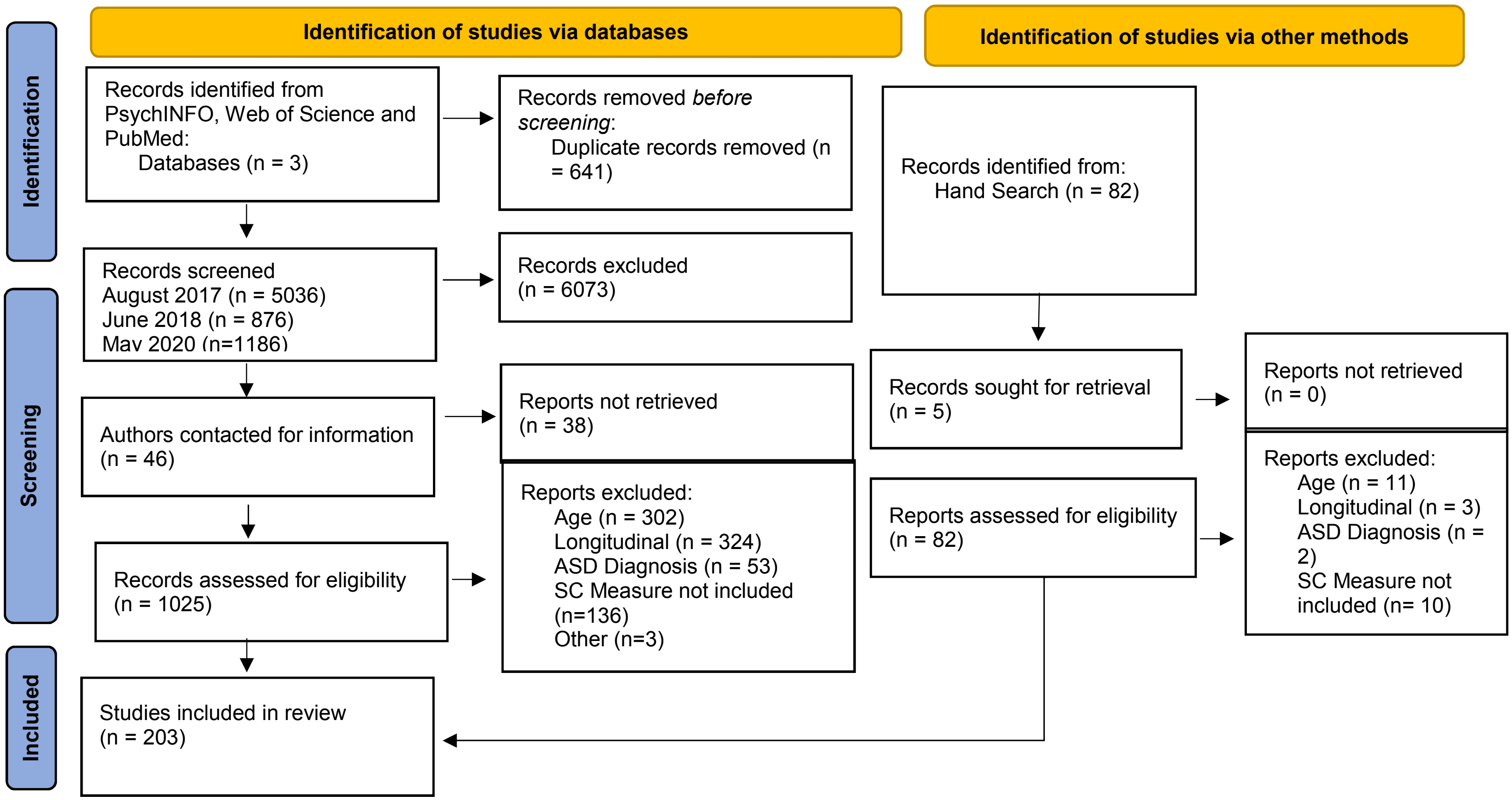
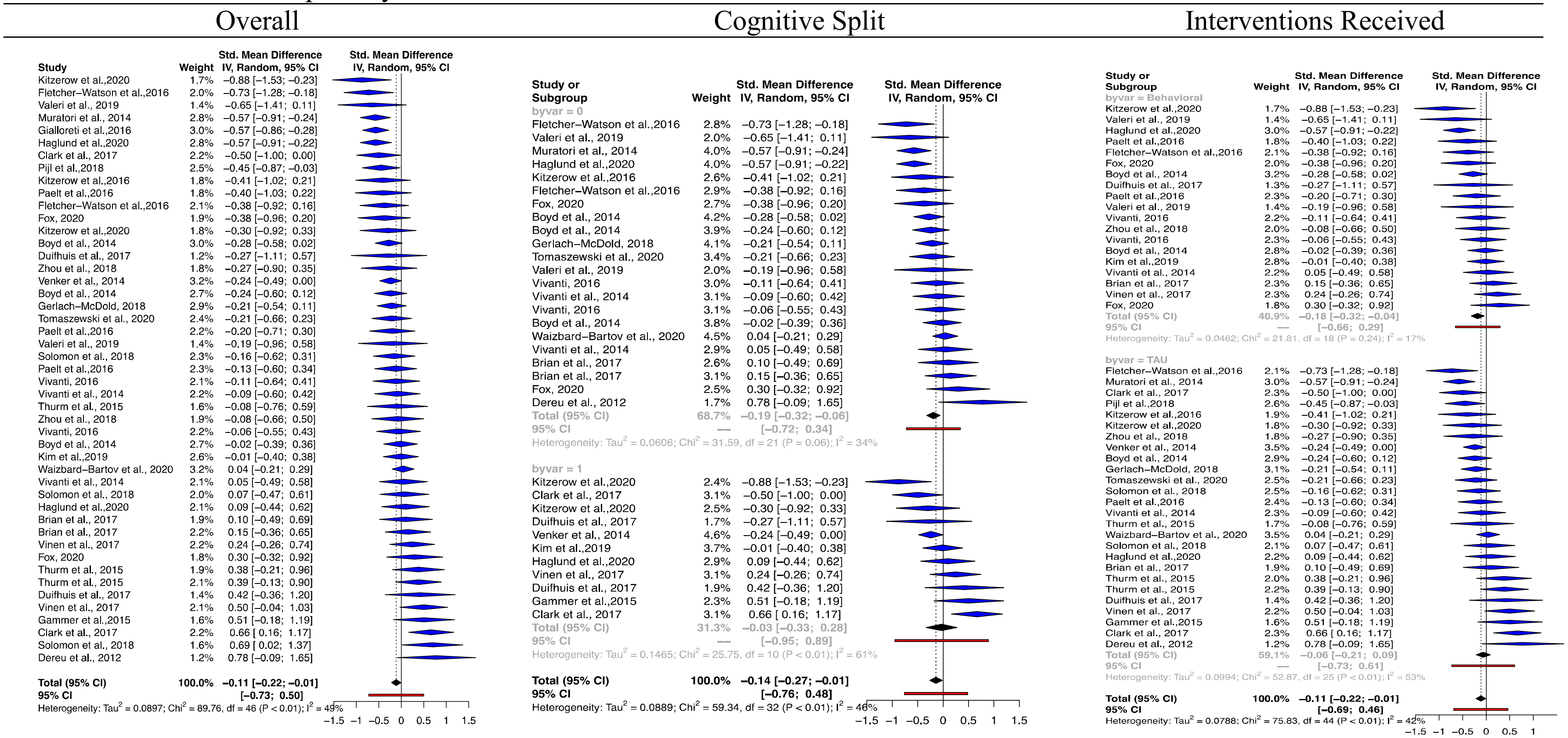


Figure 1: *PRISMA Flow Diagram for Systematic Search Procedure*

Note. From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

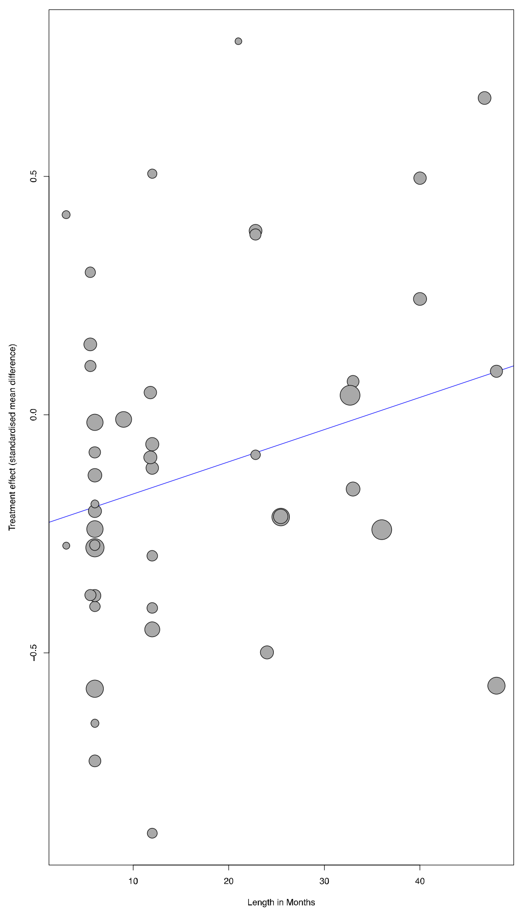
Figure 2: *ADOS Severity- Plots*

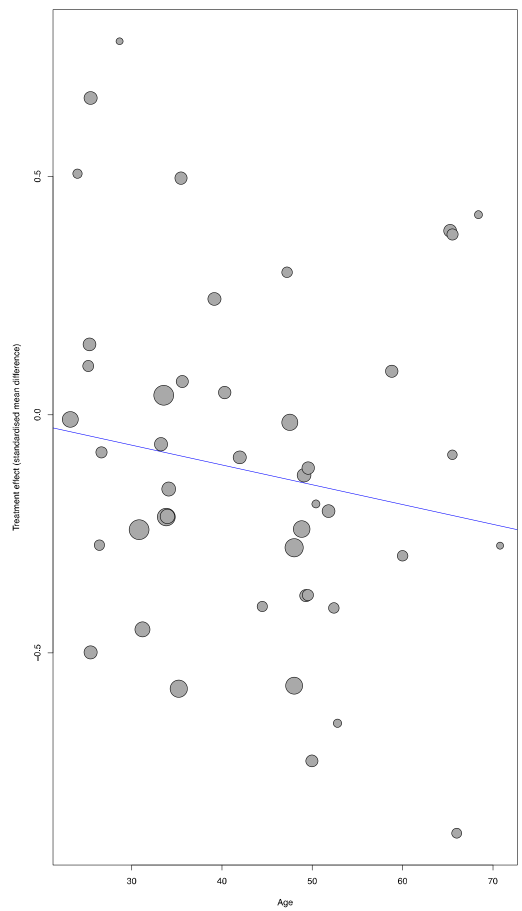
Forest Plots for Sub-Group Analyses and Overall Pooled Effect



Meta-Regression Bubble Plots

Length	Age	Year





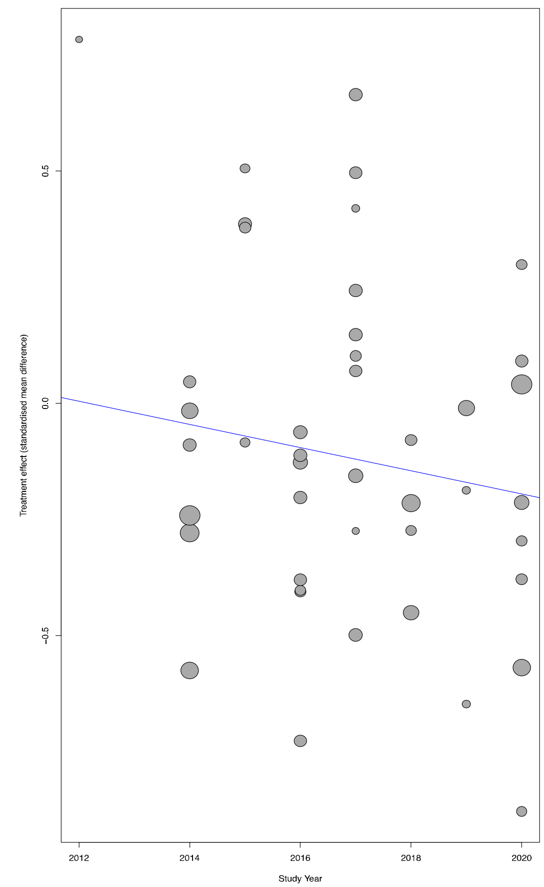
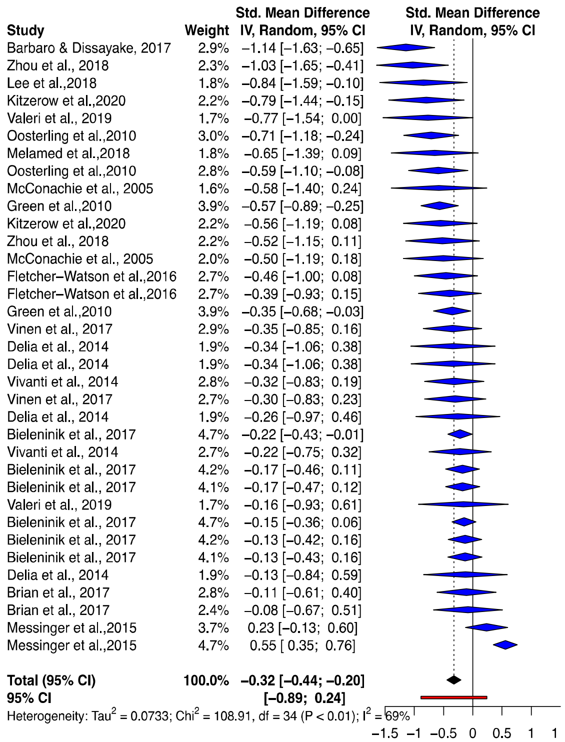


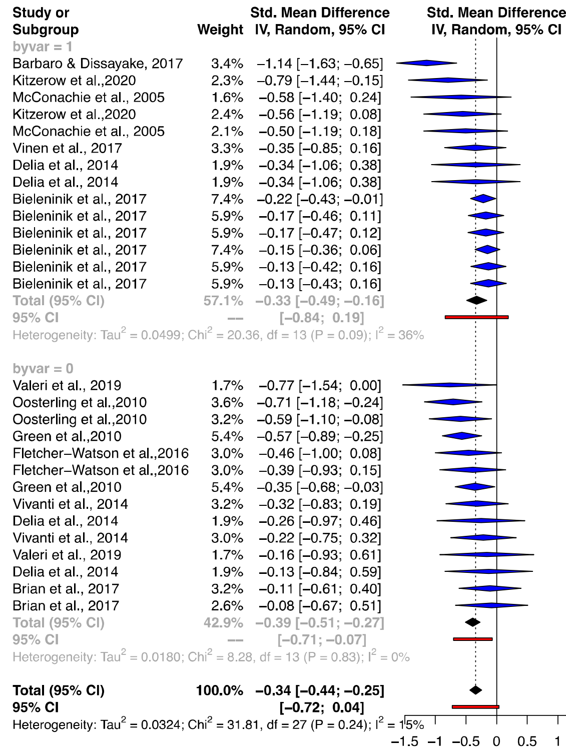
Figure 3: *ADOS Social Affect- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

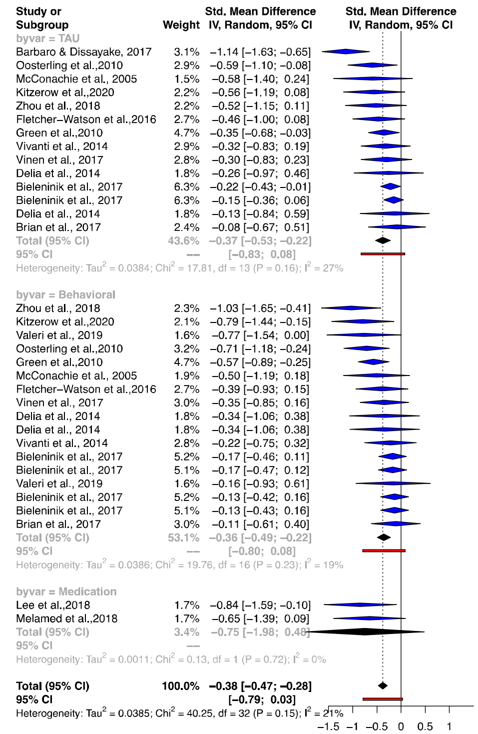
Overall



Cognitive Split



Interventions Received



Meta-Regression Bubble Plots

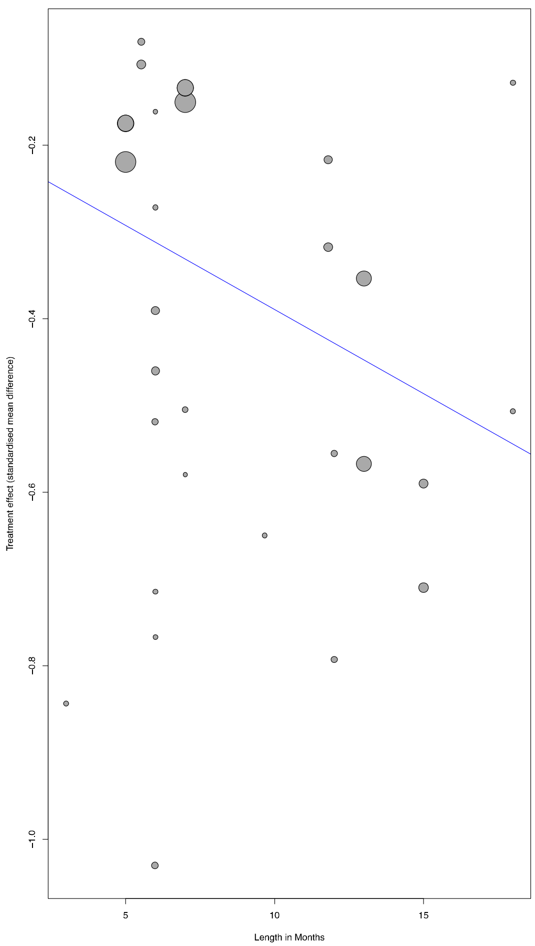
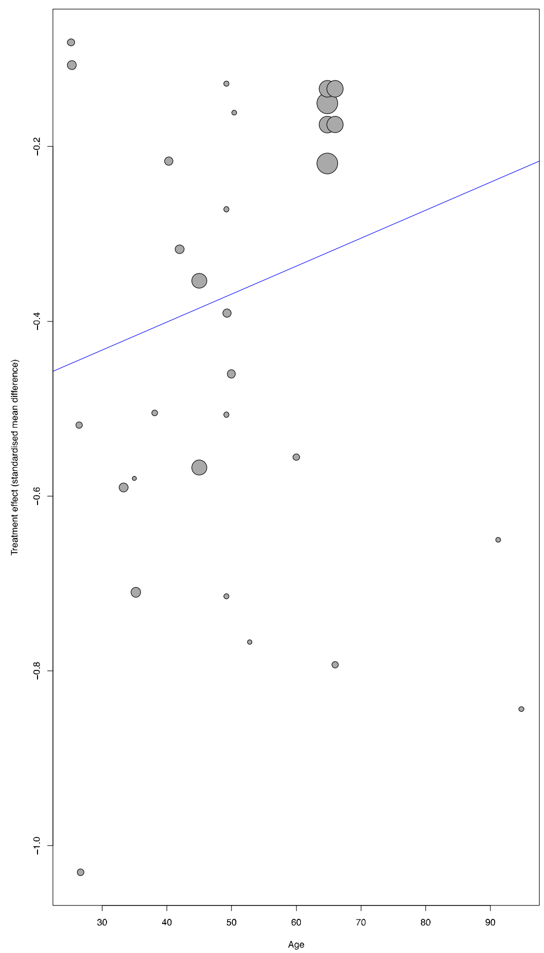
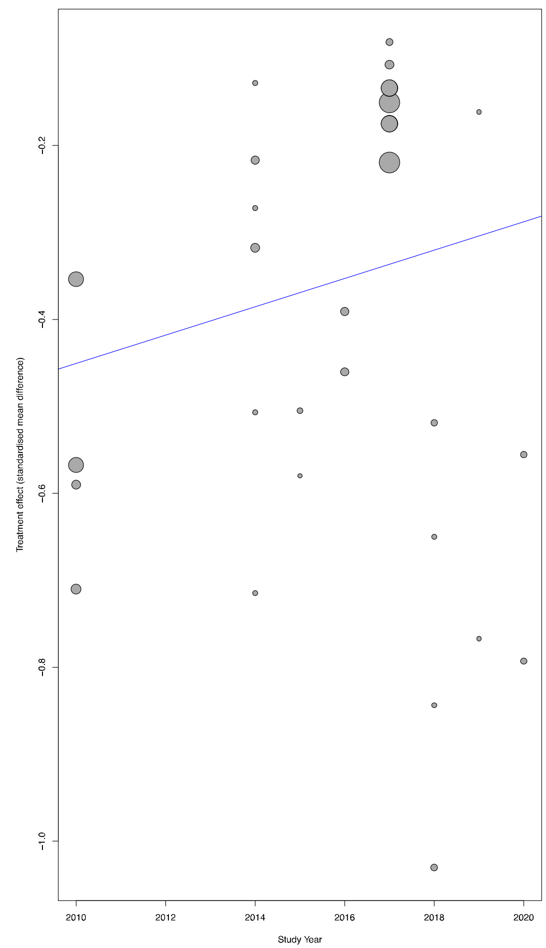
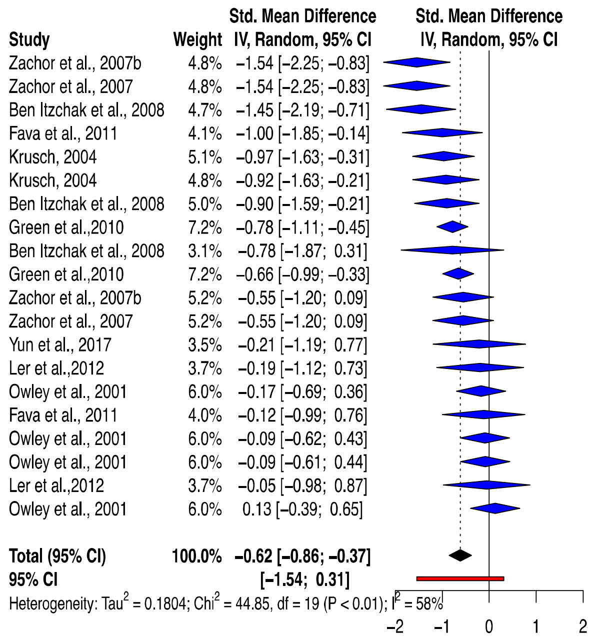
Length	Age	Year
		

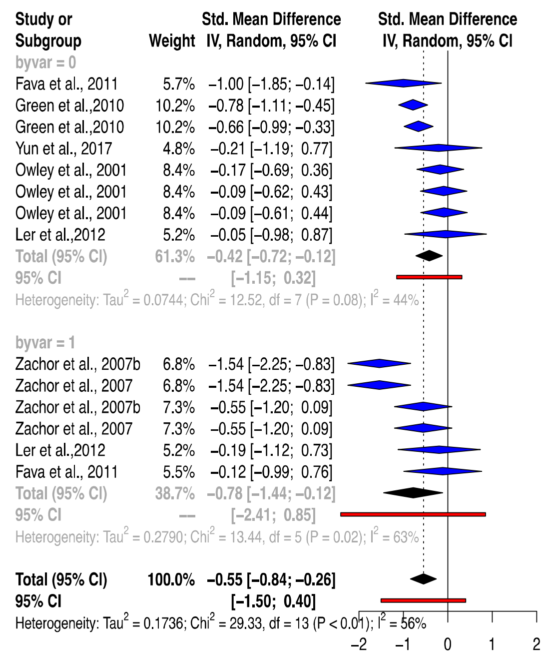
Figure 4: *ADOS Language and Communication- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

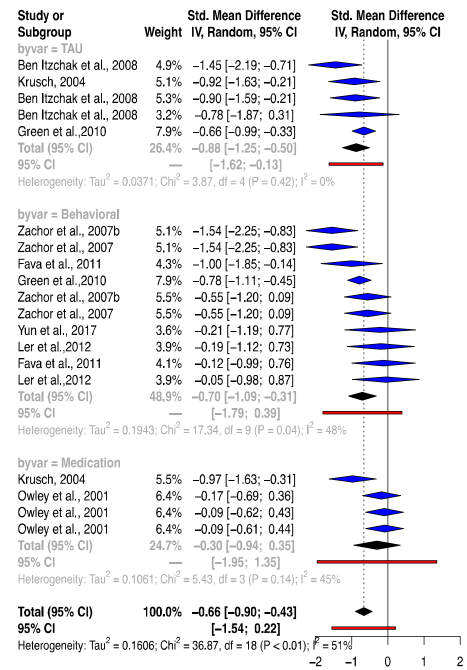
Overall



Cognitive Split

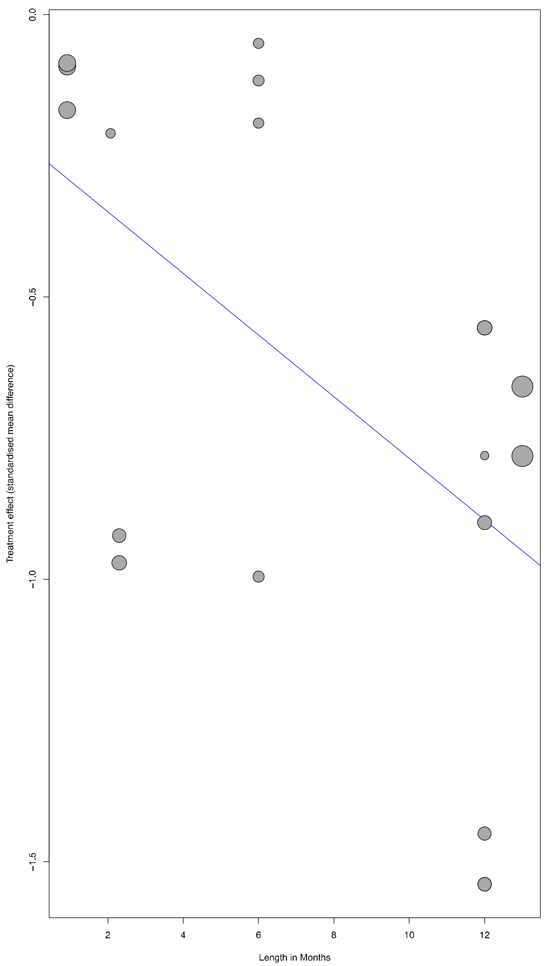


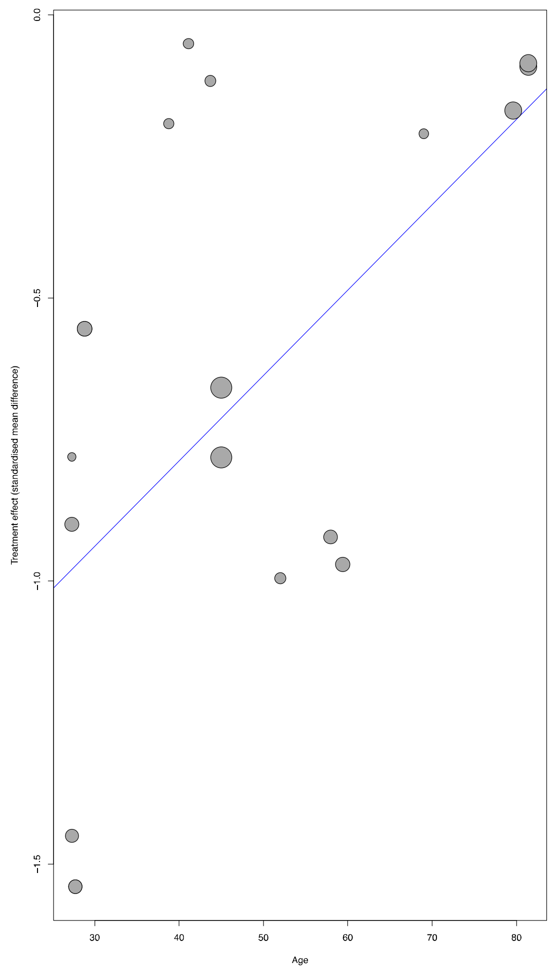
Interventions Received



Meta-Regression Bubble Plots

Length	Age	Year





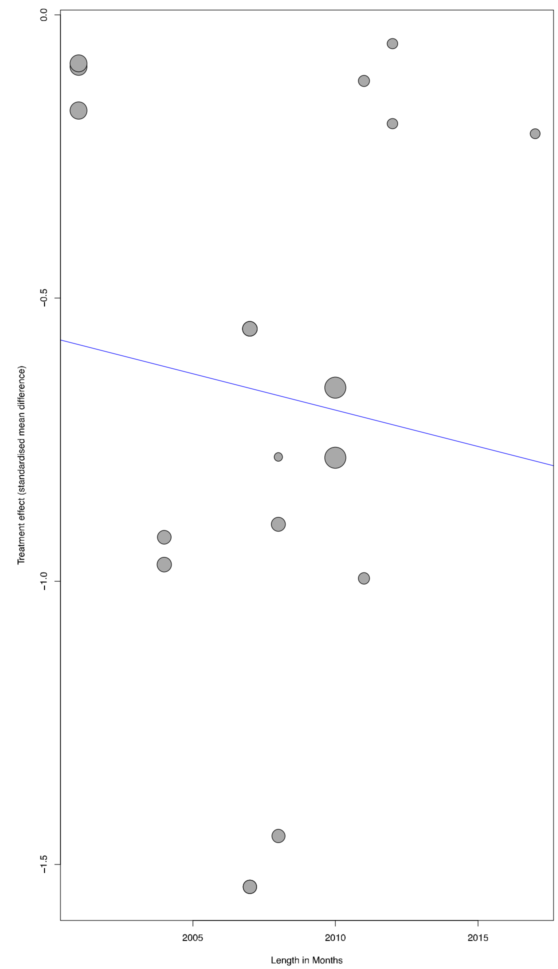
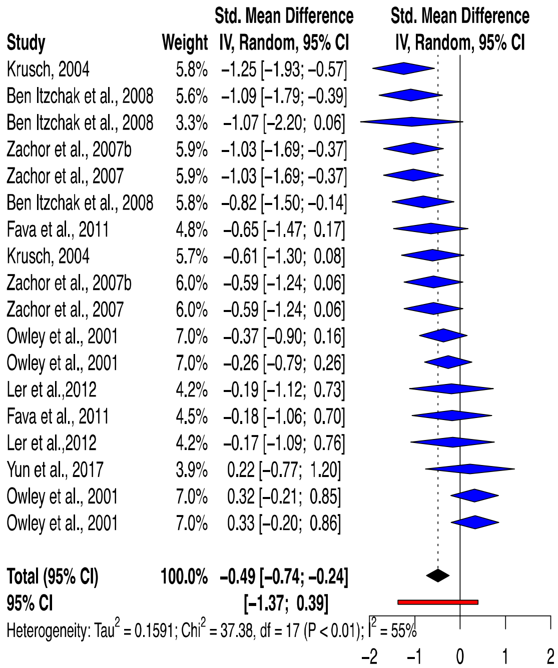
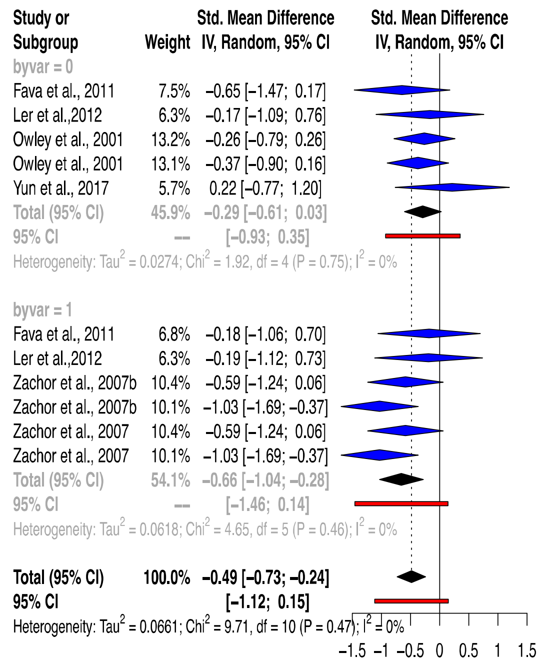


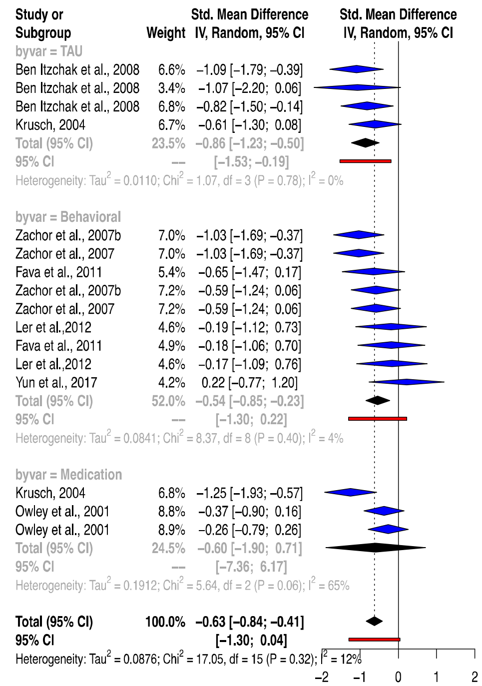
Figure 5: *ADOS Reciprocal Social Interaction- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Cognitive Split	Interventions Received







Meta-Regression Bubble Plots

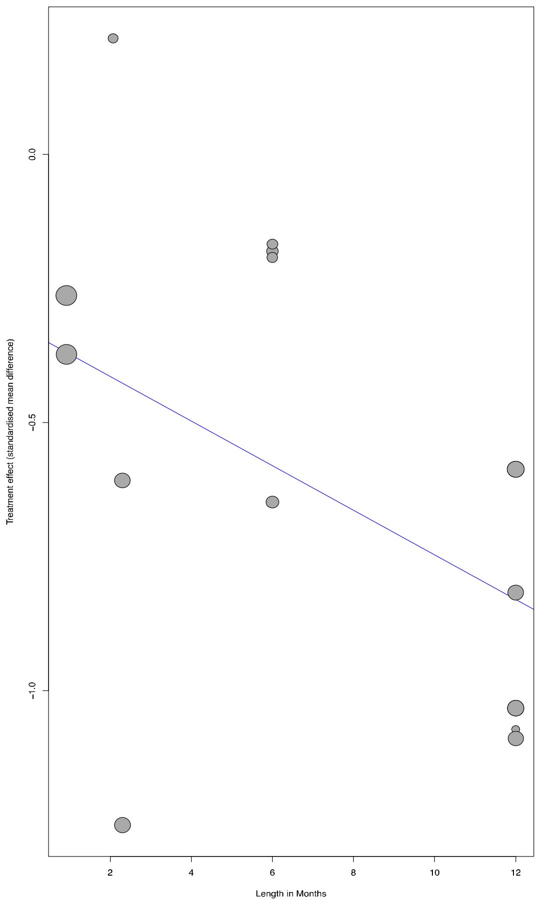
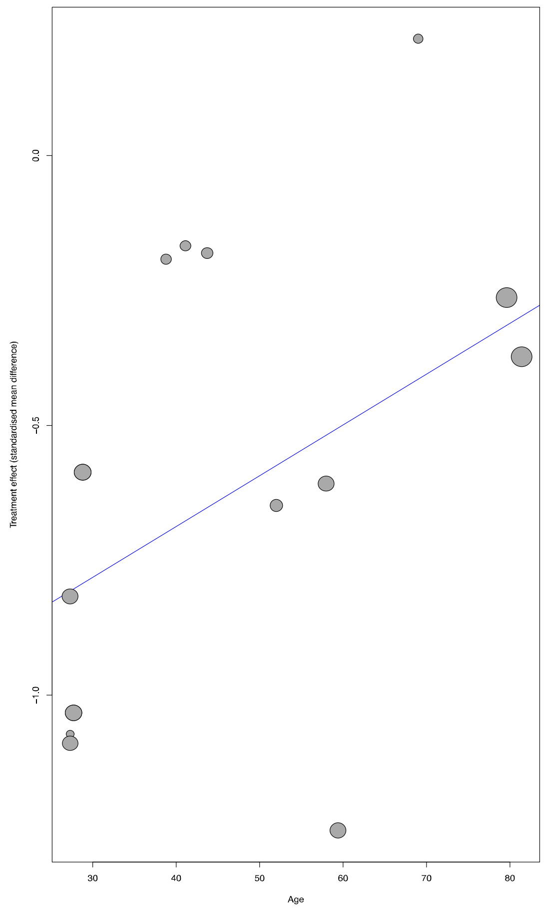
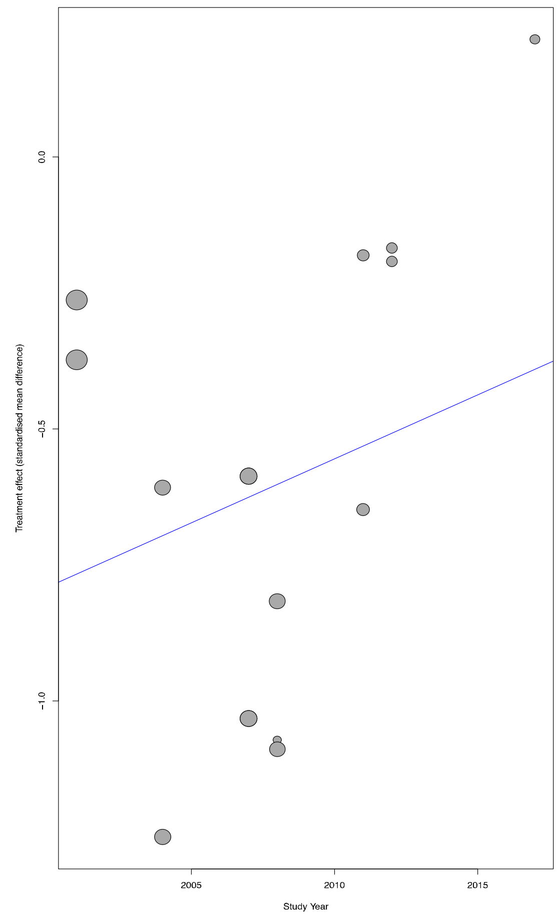
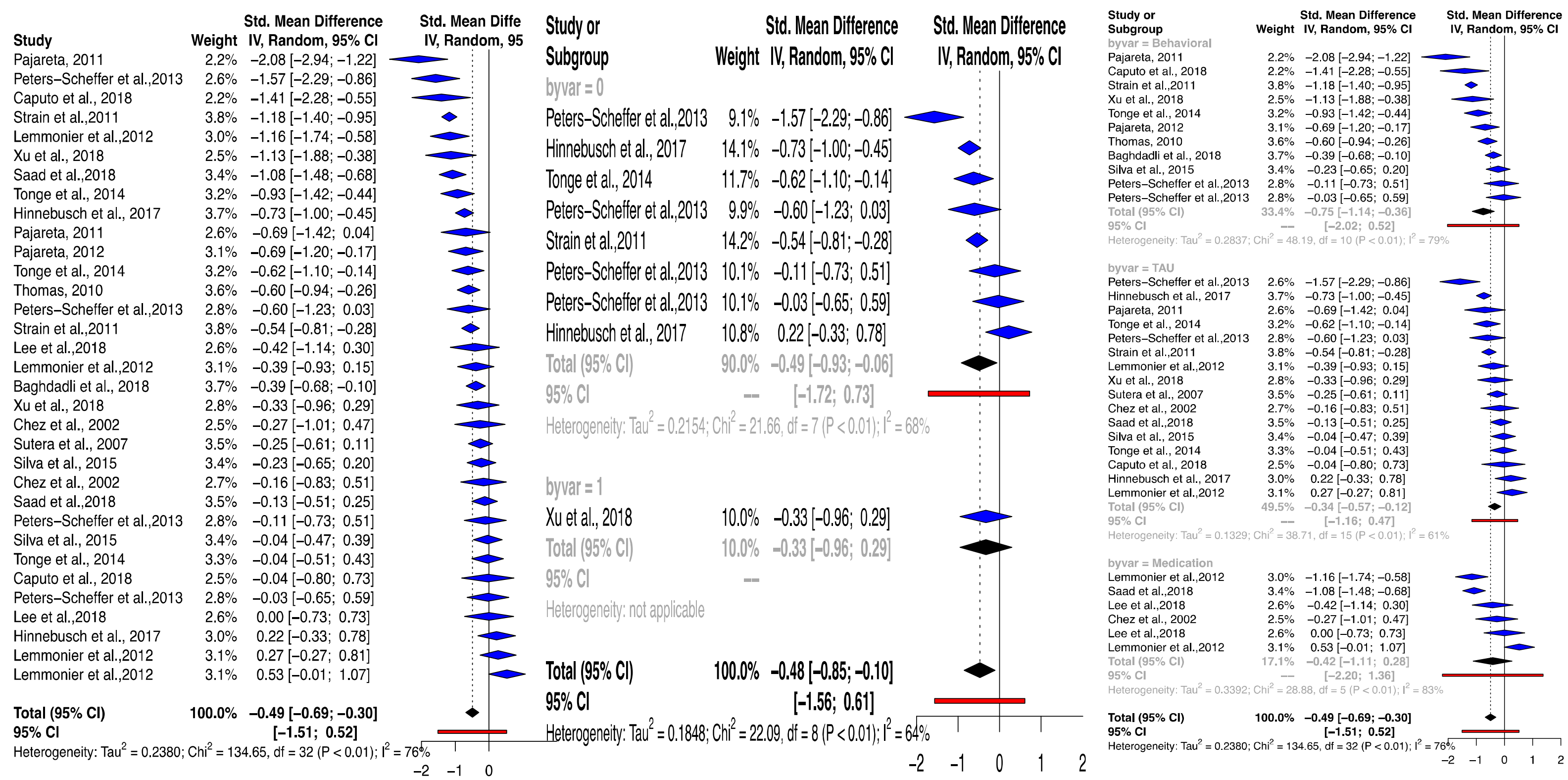
Length	Age	Year
		

Figure 6: *CARS- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall | Cognitive Split | Interventions Received



Meta-Regression Bubble Plots

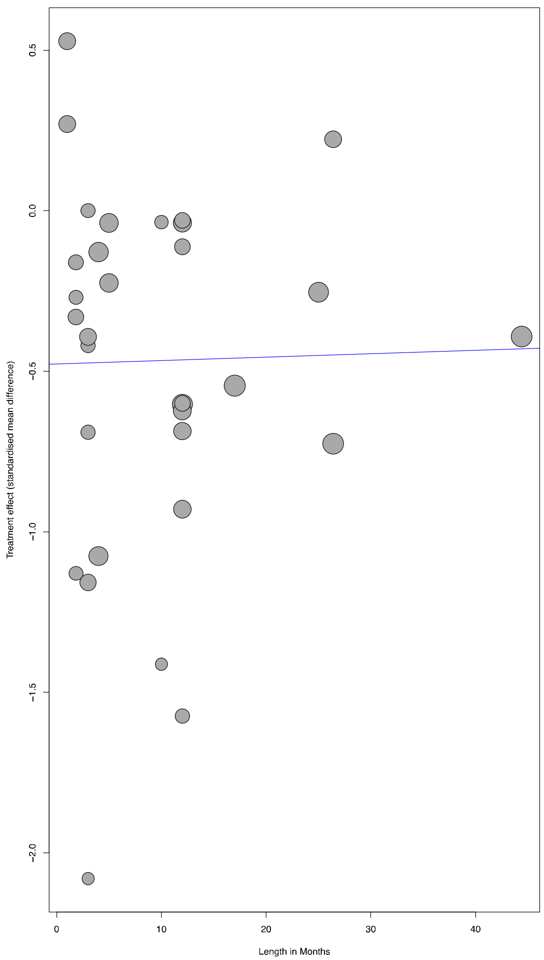
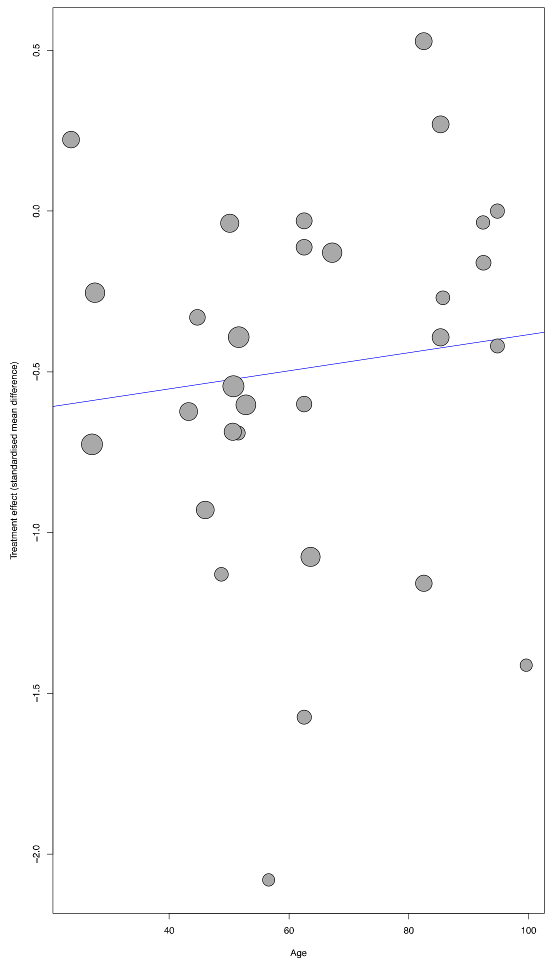
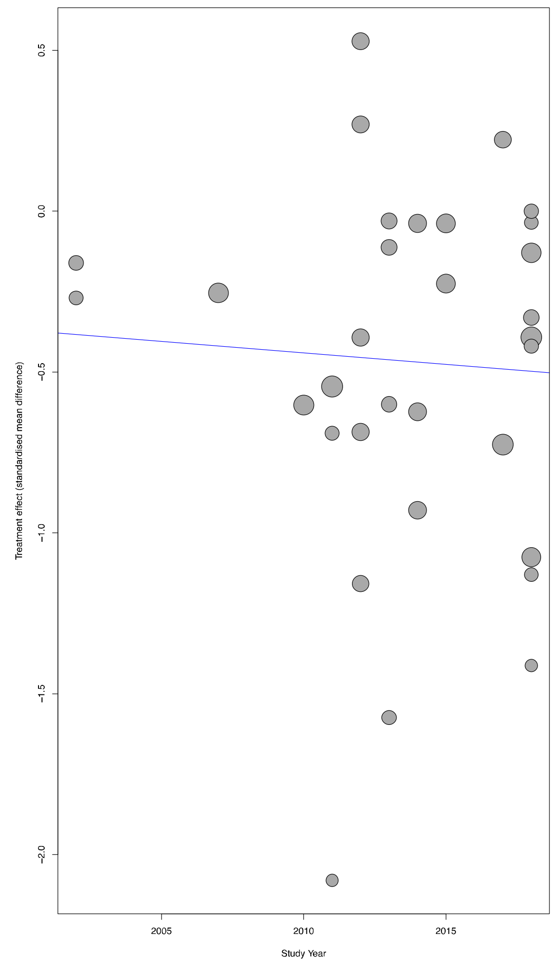
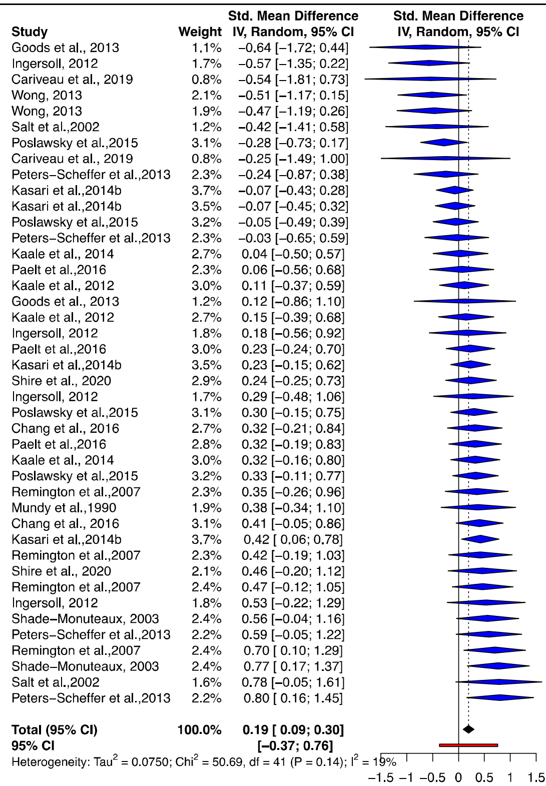
Length	Age	Year
		

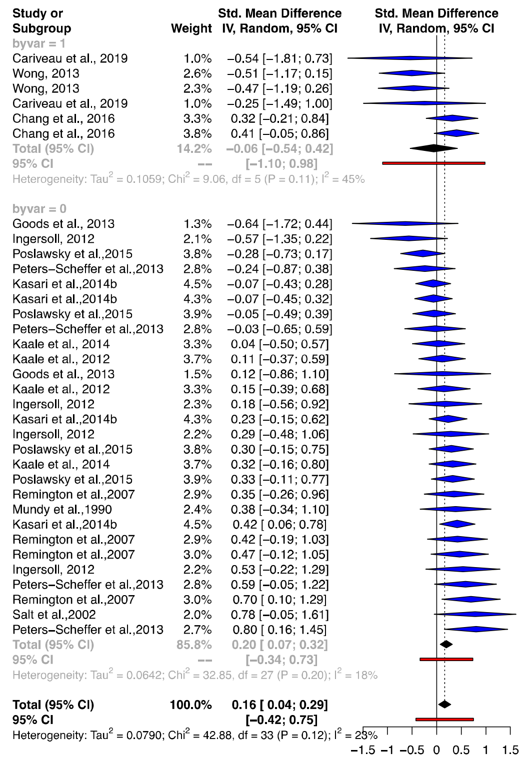
Figure 7: *ESCS IJA- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

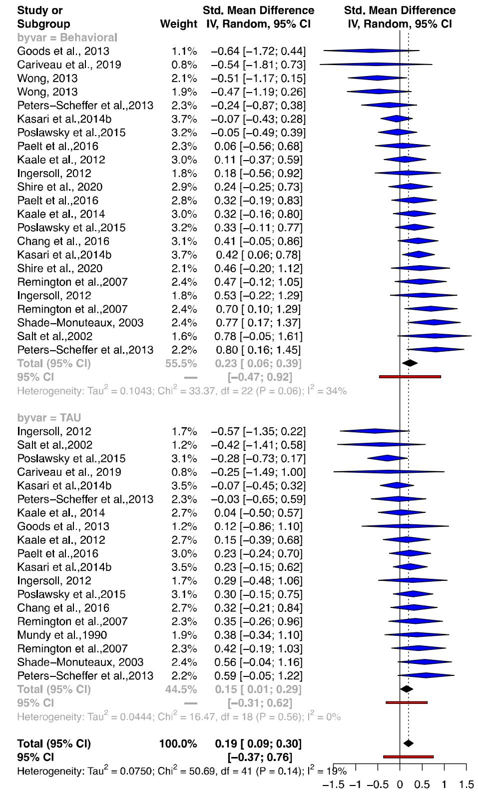
Overall



Cognitive Split



Interventions Received



Meta-Regression Bubble Plots

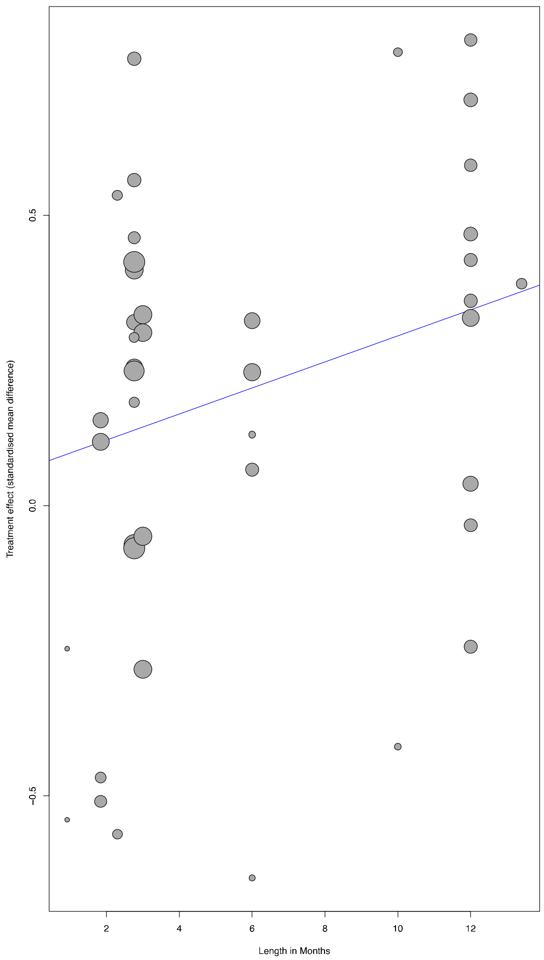
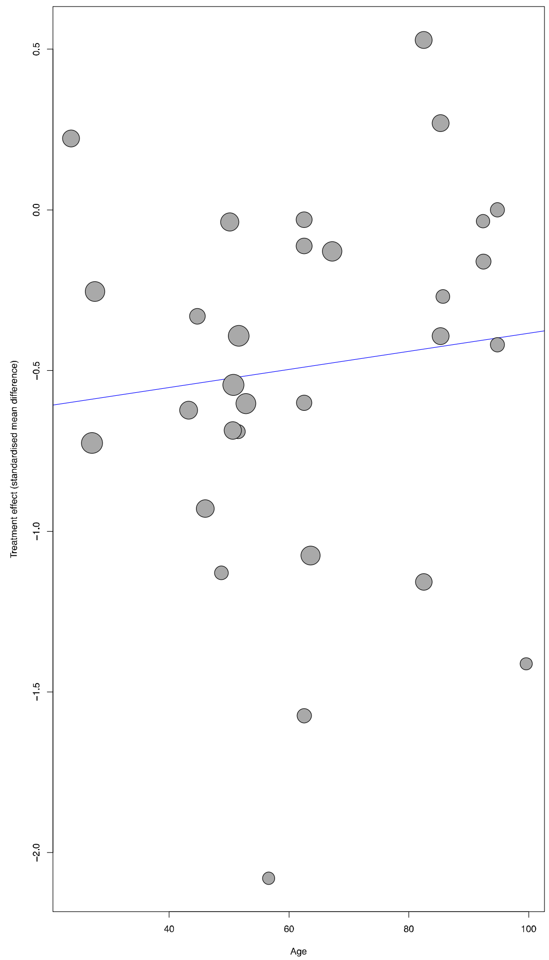
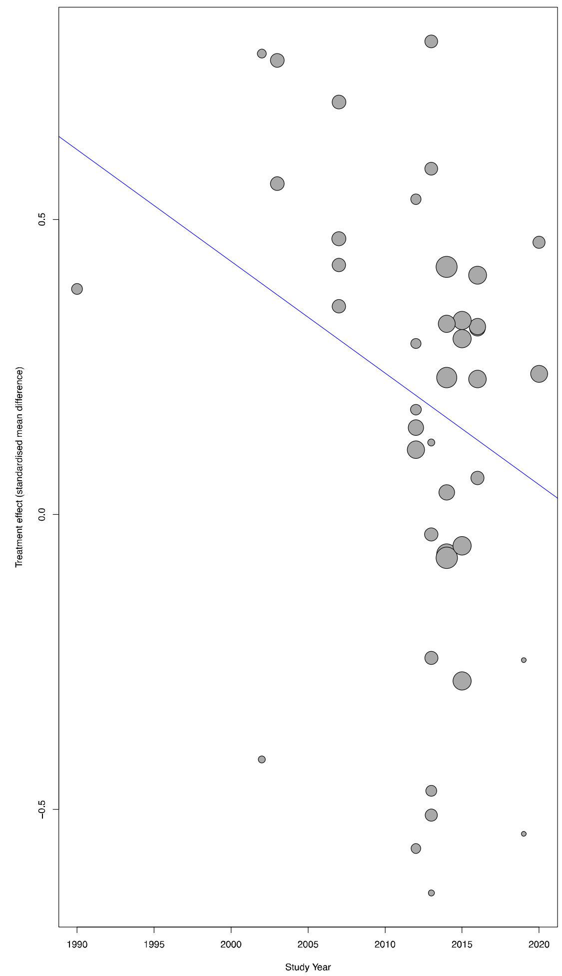
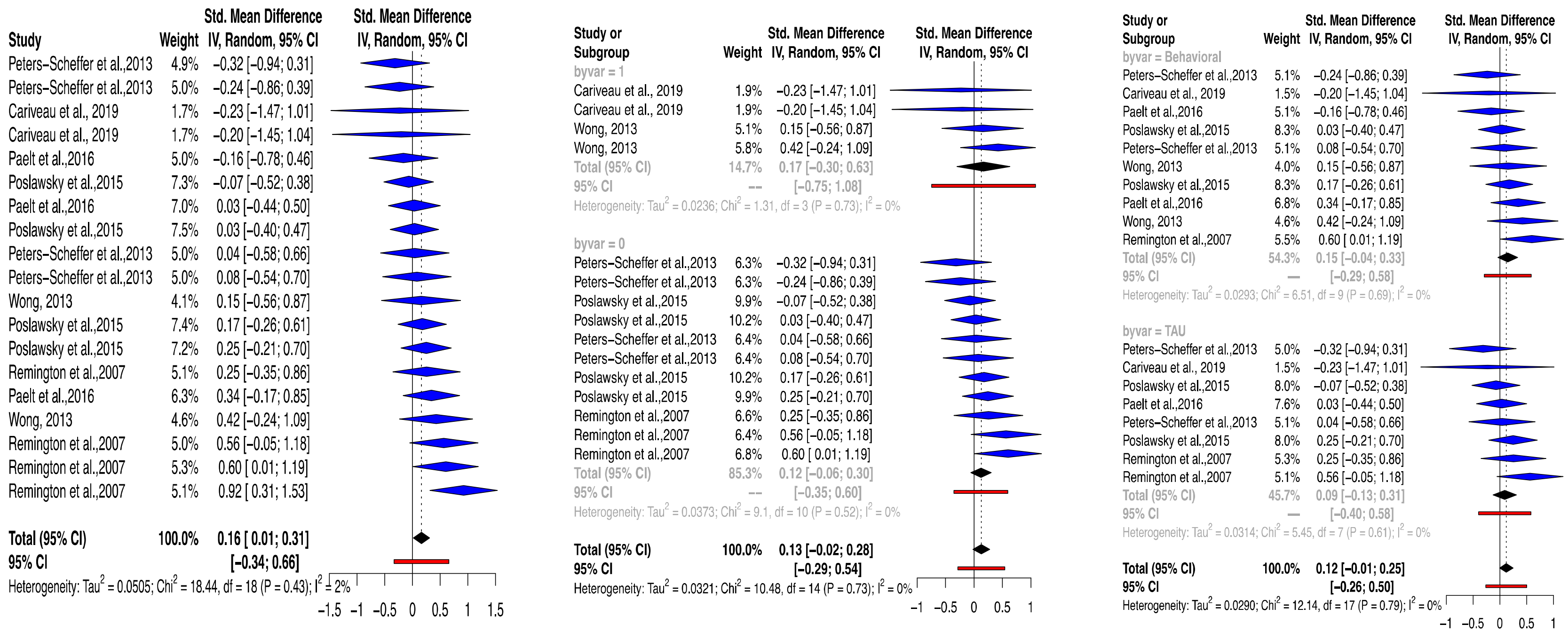
Length	Age	Year
		

Figure 8: *ESCS RJA- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Cognitive Split	Interventions Received



Meta-Regression Bubble Plots

Length	Age	Year

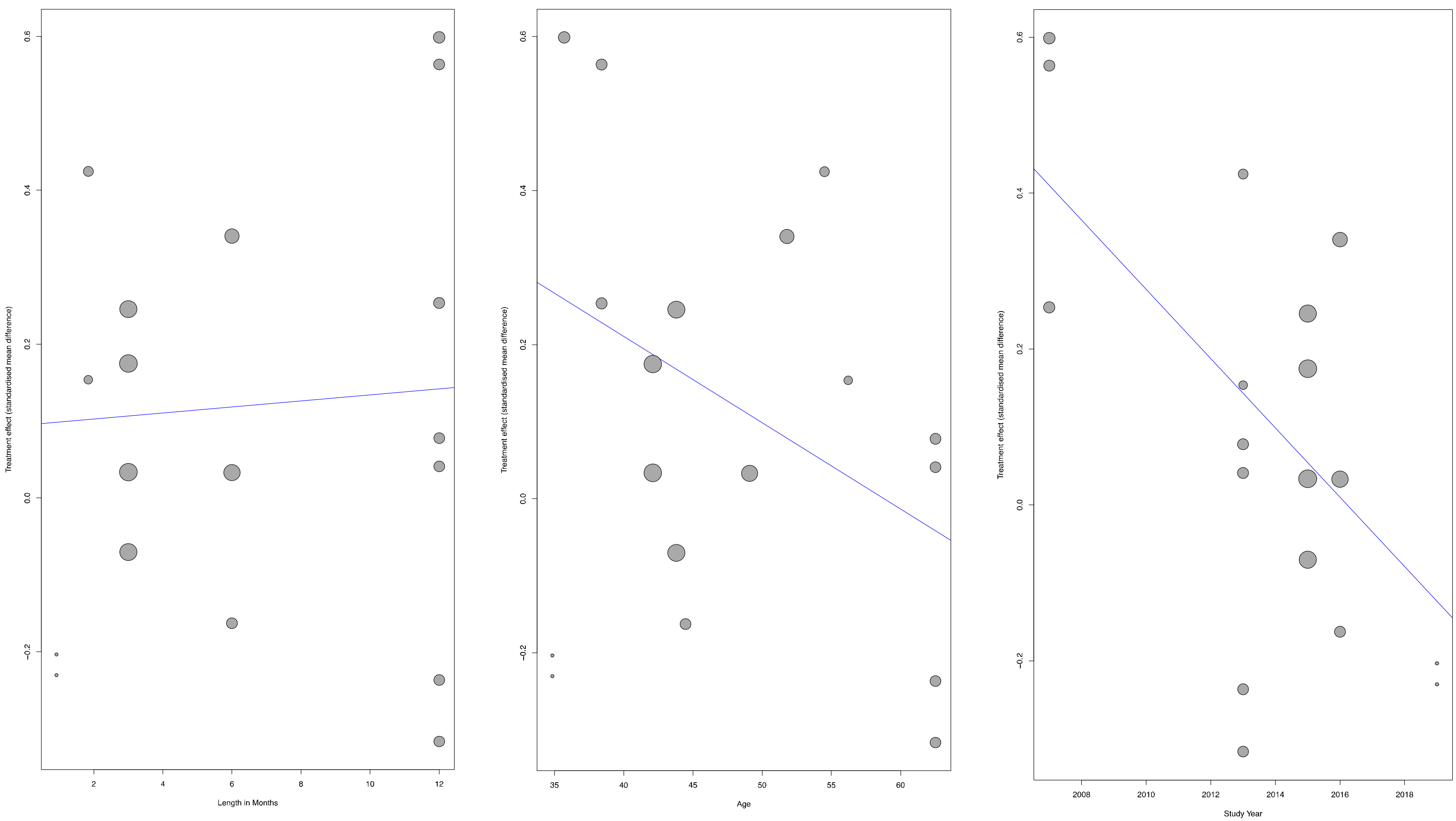
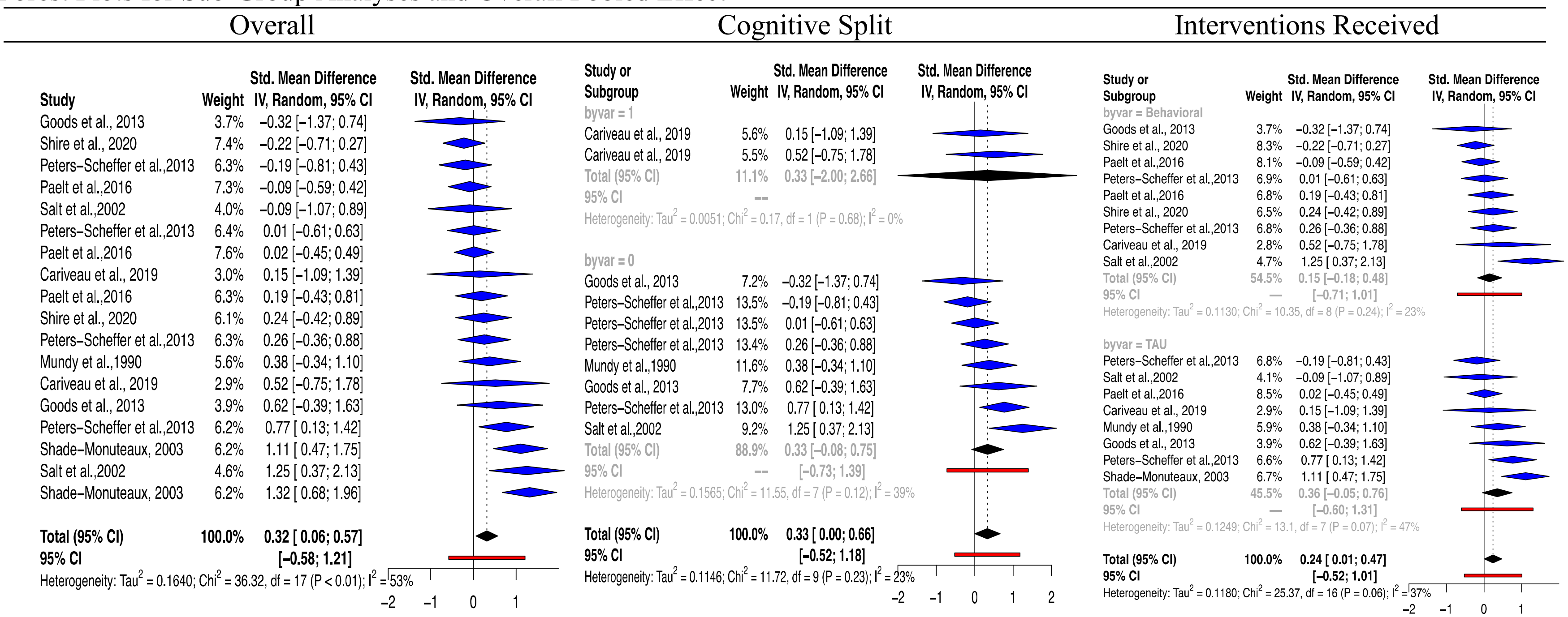


Figure 9: *ESCS Requesting- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect



Meta-Regression Bubble Plots

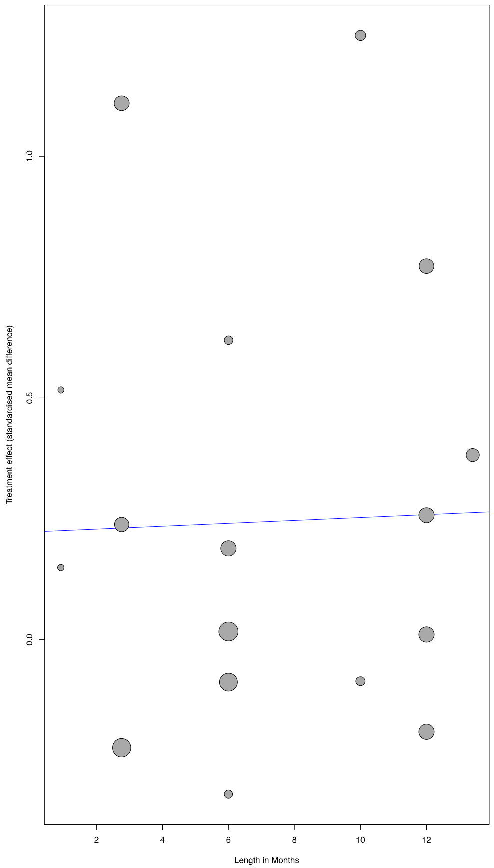
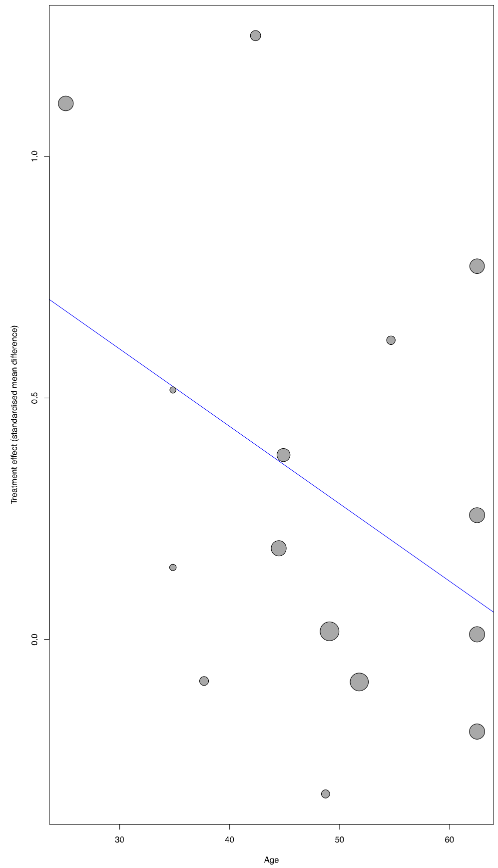
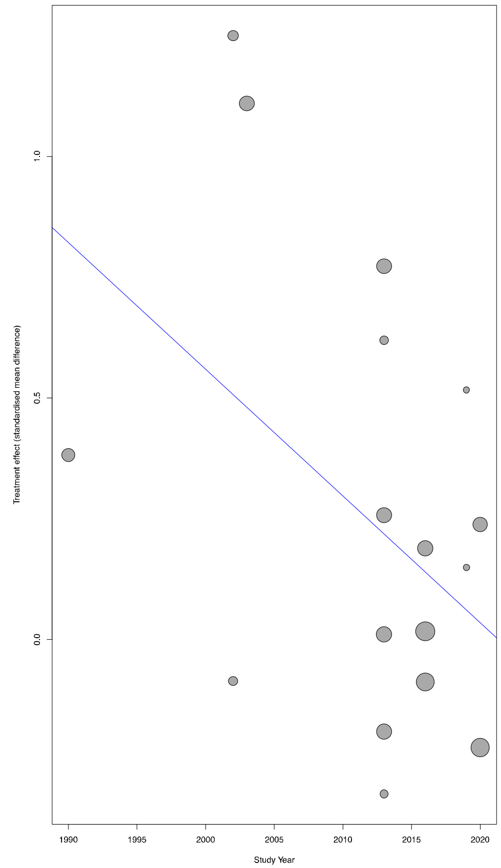
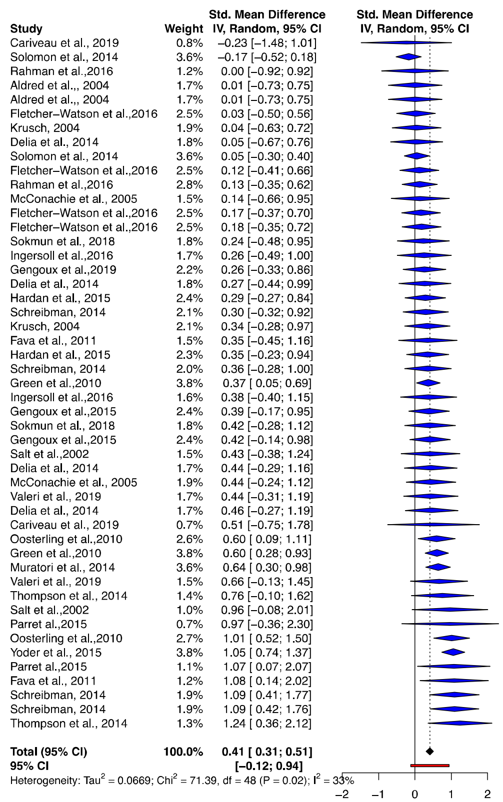
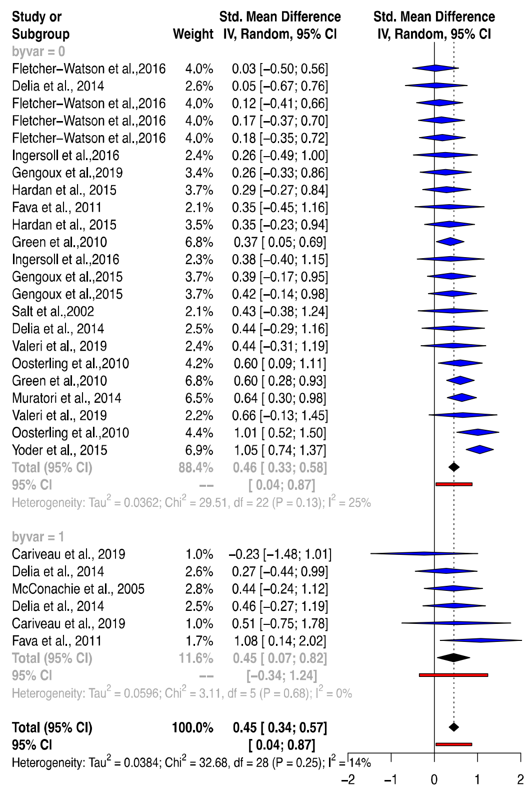
Length	Age	Year
		

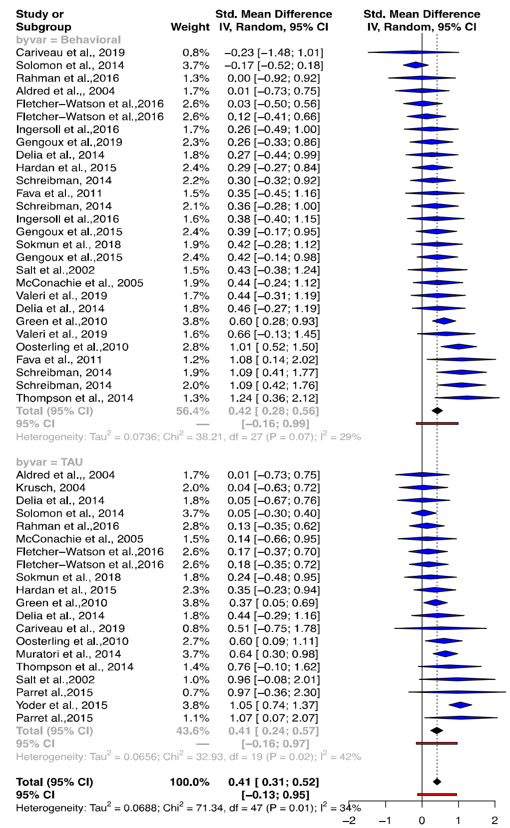
Figure 10: *MCDI Words Produced- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Cognitive Split	Interventions Received







Meta-Regression Bubble Plots

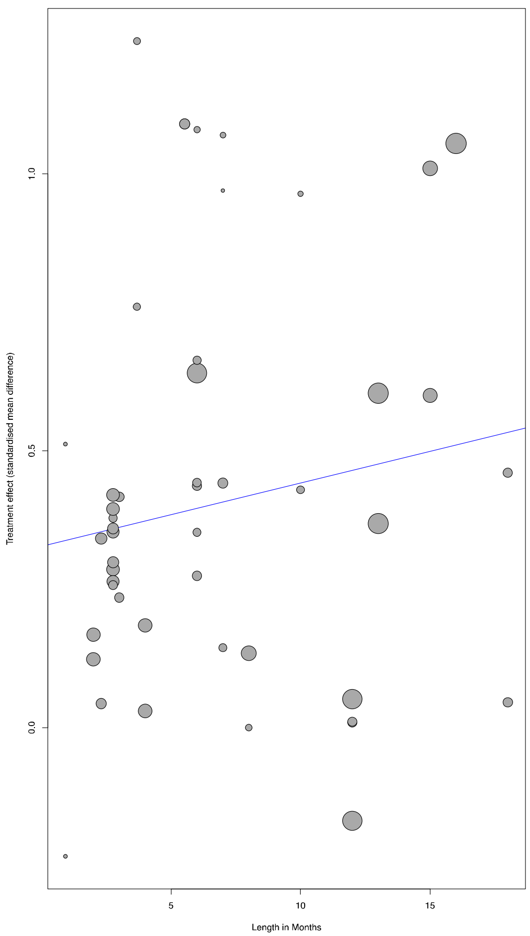
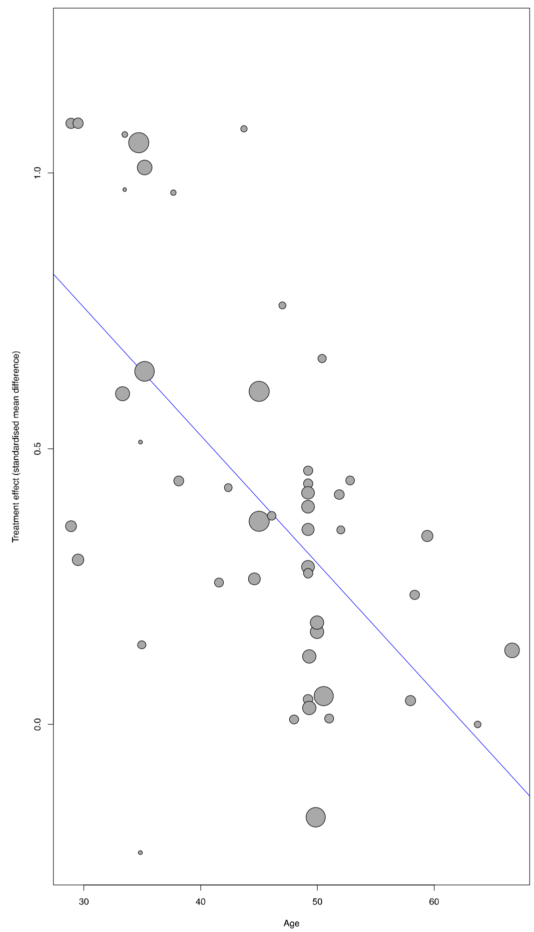
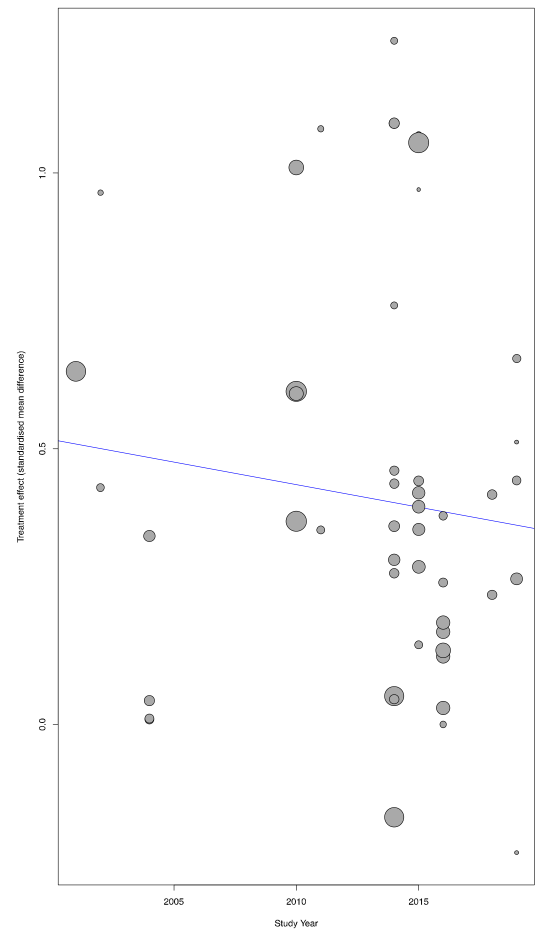
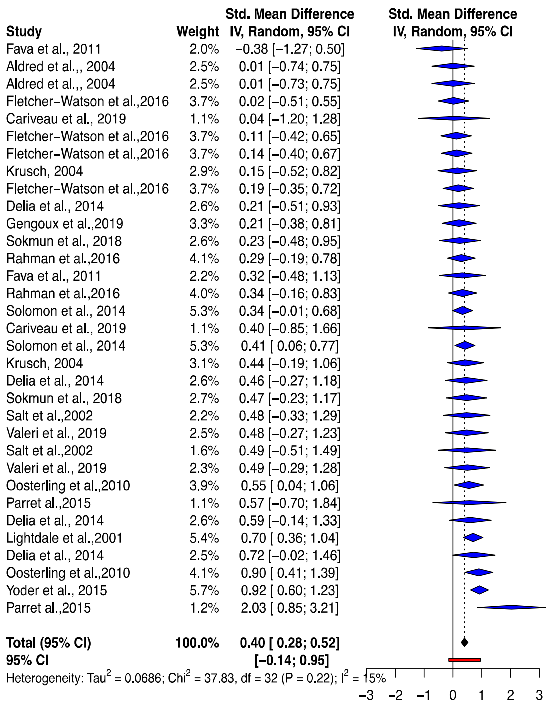
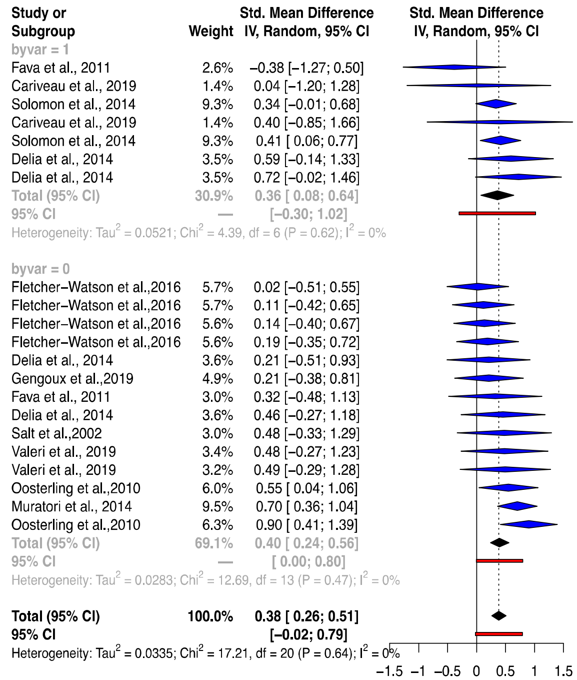
Length	Age	Year
		

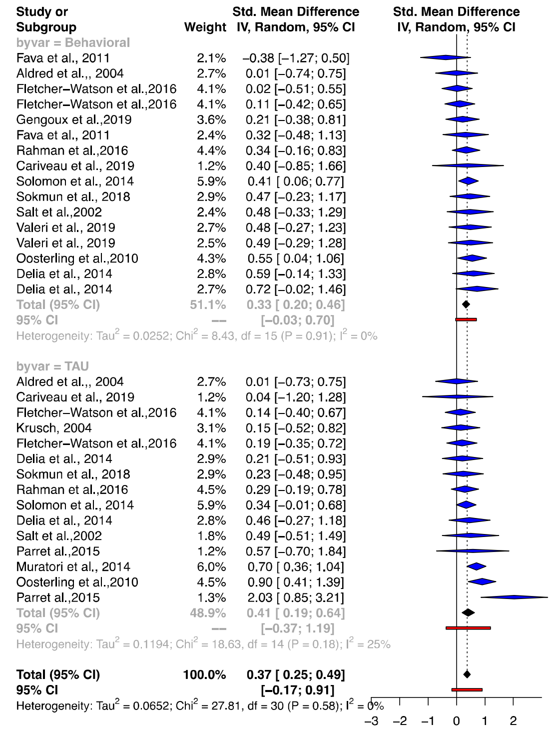
Figure 11: *MCDI Words Understood- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Cognitive Split	Interventions Received







Meta-Regression Bubble Plots

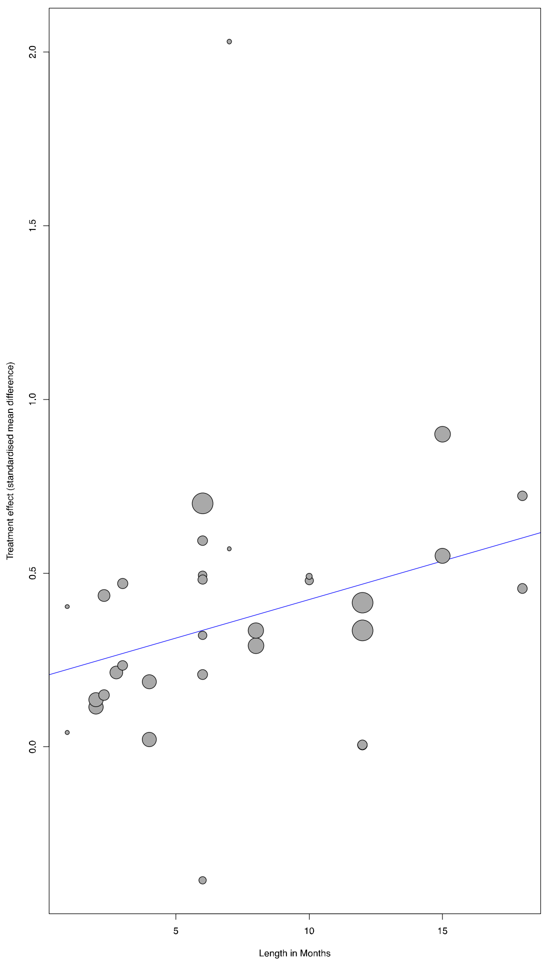
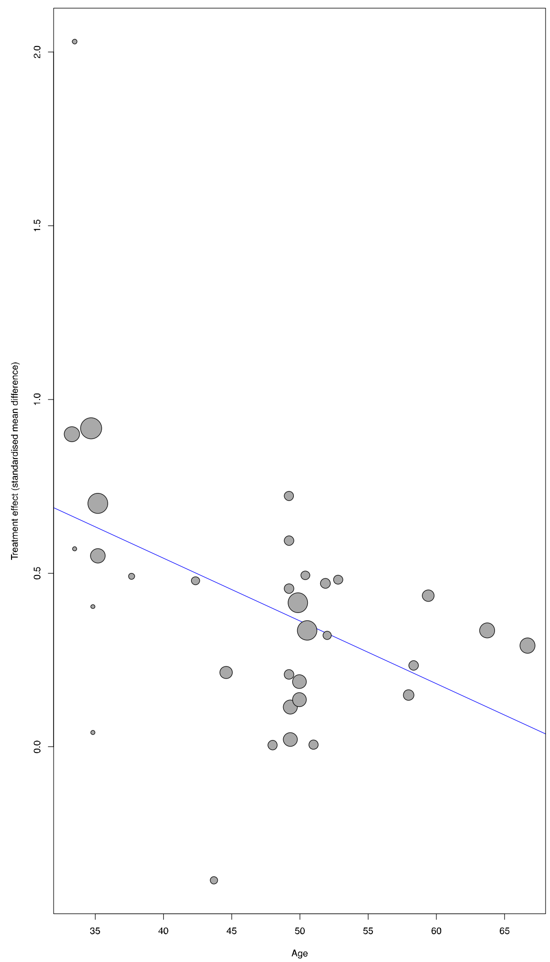
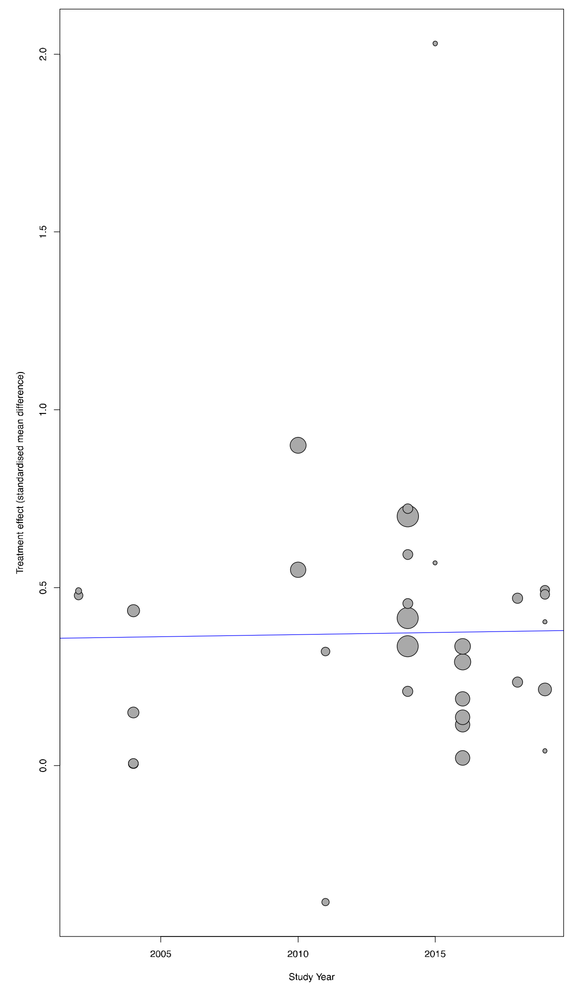
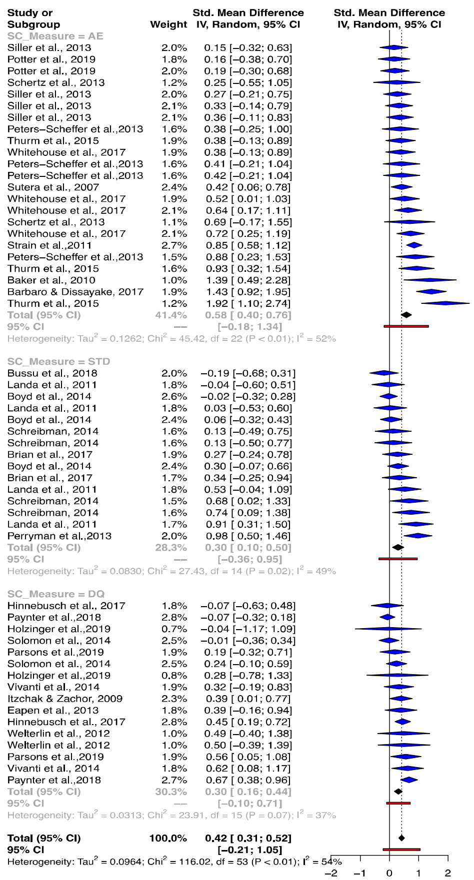
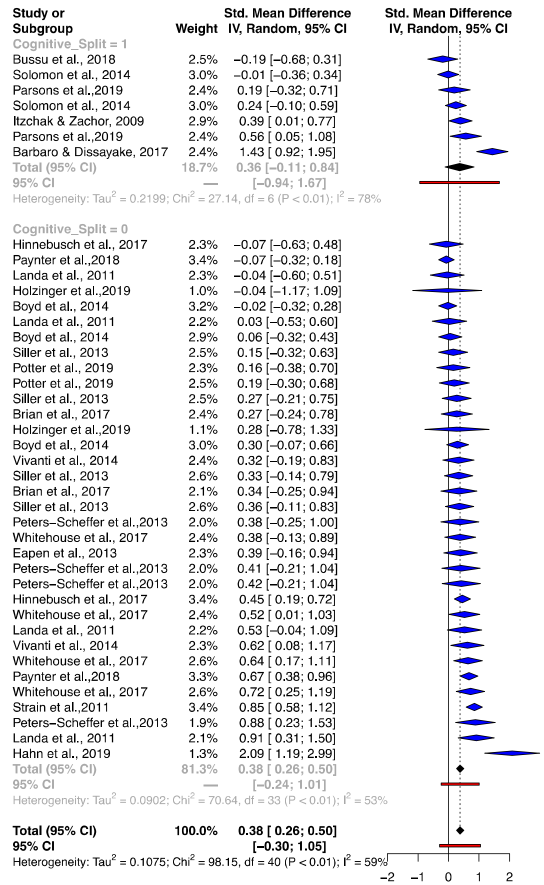
Length	Age	Year
		

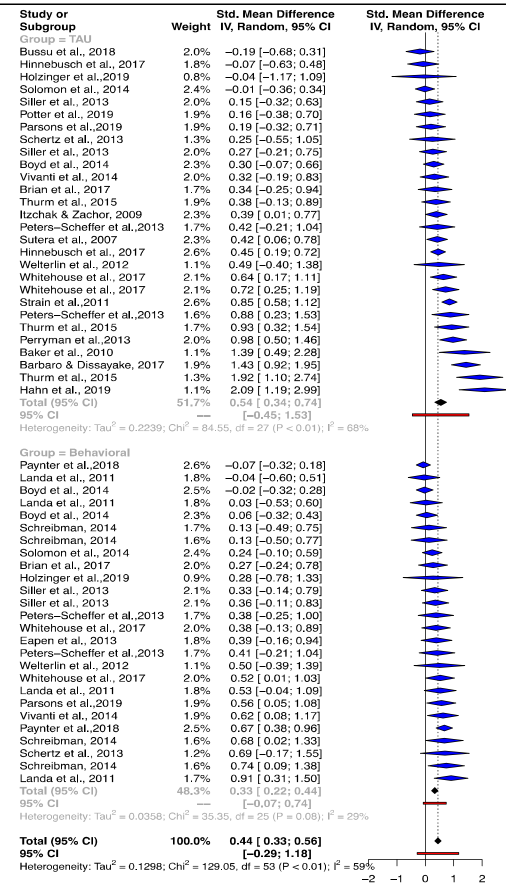
Figure 12: *Mullen Expressive Language- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Type of Measure	Cognitive Split	Interventions Received

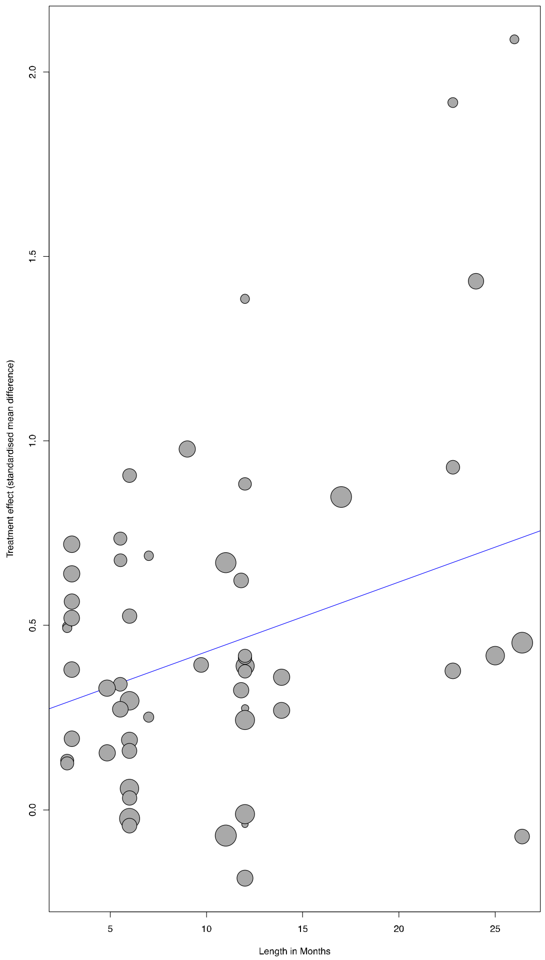


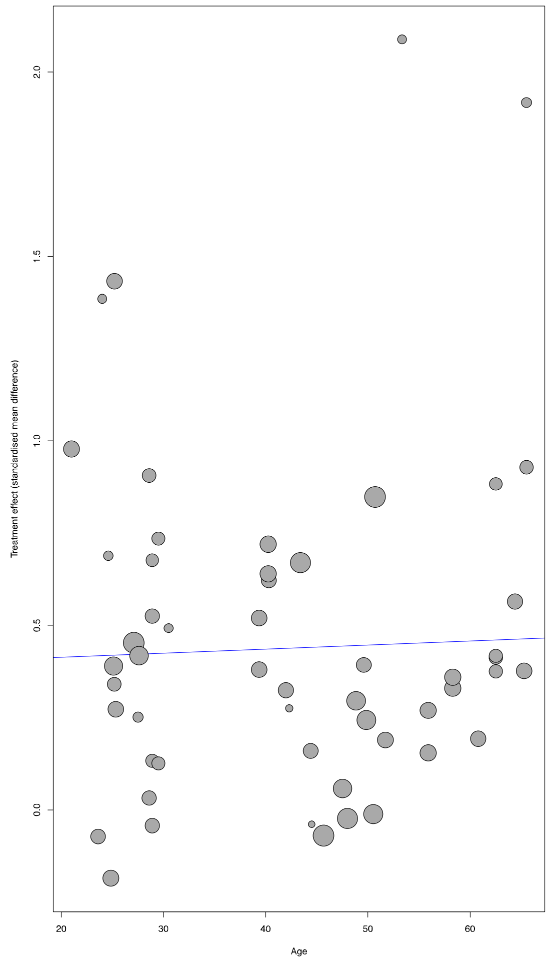




Meta-Regression Bubble Plots

Length	Age	Year





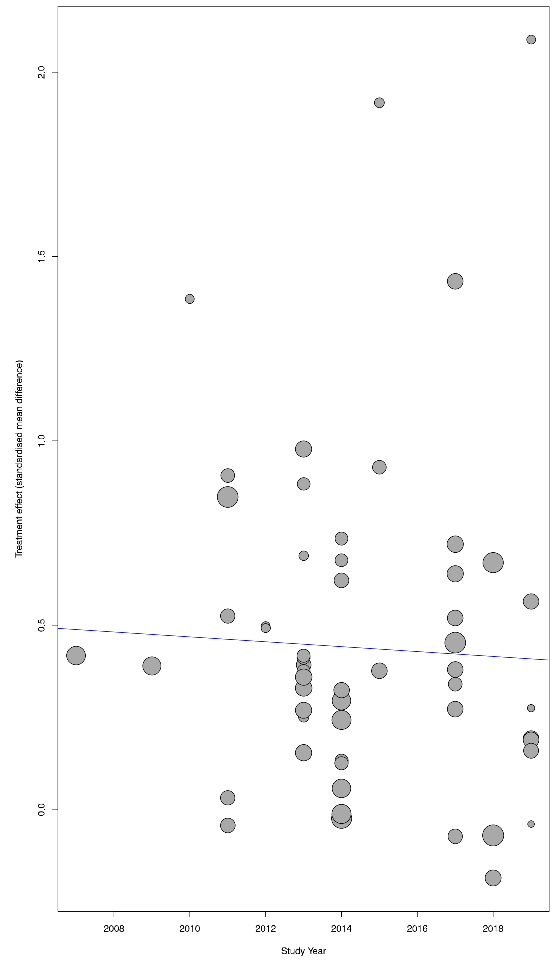
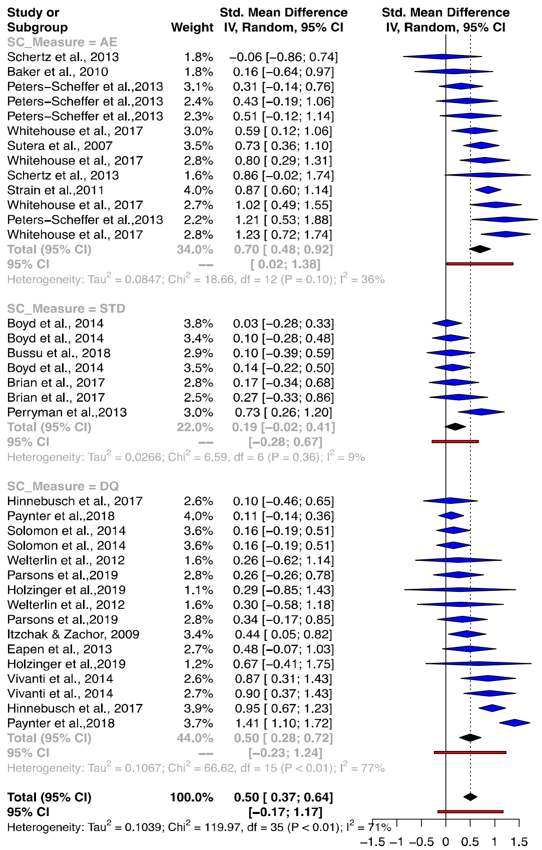
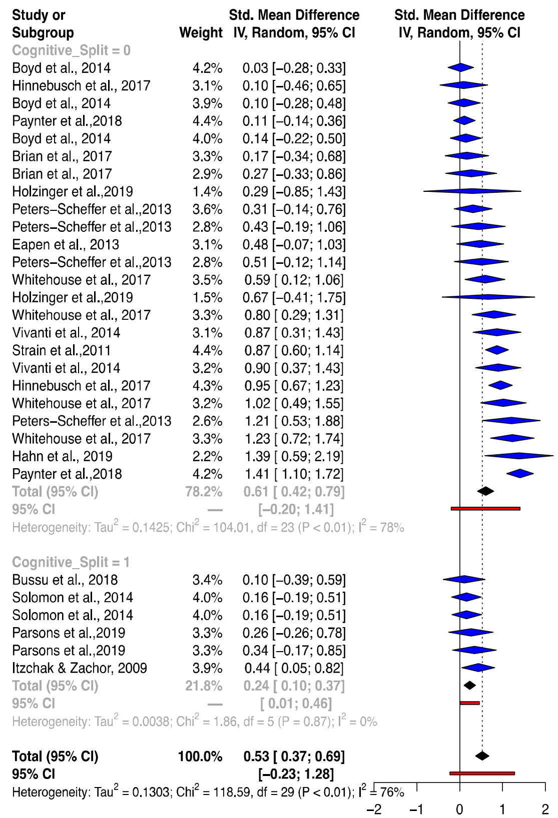


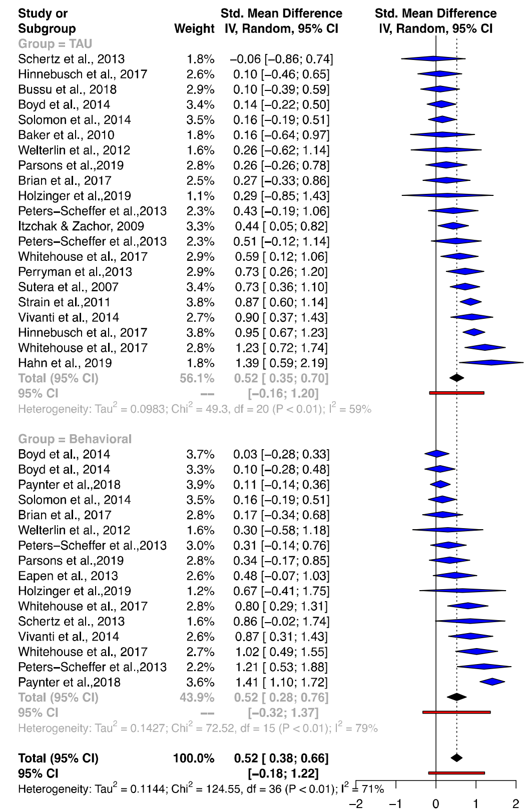
Figure 13: *Mullen Receptive Language- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Measure Type	Cognitive Split	Interventions Received

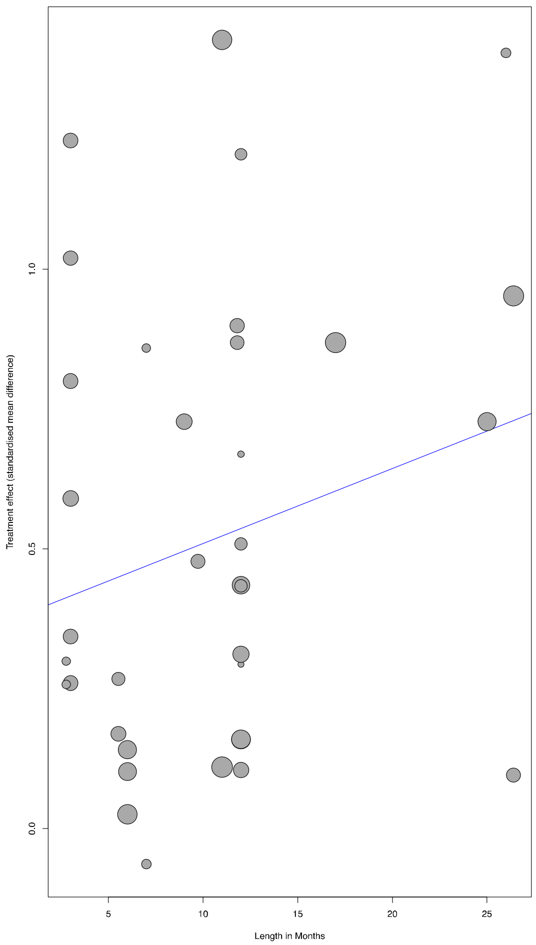


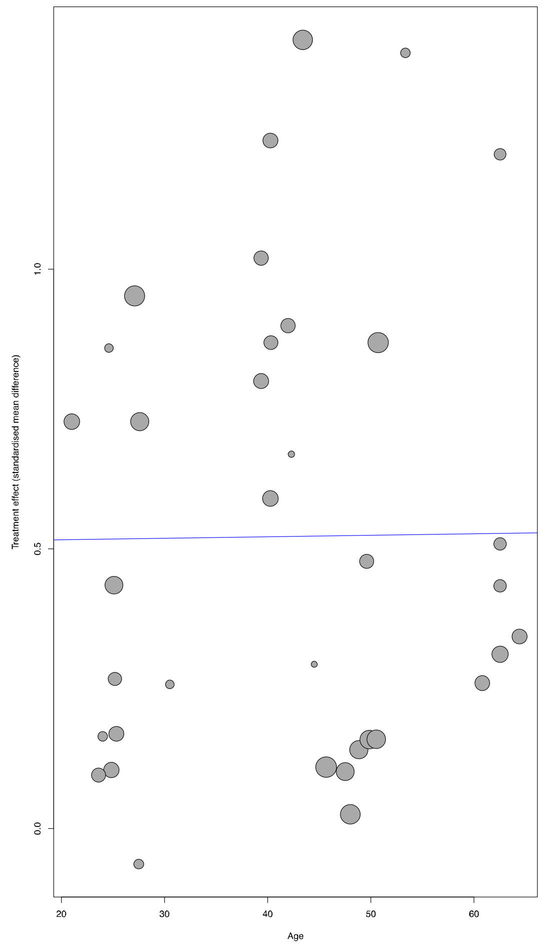




Meta-Regression Bubble Plots

Length	Age	Year





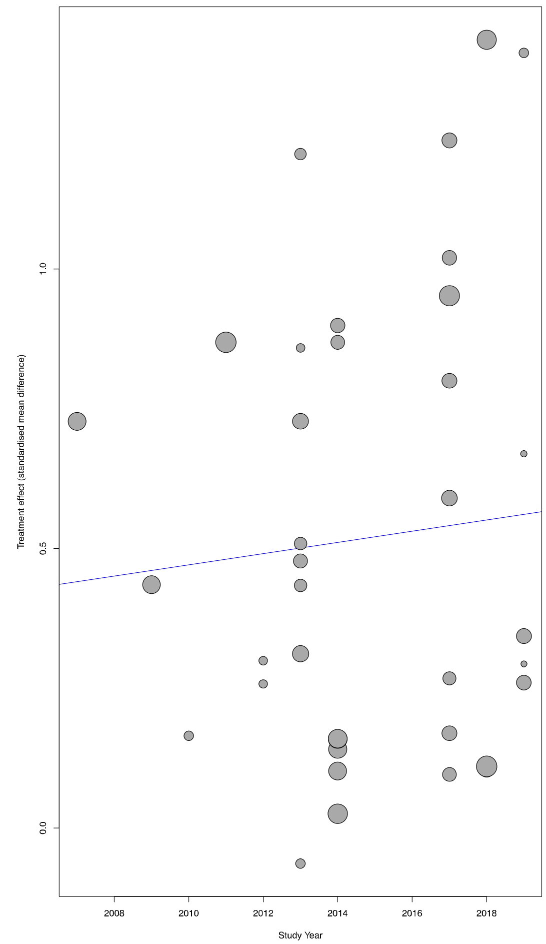
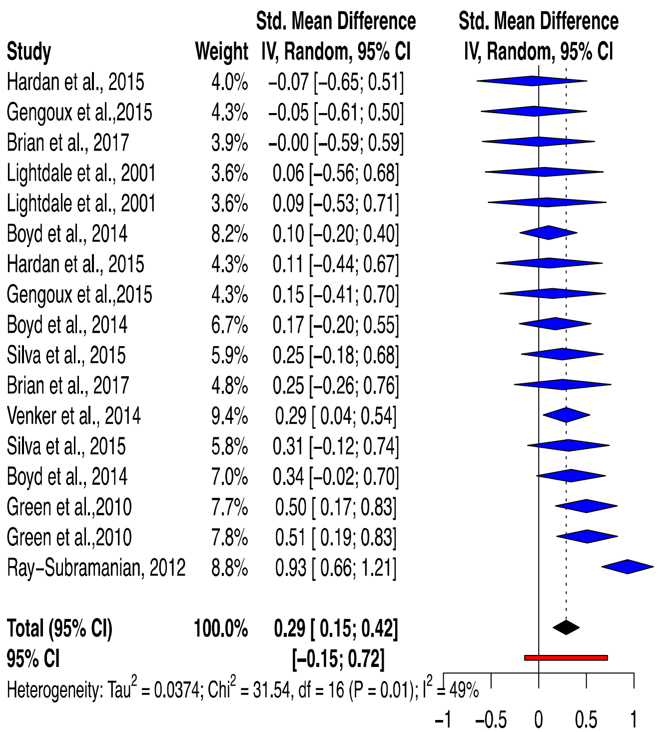
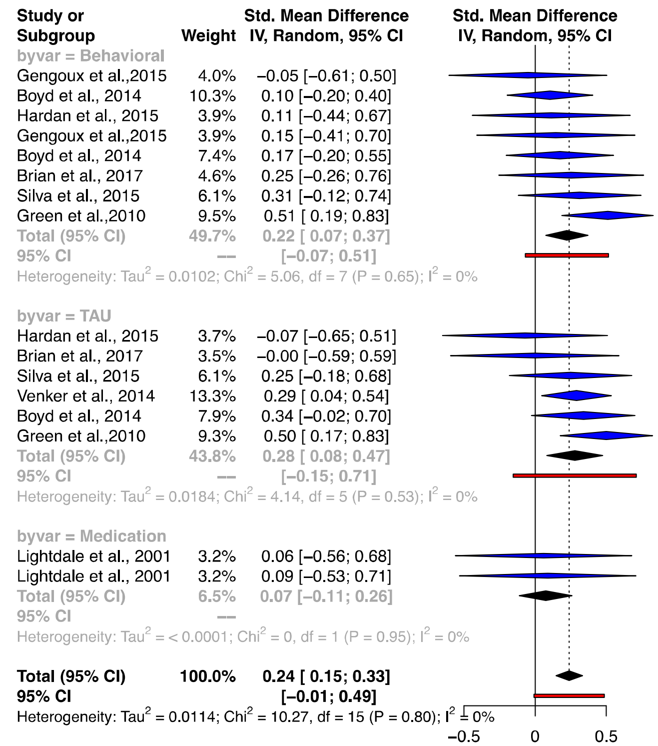


Figure 14: *PLS Expressive Language- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

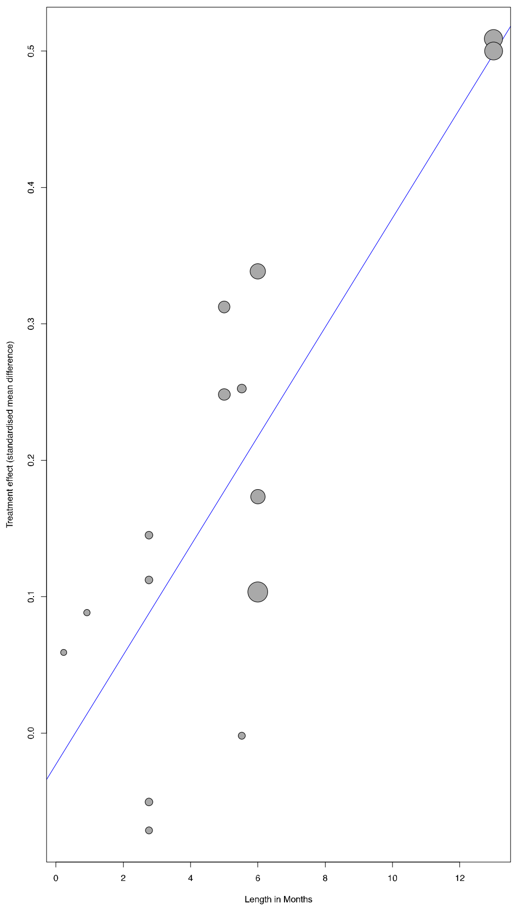
Measure Type	Interventions Received

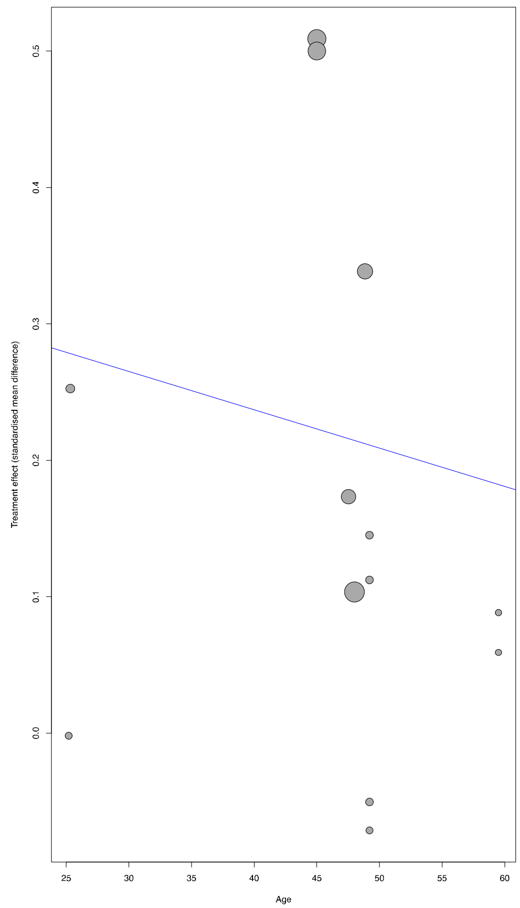




Meta-Regression Bubble Plots

Length	Age	Year





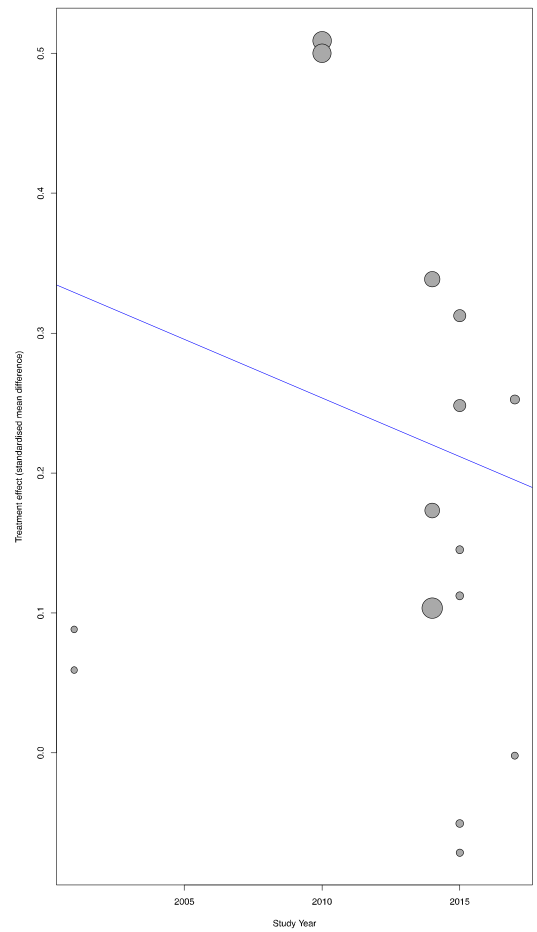
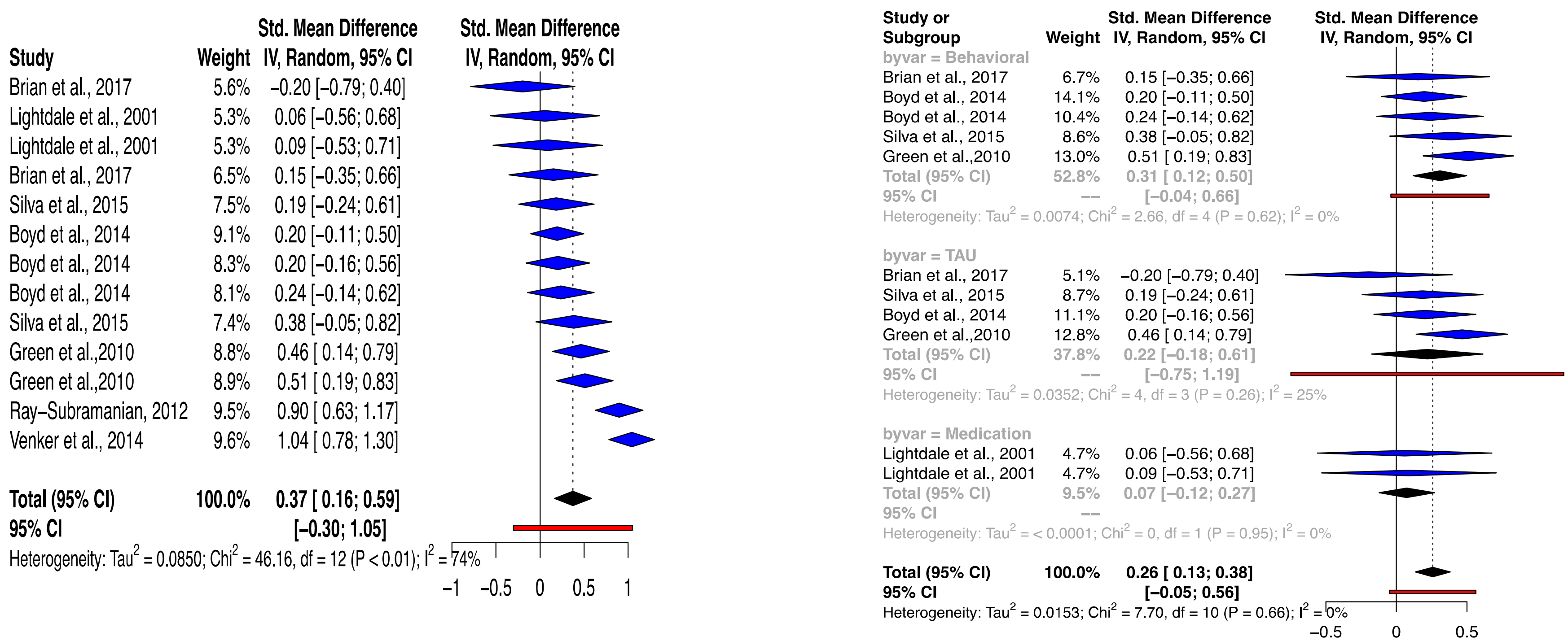


Figure 15: *PLS Receptive Language- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Interventions Received



Meta-Regression Bubble Plots

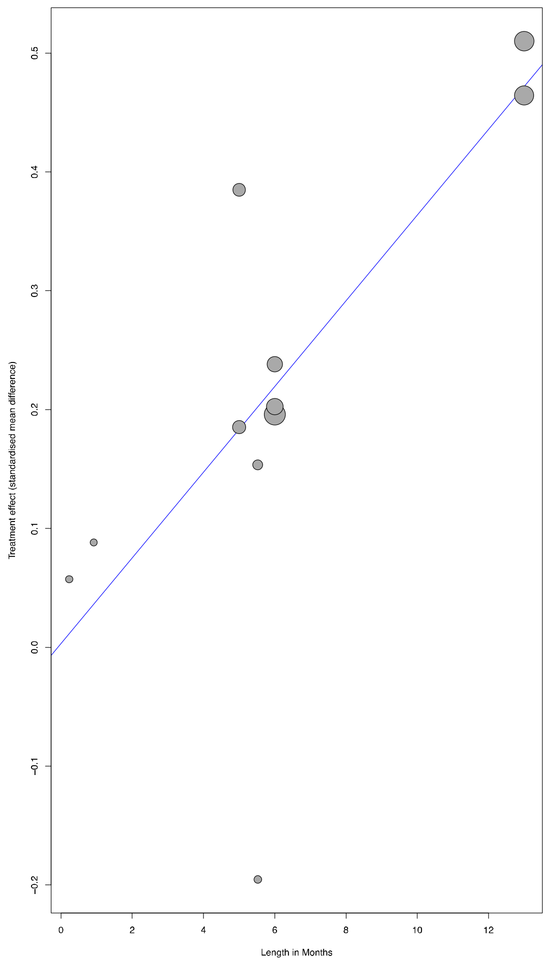
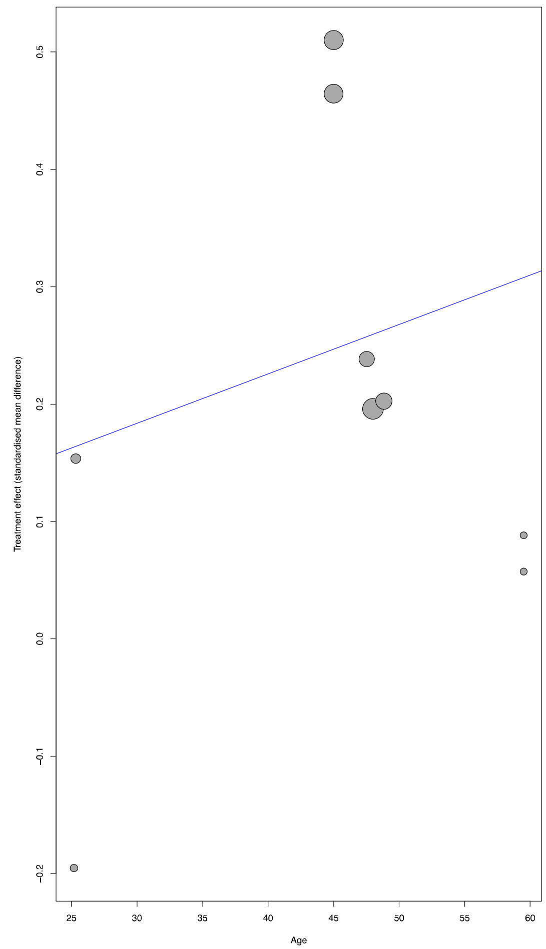
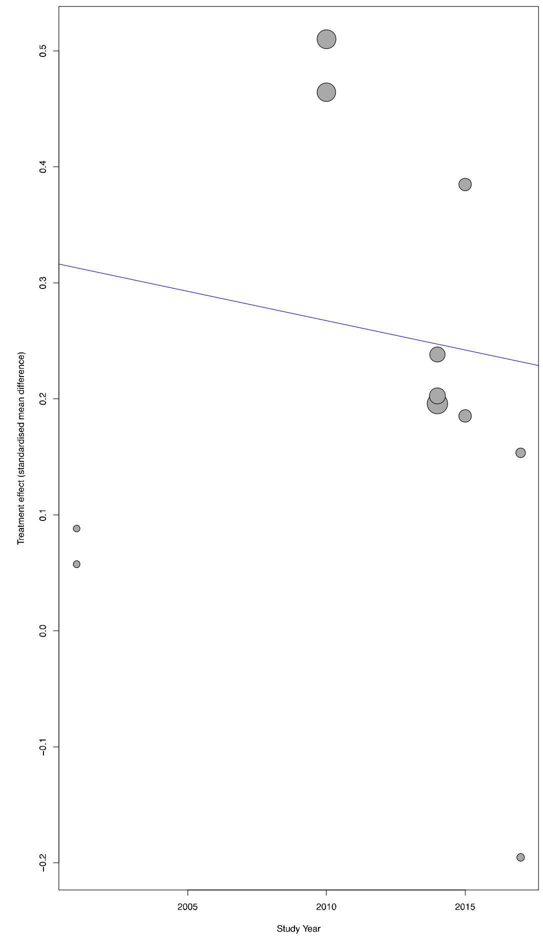
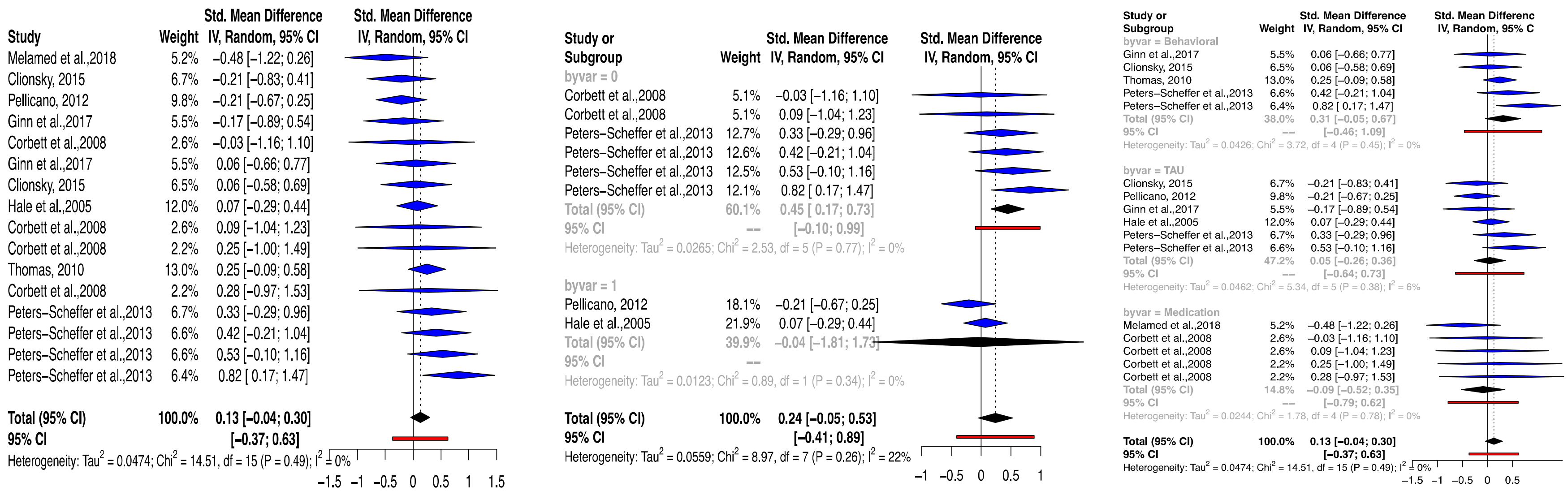
Length	Age	Year
		

Figure 16: *PPVT- Plots*

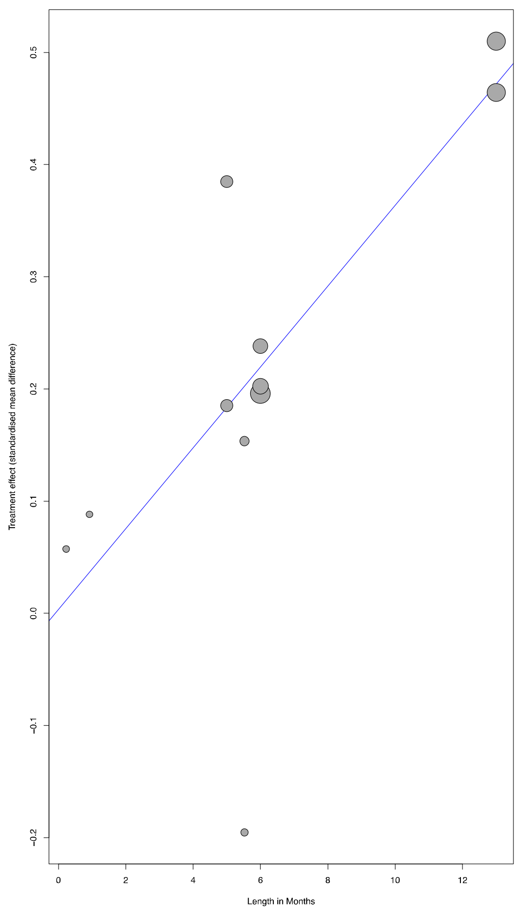
Forest Plots for Sub-Group Analyses and Overall Pooled Effect

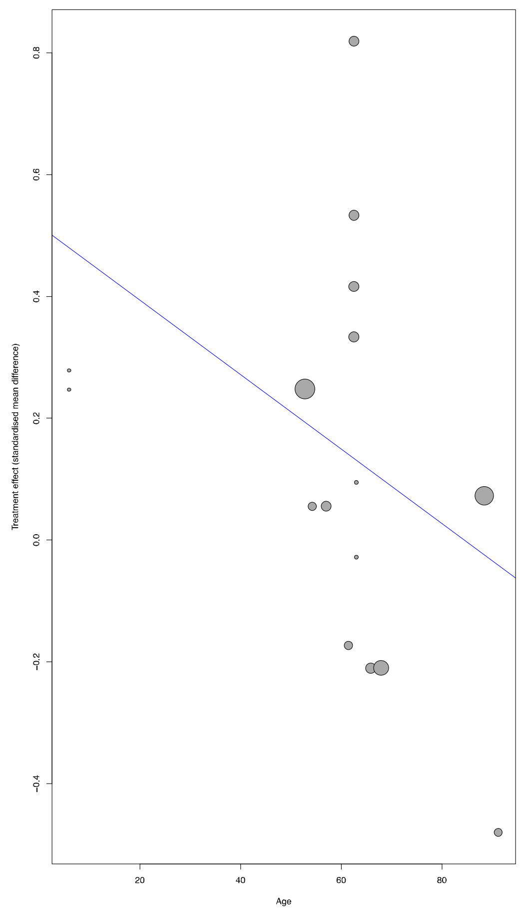
Overall | Cognitive Split | Interventions Received



Meta-Regression Bubble Plots

Length	Age	Year





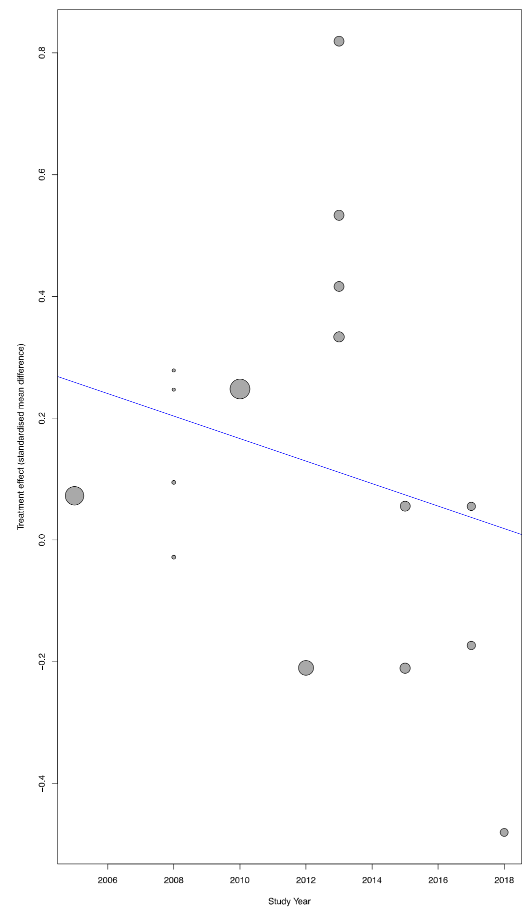
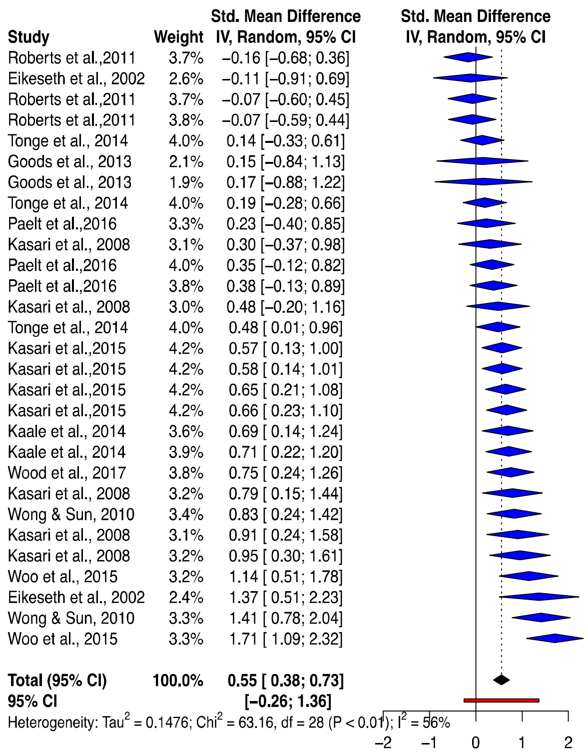
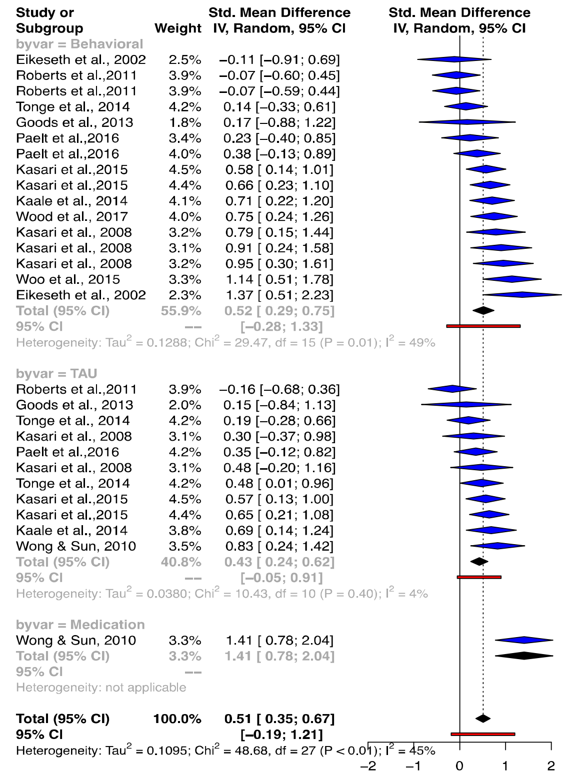


Figure 17: *Reynell Expressive Language- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Interventions Received





Meta-Regression Bubble Plots

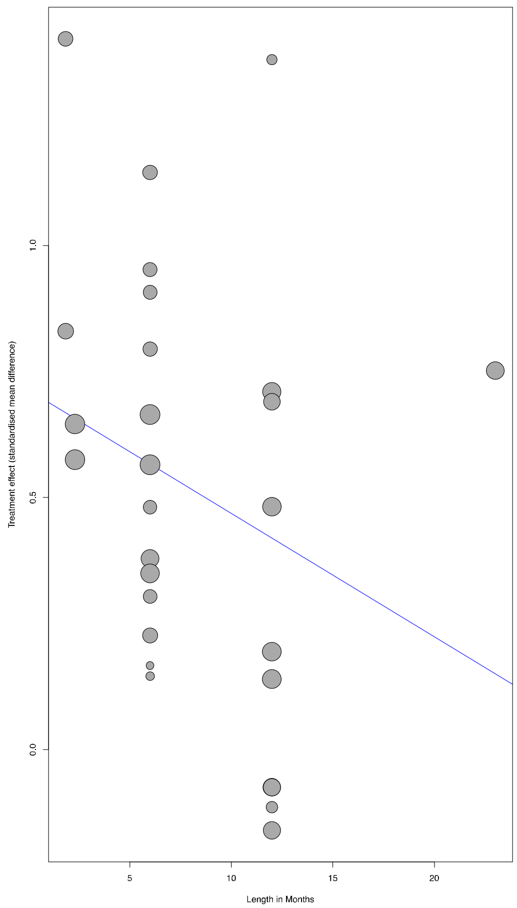
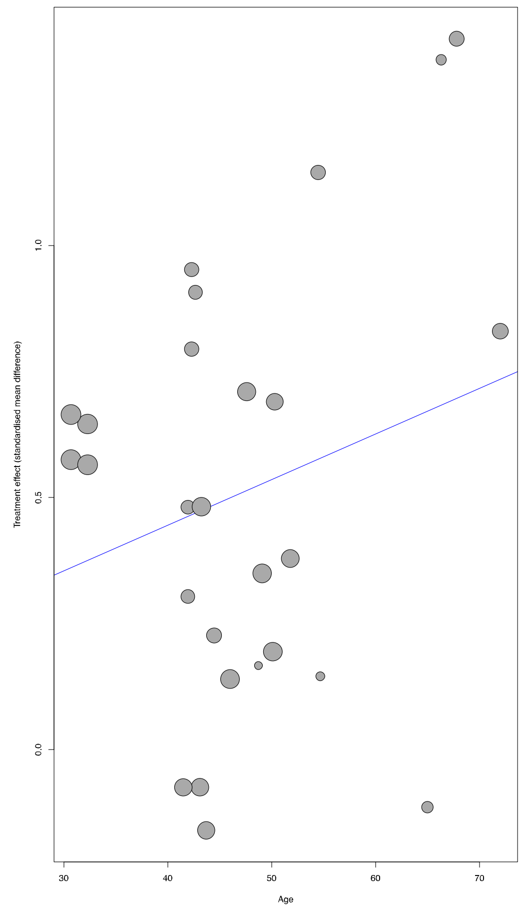
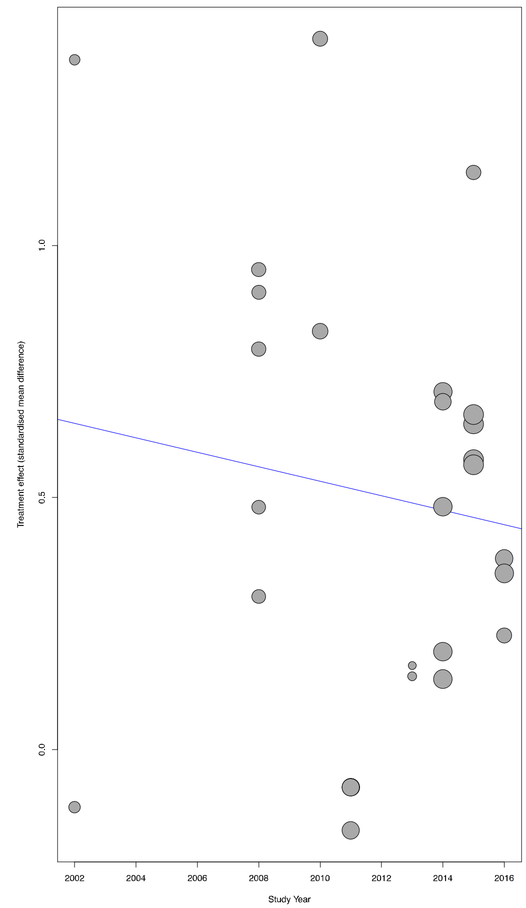
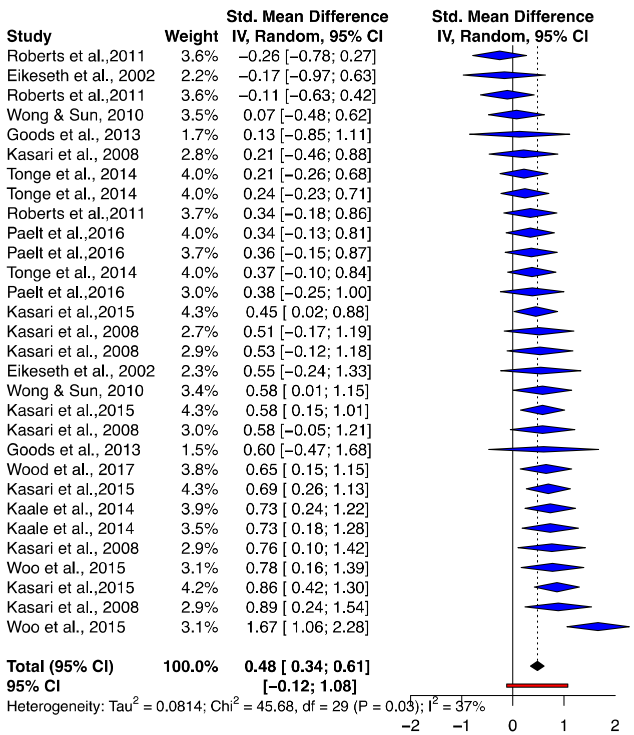
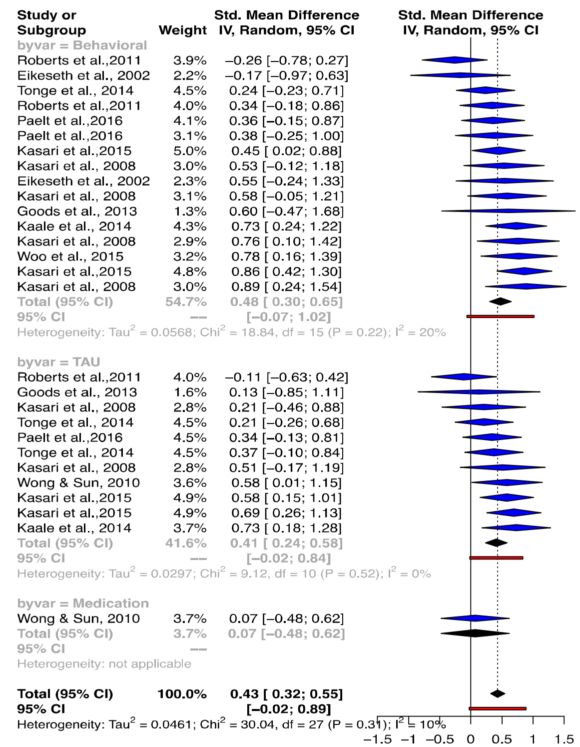
Length	Age	Year
		

Figure 18: *Reynell Receptive Language- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

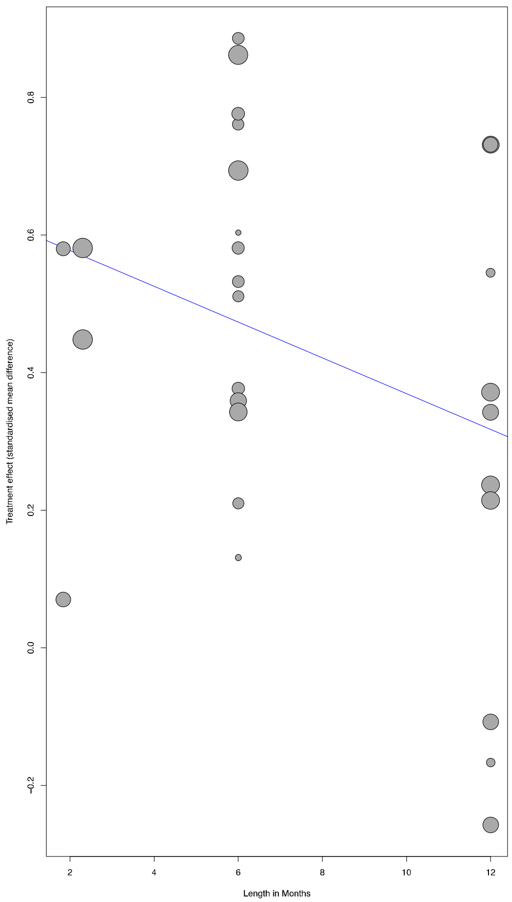
Overall	Interventions Received

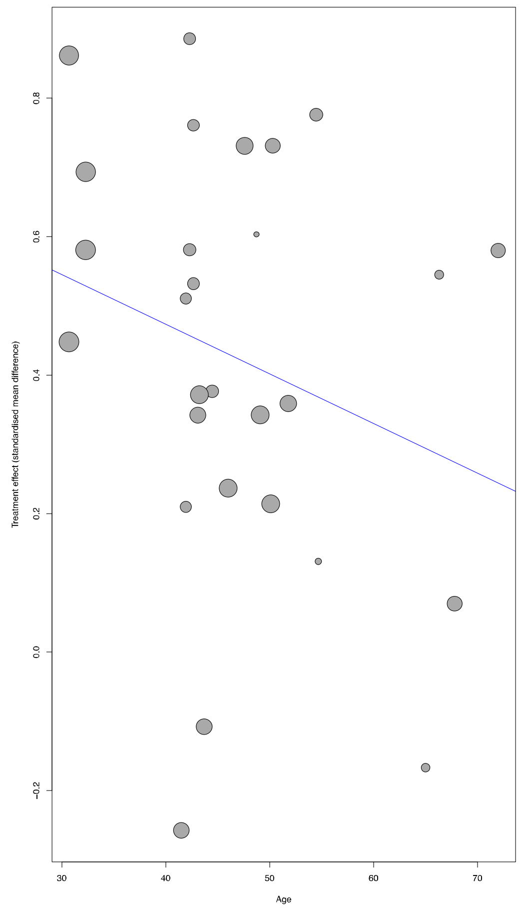




Meta-Regression Bubble Plots

Length	Age	Year





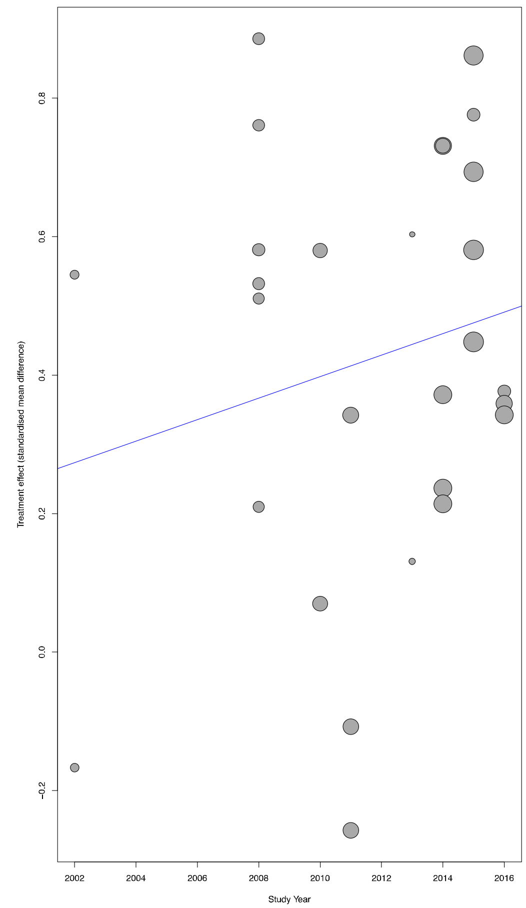
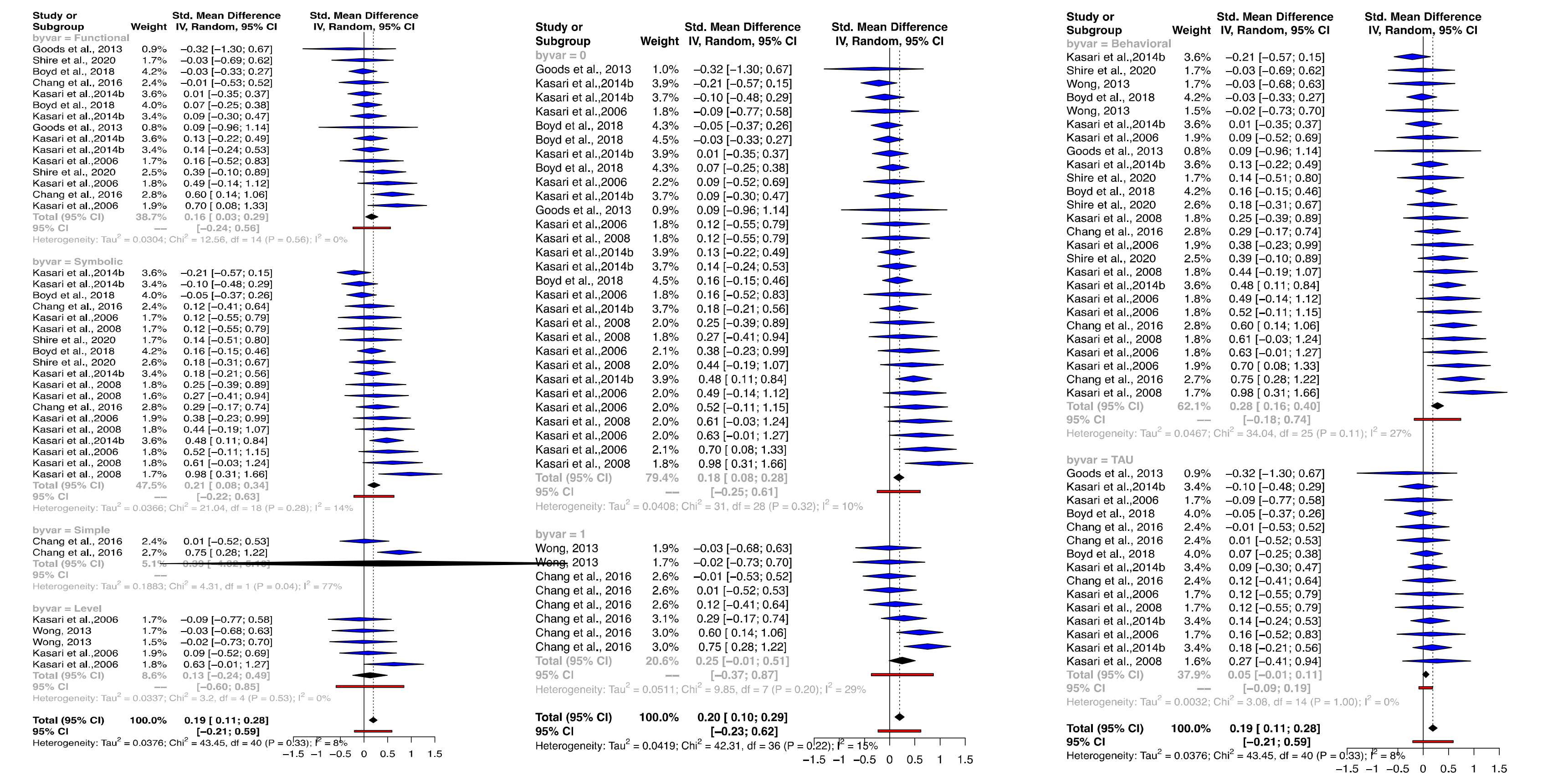


Figure 19: *SPA- Plots*

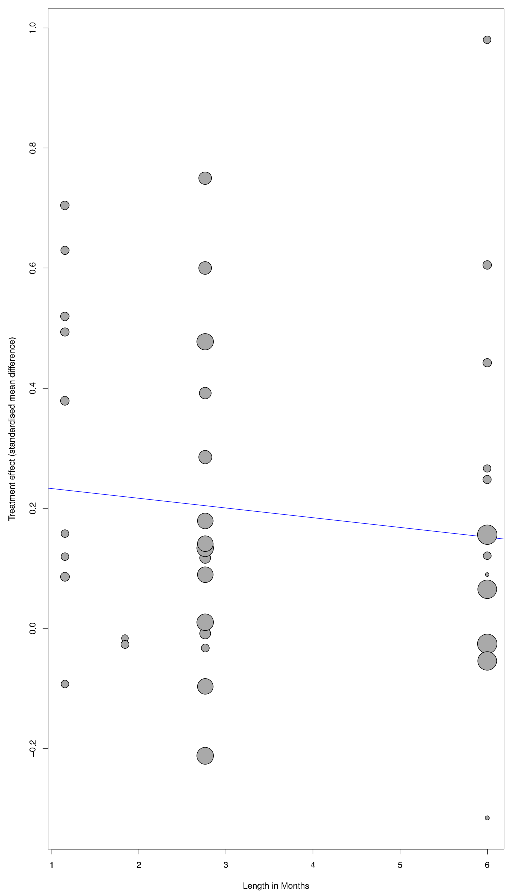
Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall | Cognitive Split | Interventions Received

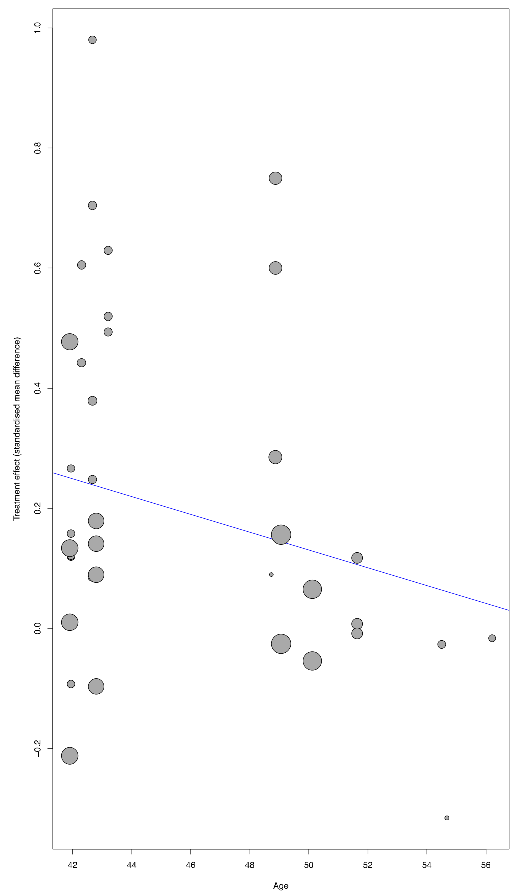


Meta-Regression Bubble Plots

Length



Age



Year

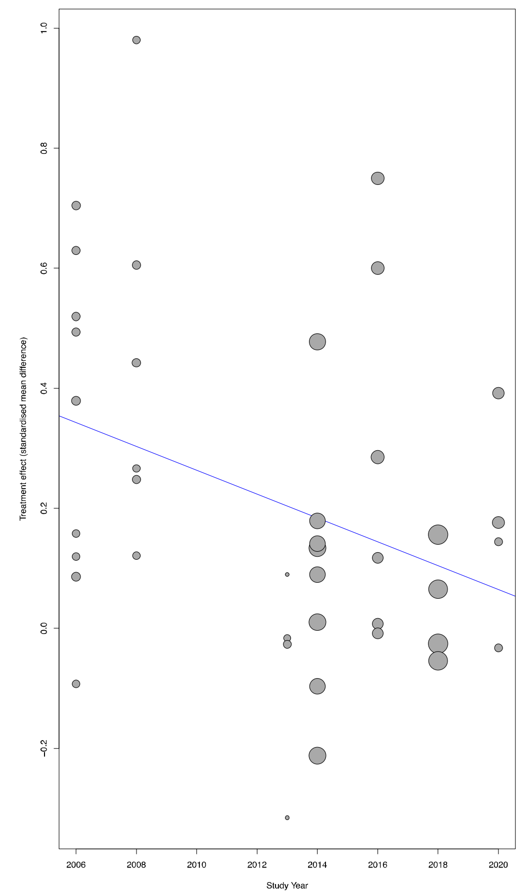
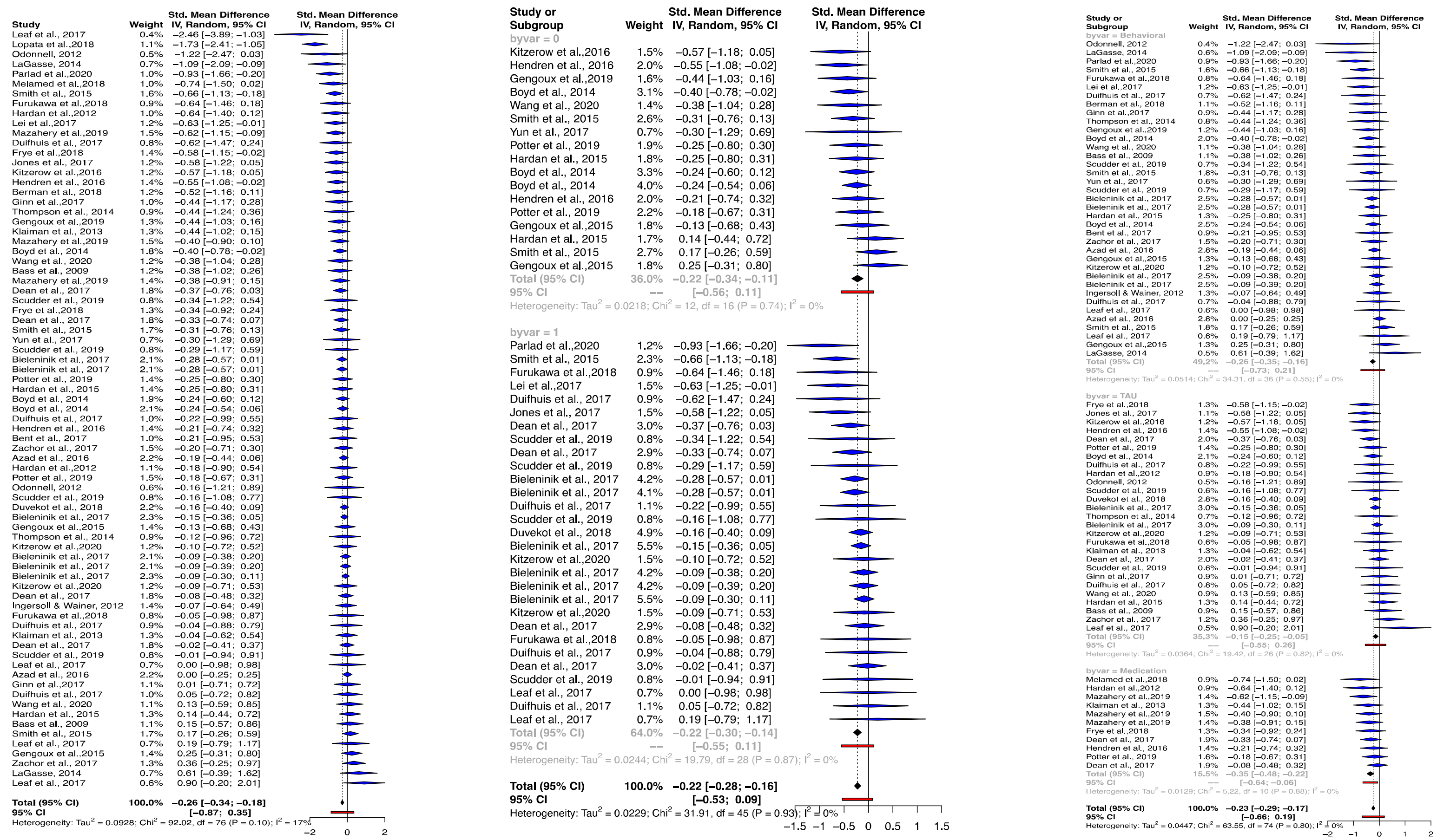


Figure 20: *SRS Total- Plots*

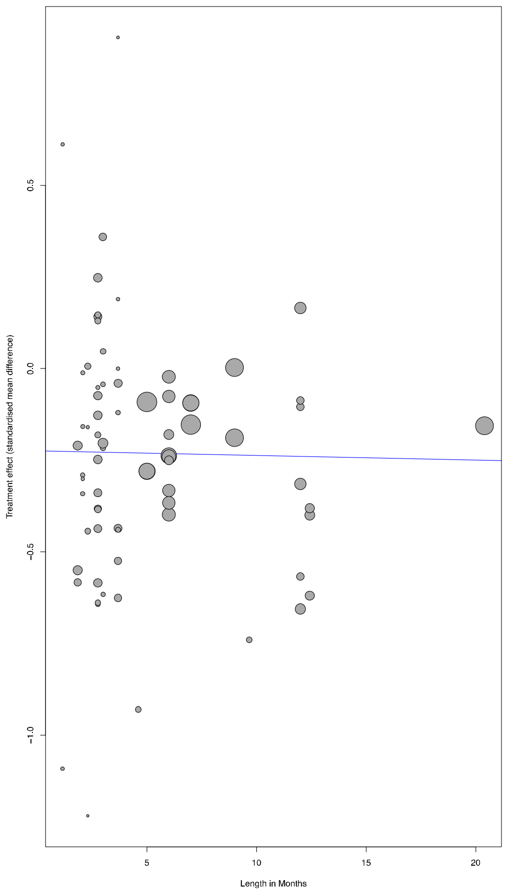
Forest Plots for Sub-Group Analyses and Overall Pooled Effect

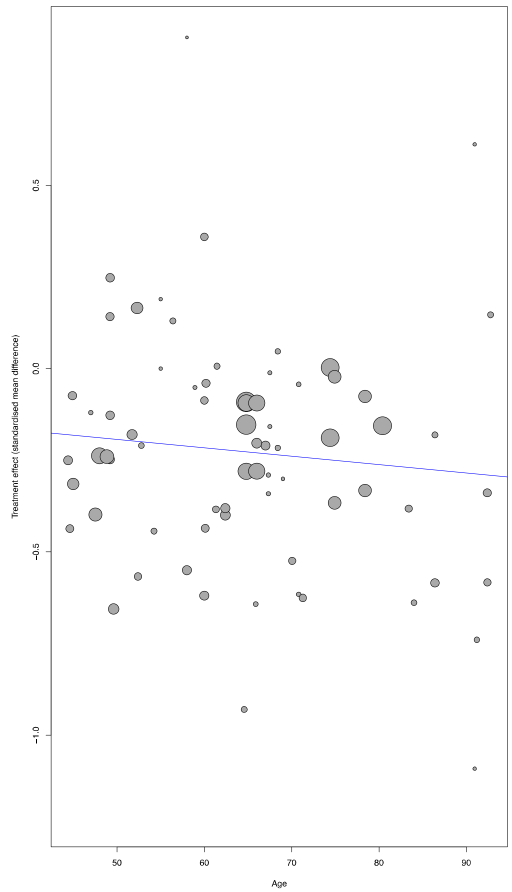
Overall	Cognitive Split	Interventions Received



Meta-Regression Bubble Plots

Length	Age	Year





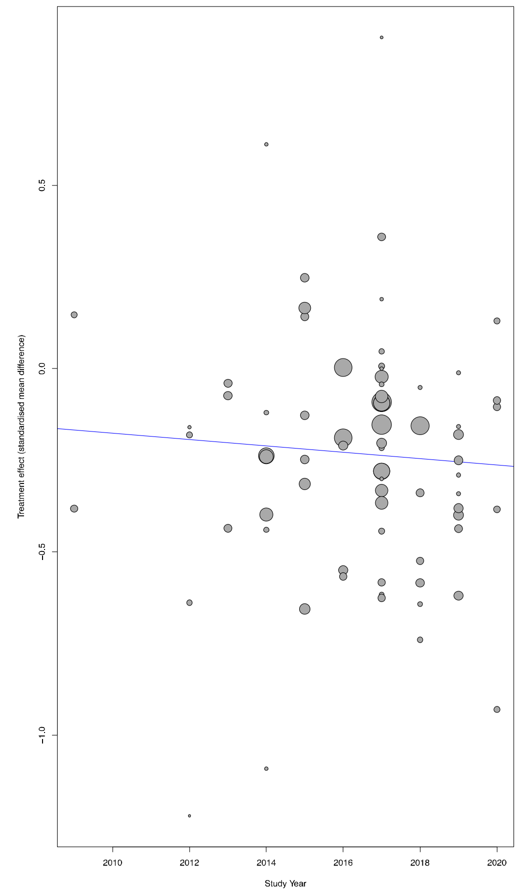
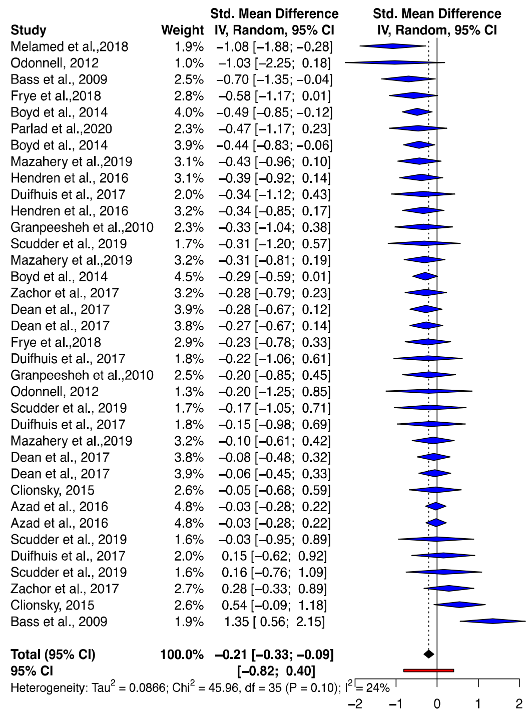


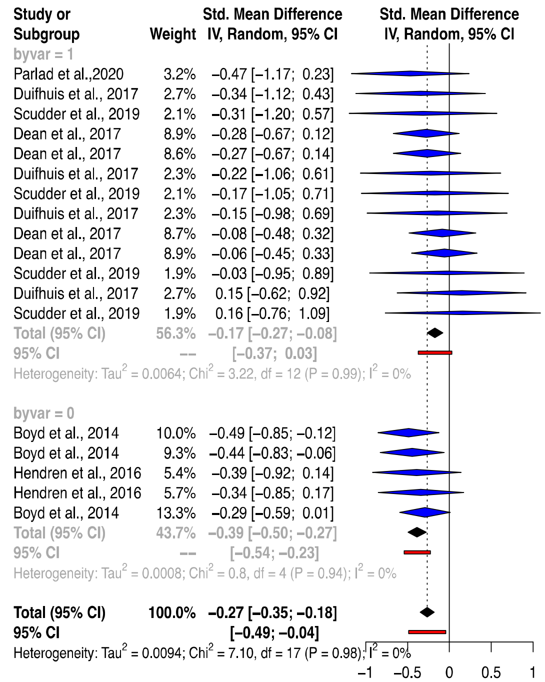
Figure 21: *SRS Cognition- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

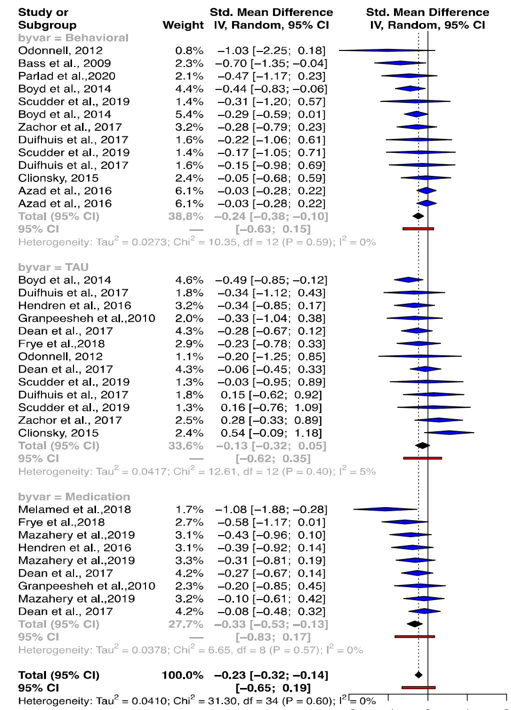
Overall



Cognitive Split

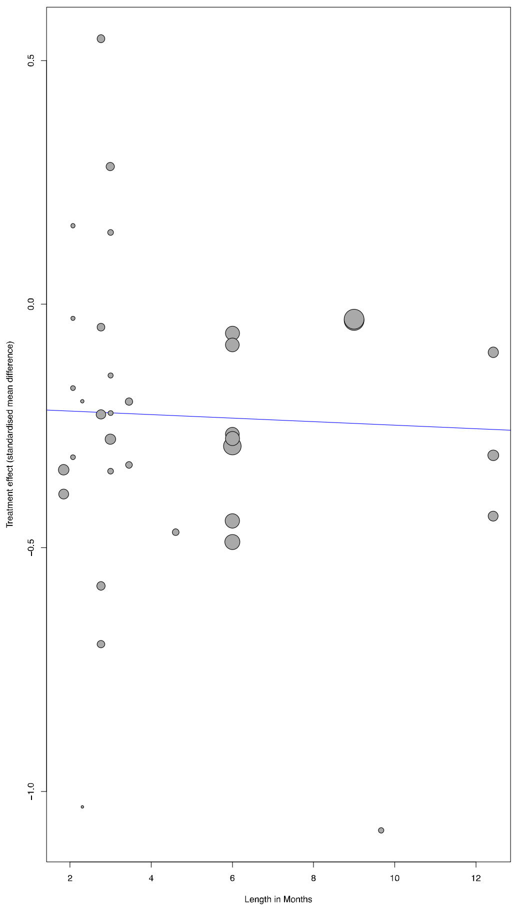


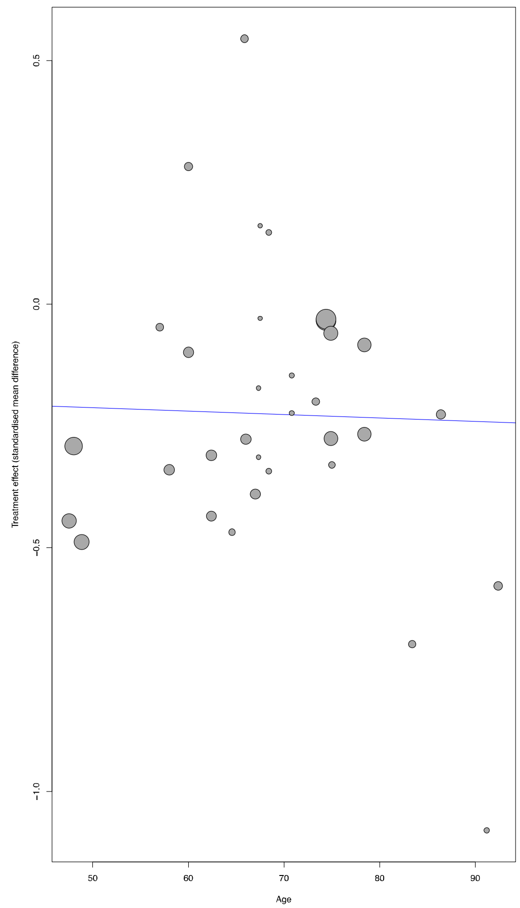
Interventions Received



Meta-Regression Bubble Plots

Length	Age	Year





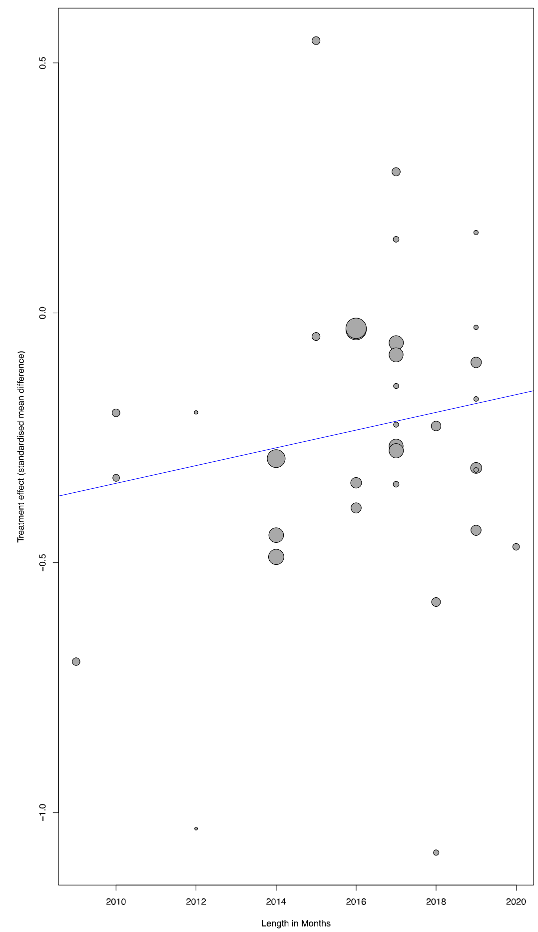
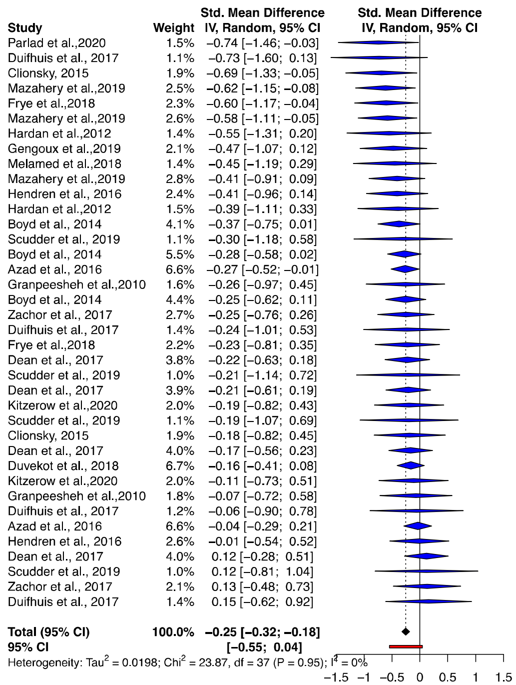
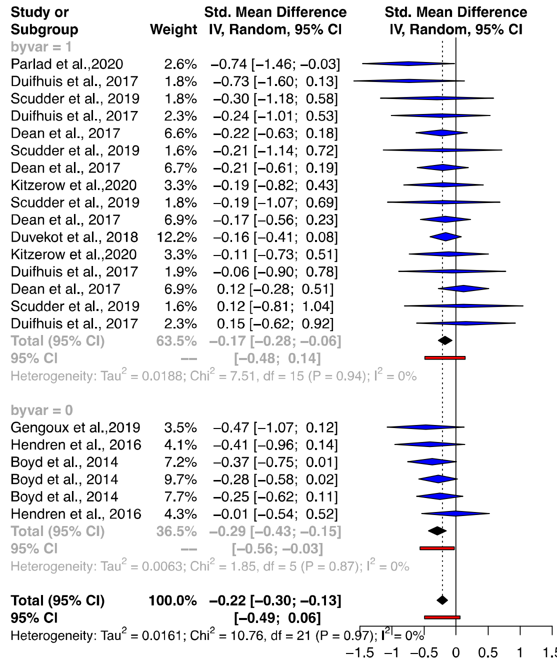


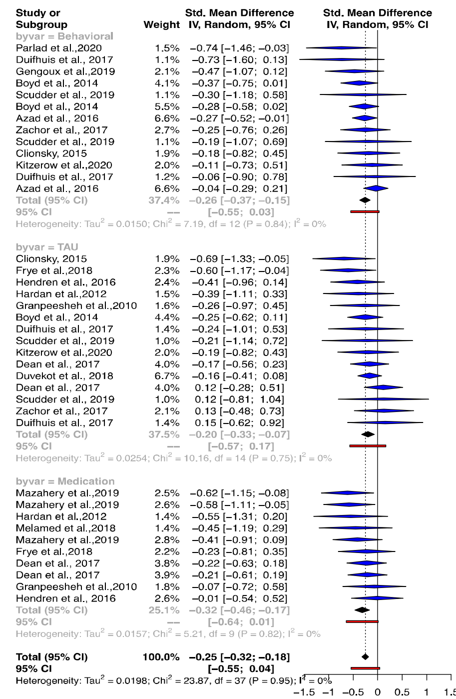
Figure 22: *SRS Social Communication- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Cognitive Split	Interventions Received







Meta-Regression Bubble Plots

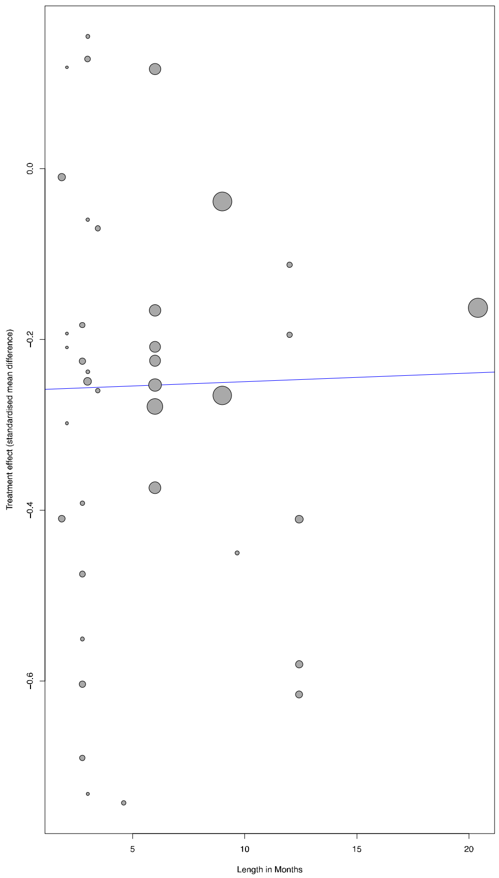
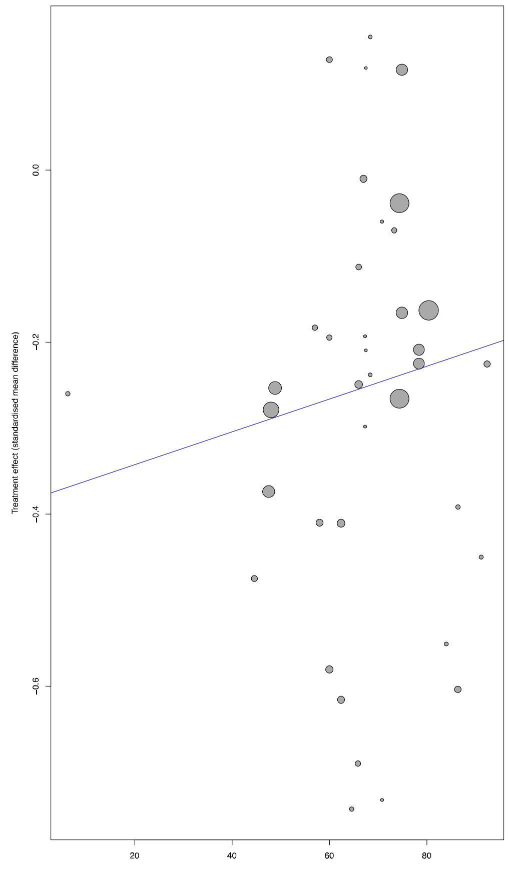
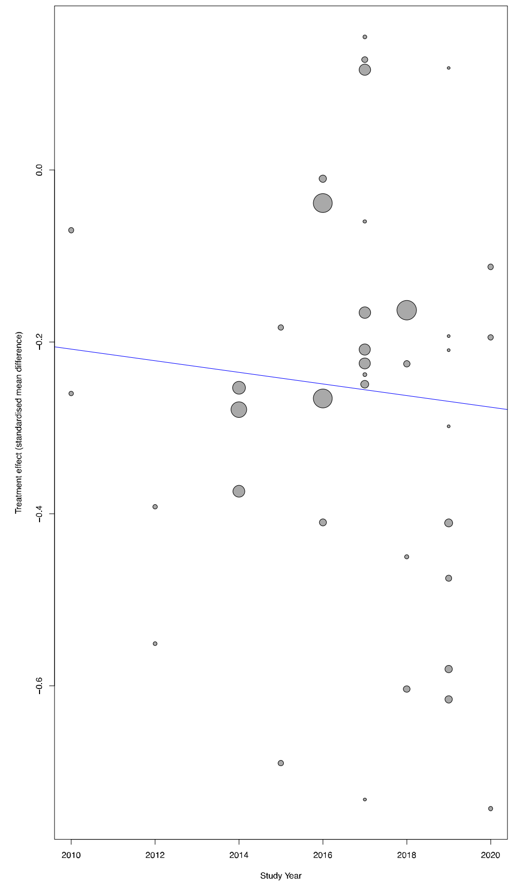
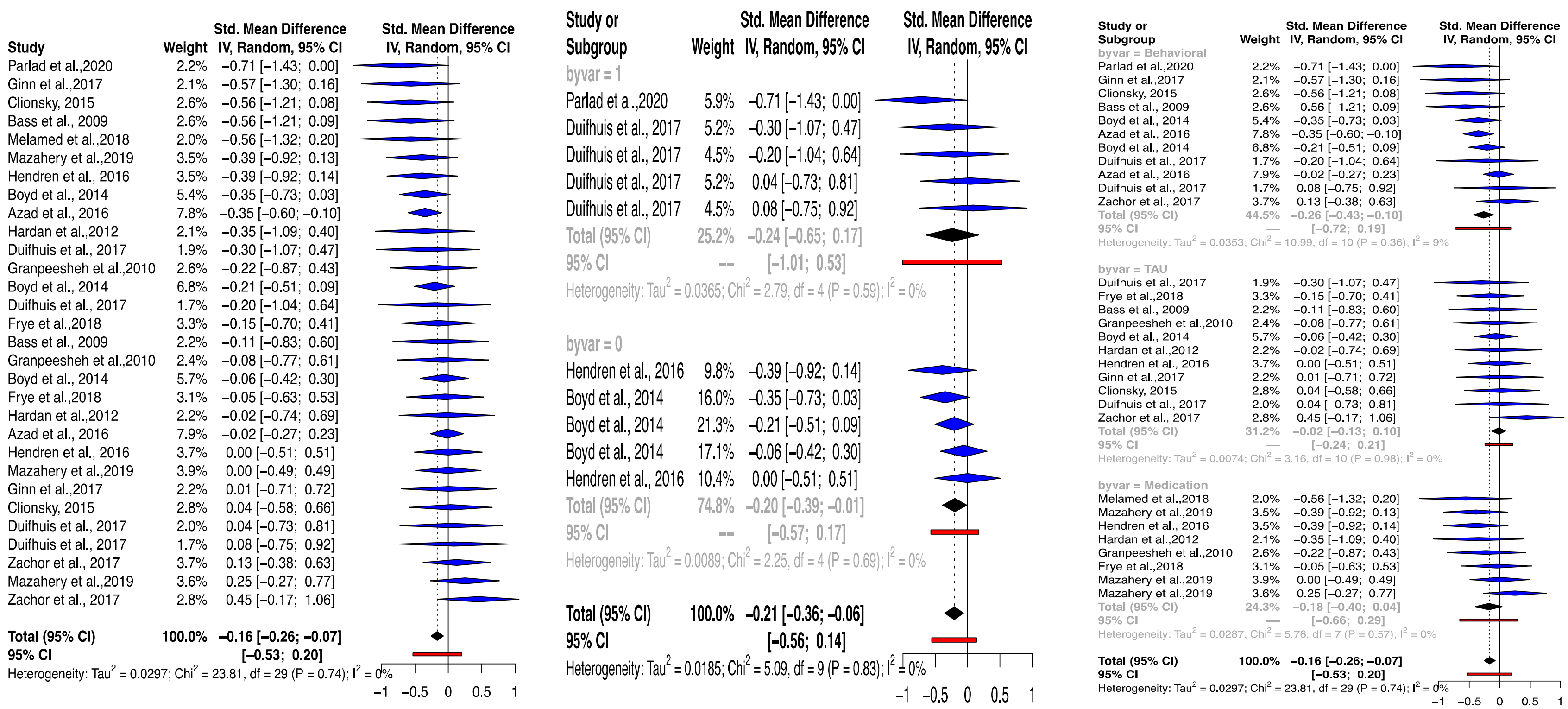
Length	Age	Year
		

Figure 23: *SRS Social Awareness- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Cognitive Split	Interventions Received



Meta-Regression Bubble Plots

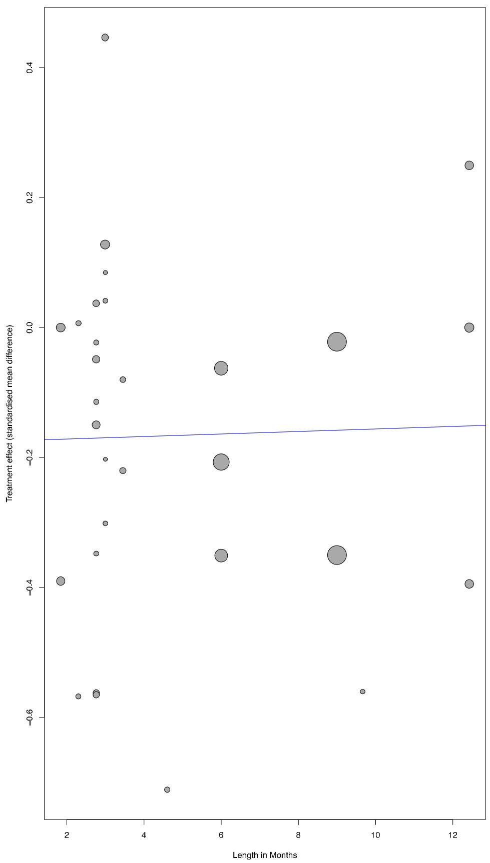
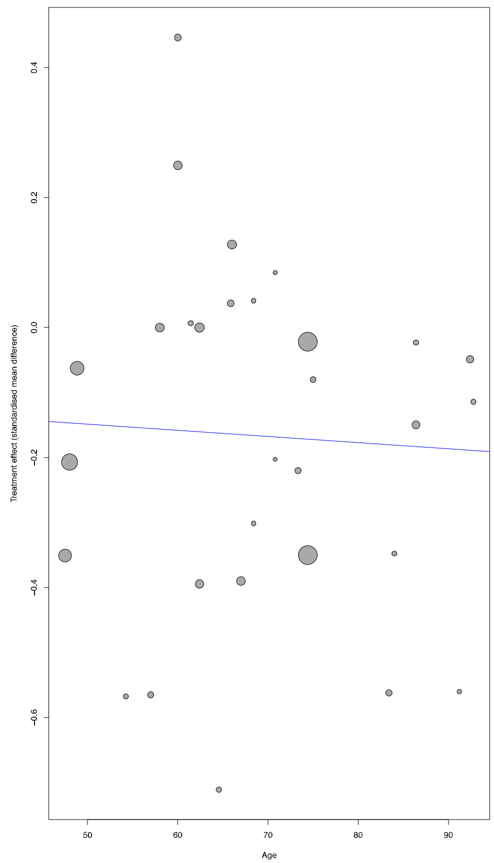
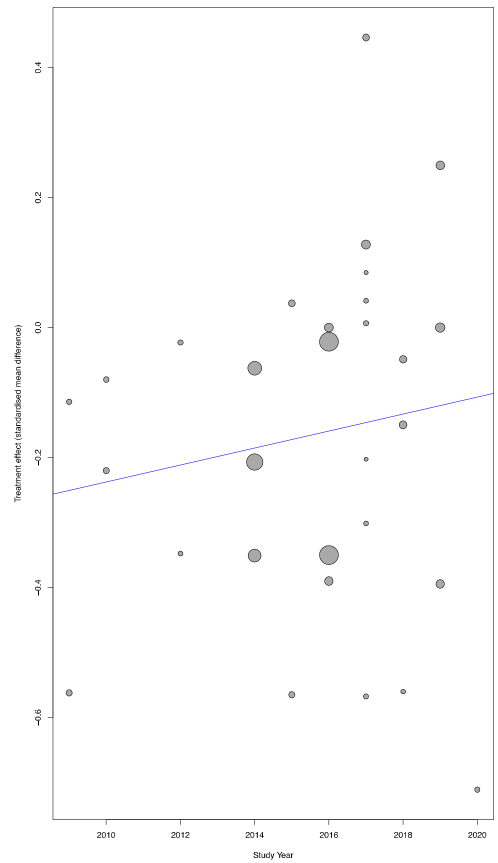
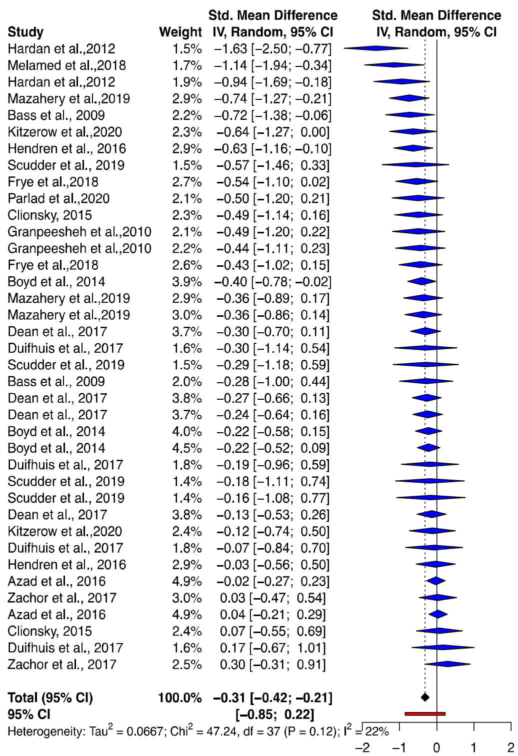
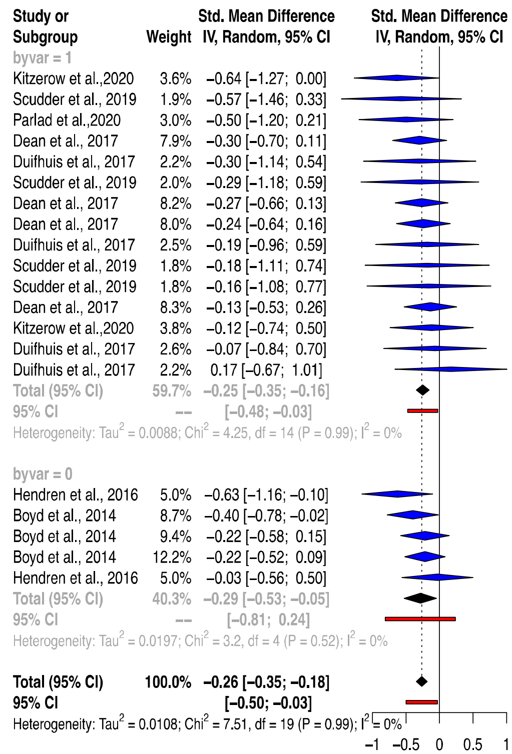
Length	Age	Year
		

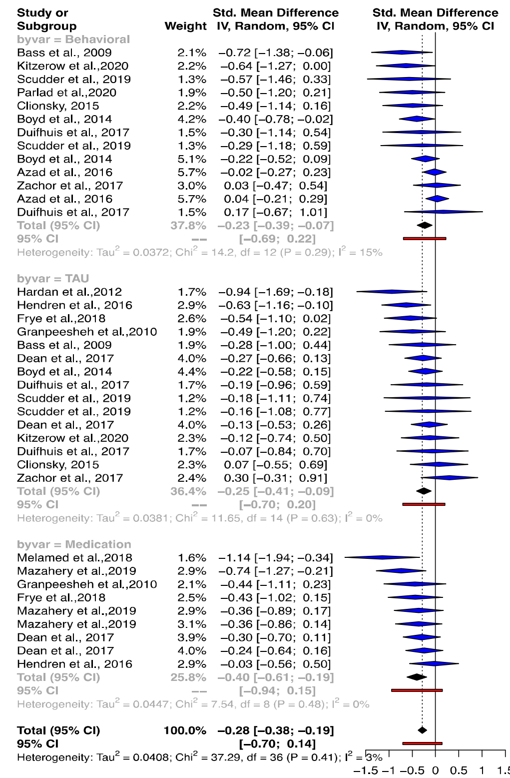
Figure 24: *SRS Social Motivation- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Cognitive Split	Interventions Received







Meta-Regression Bubble Plots

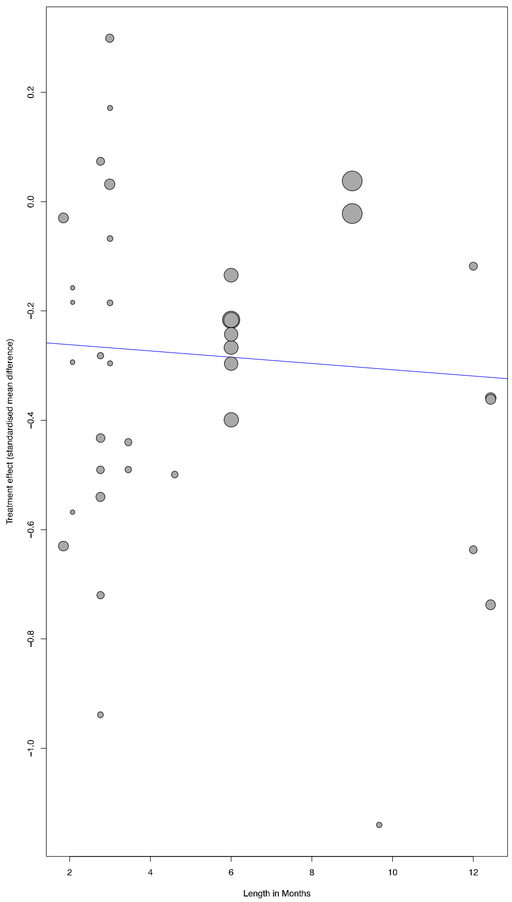
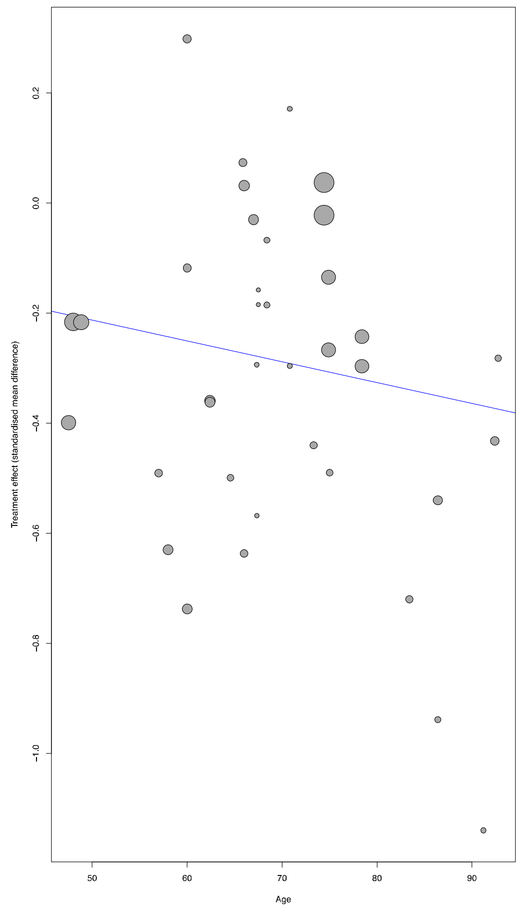
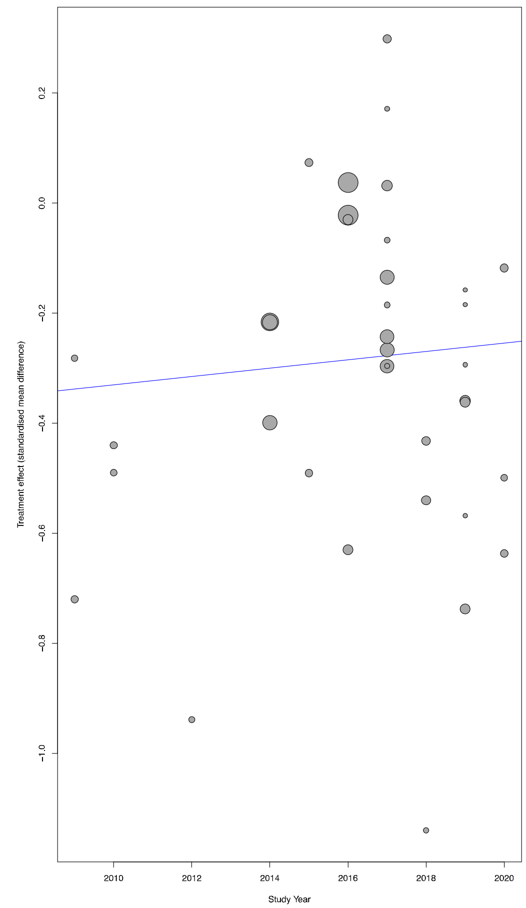
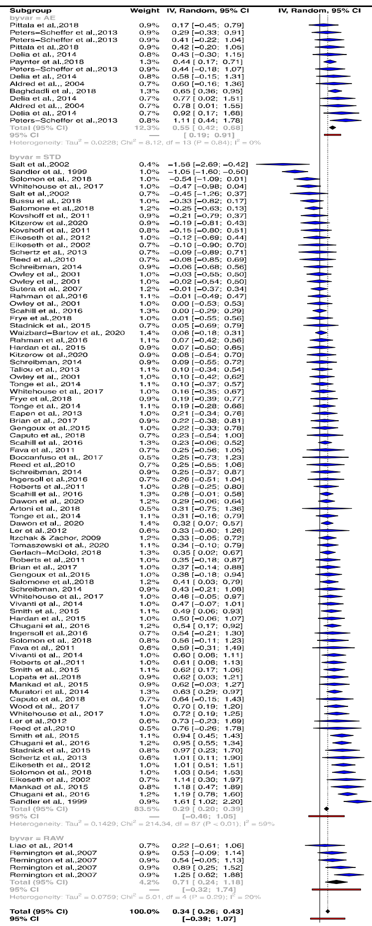
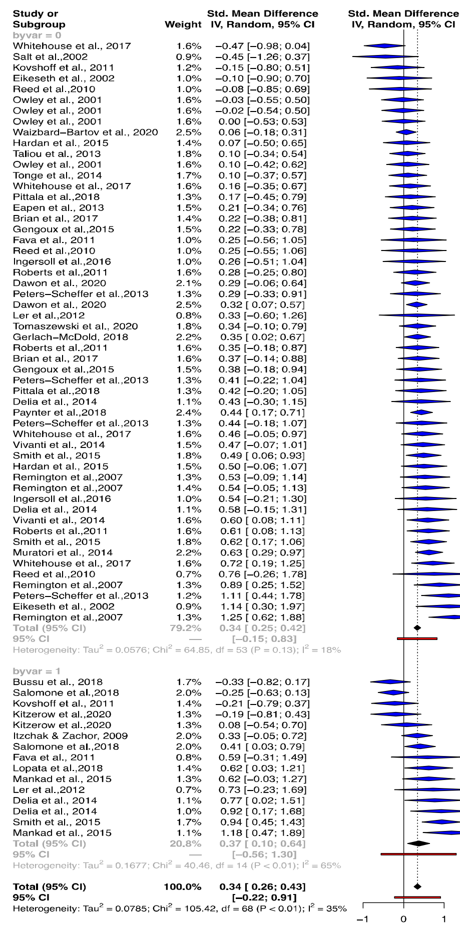
Length	Age	Year
		

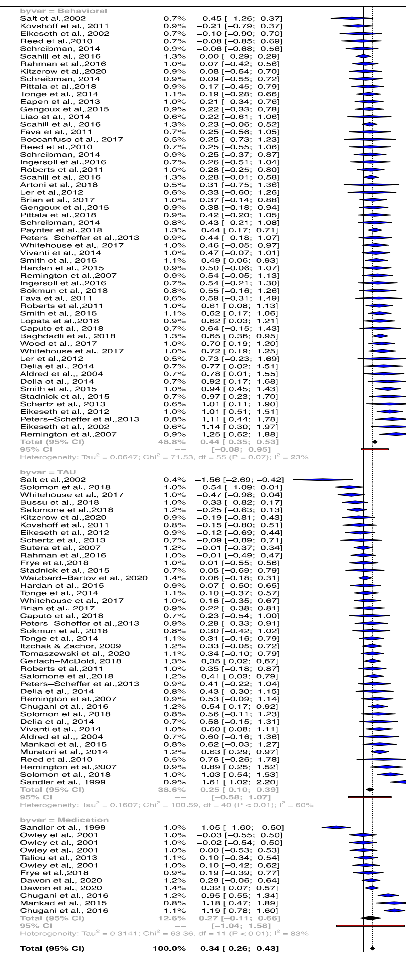
Figure 25: *Vineland Communication- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

Overall	Cognitive Split	Interventions Received

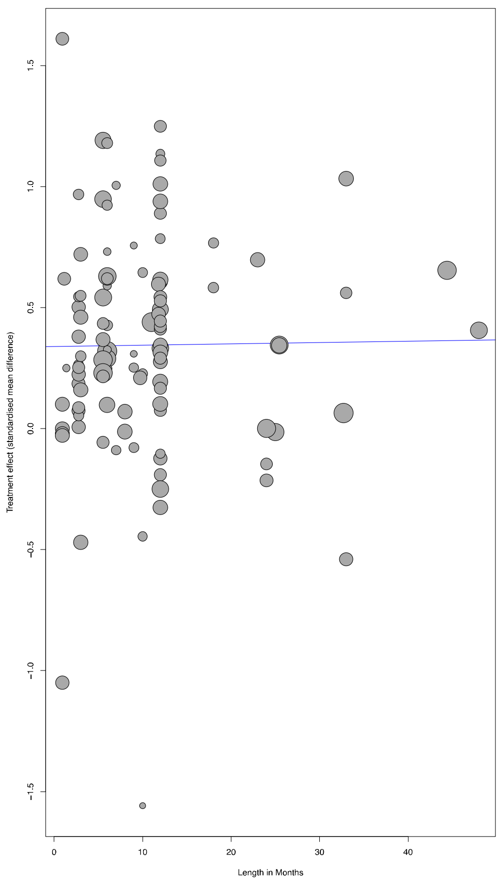


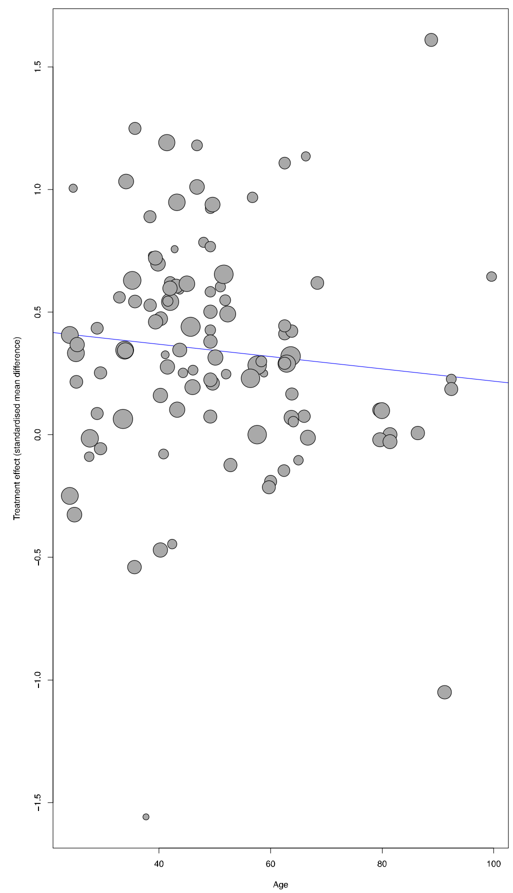




Meta-Regression Bubble Plots

Length	Age	Year





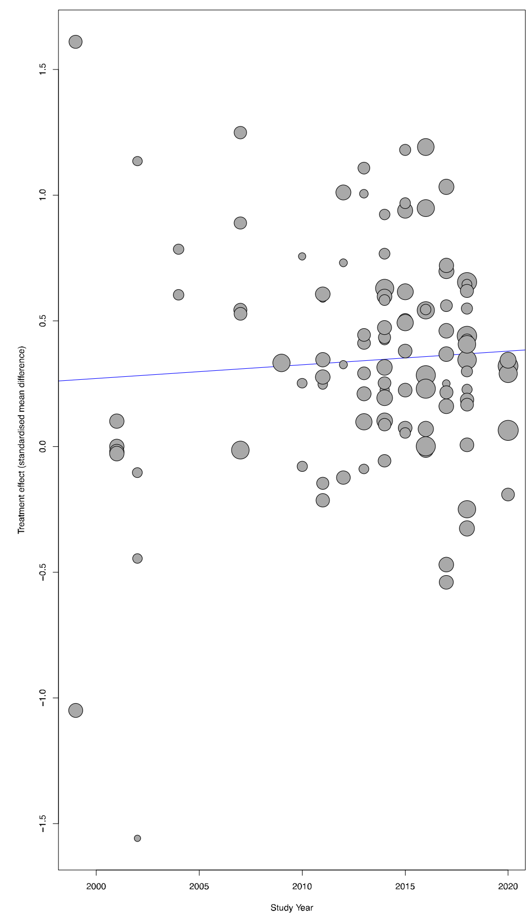
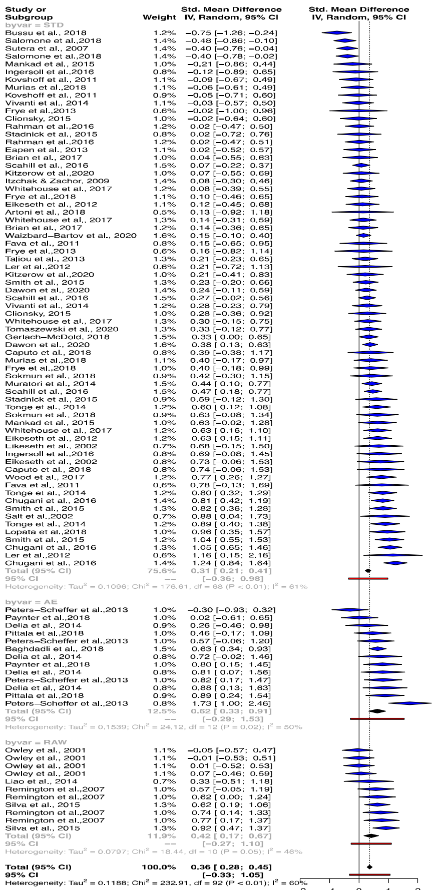
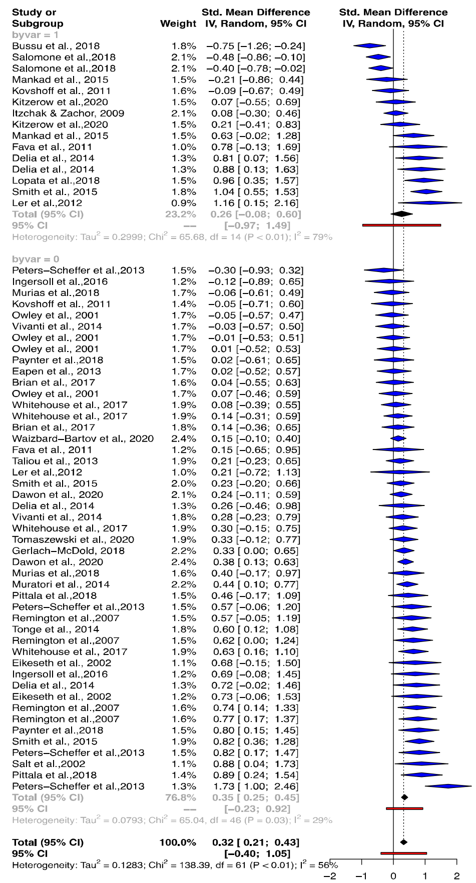


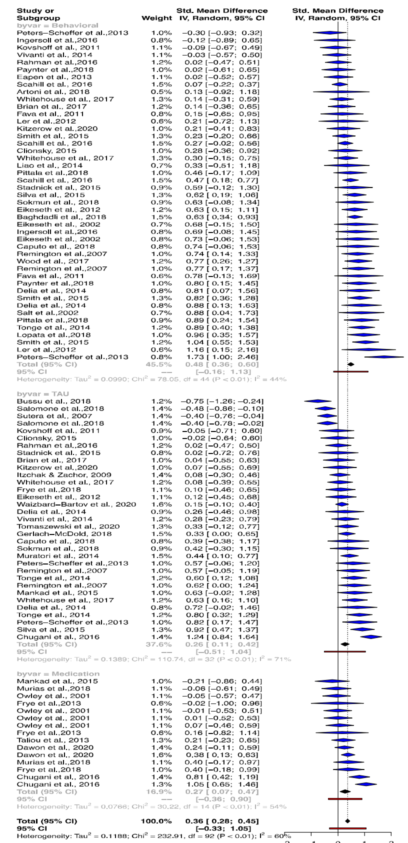
Figure 26: *Vineland Socialization- Plots*

Forest Plots for Sub-Group Analyses and Overall Pooled Effect

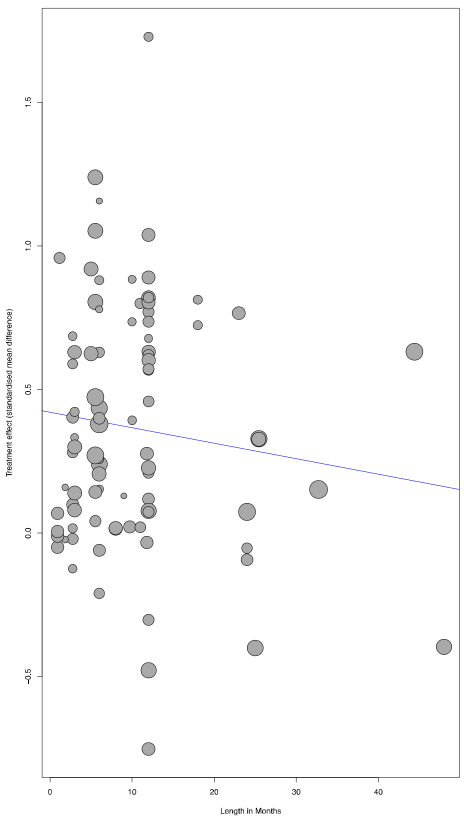
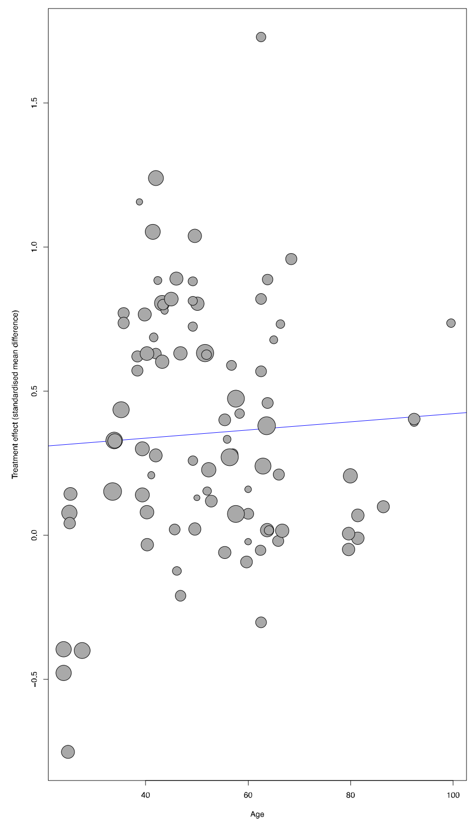
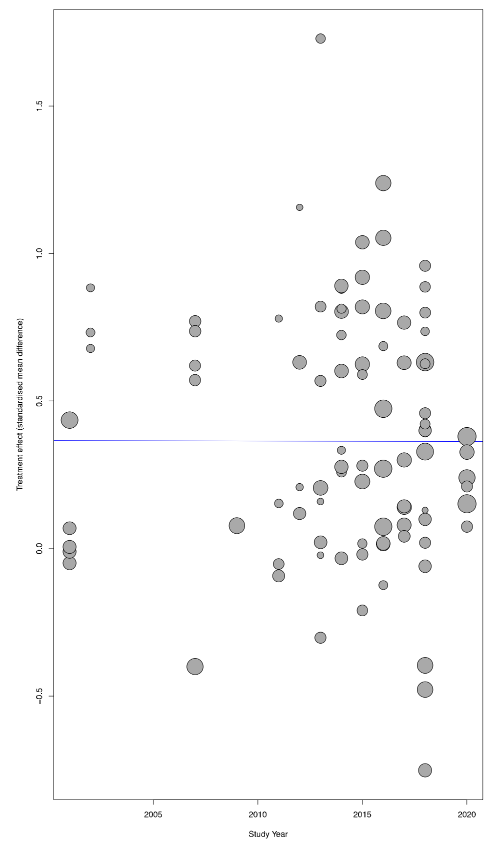
Type of Measure	Cognitive Split	Interventions Received







Meta-Regression Bubble Plots

Length	Age	Year
		

Appendix A- Supplementary Material

A- 1: *Search Terms and Strategy*

The syntax used was modified for each individual database. The search syntax below represents a title and abstract search for the PsychINFO Database. First, broken up for clarity and then provided as it was entered.

Syntax broken up by category
"Pervasive development* disorder*" OR Autis* OR PDD OR PDD-NOS OR Asperg* OR asd
AND
(Child* OR infan* OR kindergarten* OR pediatric OR toddler OR pre-school* OR preschool* OR "primary school*" OR "elementary school*")
AND
("free play" OR "parent child interaction" OR "caregiver child interaction" OR "caregiver play interaction" OR PCX OR CCX OR "mother child interaction" OR "behavioral rating" OR "MacArthur Communicative Development Inventory" OR MCDI OR "Griffiths Scale of Infant Development" OR VABS OR "Vineland Adaptive Behavior Scale" OR "preschool language scales" OR PLS OR "Communication symbolic behavior scales" OR CSBS OR "Mullen Scales of Early Learning" OR MSEL OR "Leiter International Performance Scale" OR Leiter OR "Early intervention developmental profile" OR EIDP OR "Preschool Developmental Learning Accomplishments Profile" OR "Comprehensive Assessment of Spoken Language" OR CASL OR GMDS OR "Griffiths Mental Development Scales" OR BPVS OR "British Picture Vocabulary Scale" OR SICD OR "Sequenced Inventory of Communicative Development" OR DANVA OR "Diagnostic Analysis of Non-Verbal Accuracy" OR WASI OR "Wechsler Abbreviated Scale of Intelligence" OR SSRS OR "Social Skills Rating System" OR JAMES OR "Joint Attention Measure from ESCS" OR "Early Social Communication Scales" OR ESCS OR "Natural Language Sample" OR "structured play assessment" OR SPA OR "Social Communication Questionnaire" OR SCQ OR "Social responsiveness scale" OR "Imitation Battery" OR "Imitation disorders evaluation scale" OR "pre-verbal communication schedule" OR "social communication behavior codes" OR "Clinical Evaluation of Language Fundamentals" OR CELF OR "Expressive one-word picture vocabulary test" OR EOWPVT OR "receptive one-word picture vocabulary test" OR ROWPVT OR "Illinois Test of Psycholinguistic Abilities" OR "Peabody picture vocabulary test" OR PPVT OR "Battelle Developmental Inventory" OR "Bayley Scales of Infant Development" OR "Stanford-Binet Intelligence Scale" OR "autism diagnostic observation schedule" OR ADOS)

Combined Search Syntax
TS=("Pervasive development* disorder*" OR Autis* OR PDD OR PDD-NOS OR Asperg* OR asd) "AND" (Child* OR infan* OR kindergarten* OR pediatric OR toddler OR pre-school* OR preschool* OR "primary school*" OR "elementary school*") "AND" ("free play" OR "parent child interaction" OR "caregiver child interaction" OR "caregiver play interaction" OR PCX OR CCX OR "mother child interaction" OR "behavioral rating" OR "MacArthur Communicative Development Inventory" OR MCDI OR "Griffiths Scale of Infant Development" OR VABS OR "Vineland Adaptive Behavior Scale" OR "preschool language scales" OR PLS OR "Communication symbolic behavior scales" OR CSBS OR "Mullen Scales of Early Learning" OR MSEL OR "Leiter International Performance Scale" OR Leiter OR "Early intervention developmental profile" OR EIDP OR "Preschool Developmental Learning Accomplishments Profile" OR "Comprehensive Assessment of Spoken Language" OR CASL OR GMDS OR "Griffiths Mental Development Scales" OR BPVS OR "British Picture Vocabulary Scale" OR SICD OR "Sequenced Inventory of Communicative Development" OR DANVA OR "Diagnostic Analysis of Non-Verbal Accuracy" OR WASI OR "Wechsler Abbreviated Scale of Intelligence" OR SSRS OR "Social Skills Rating System" OR JAMES OR "Joint Attention Measure from ESCS" OR "Early Social Communication Scales" OR ESCS OR "Natural Language Sample" OR "structured play assessment" OR SPA OR "Social Communication Questionnaire" OR SCQ OR "Social responsiveness scale" OR "Imitation Battery" OR "Imitation disorders evaluation scale" OR "pre-verbal communication schedule" OR "social communication behavior codes" OR "Clinical Evaluation of Language Fundamentals" OR CELF OR "Expressive one-word picture vocabulary test" OR EOWPVT OR "receptive one-word picture vocabulary test" OR ROWPVT OR "Illinois Test of Psycholinguistic Abilities" OR "Peabody picture vocabulary test" OR PPVT OR "Battelle Developmental Inventory" OR "Bayley Scales of Infant Development" OR "Stanford-Binet Intelligence Scale" OR "autism diagnostic observation schedule" OR ADOS)

A- 2: Quality Assessment Tool

Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies

Criteria	**Yes**	**No**	**Other (CD, NR, NA)***
1. Was the research question or objective in this paper clearly stated?			
2. Was the study population clearly specified and defined?			
3. Was the participation rate of eligible persons at least 50%?			
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study pre-specified and applied uniformly to all participants?			
5. Was a sample size justification, power description, or variance and effect estimates provided?			
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?			
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10. Was the exposure(s) assessed more than once over time?			
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
12. Were the outcome assessors blinded to the exposure status of participants?			
13. Was loss to follow-up after baseline 20% or less?			
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?			
Quality Rating (Good, Fair, or Poor) (see guidance)			

Rater #1 initials:
Rater #2 initials:
Additional Comments (If POOR, please state why):

*CD, cannot determine; NA, not applicable; NR, not reported

A- 3: Names of Included Social Communication Measures

ABAS-Social Subscale	Bayley Communication	ESCS IJA	MCDI Comprehension	PEP-R Verbal Cognitive	SSRS-Positive Social Behaviors
ABBLS-R	Bayley Social Emotional	ESCS RJA	MCDI Expressive	Reynell Total	SSRS-Negative Social Behaviors
ABC Total	BOSCC Social Affect	ESCS Requesting	MCDI Gestures	Reynell Expressive	Symbolic Play Test
ABC Social Withdrawal	BOSCC Total	Expressive Vocabulary Test	MCDI MLU	Reynell Receptive	TOLD3 Language Quotient
ABC Inappropriate Language	CARS	FEAS	PLS Receptive	ROWPVT	TOPP-Pretend Play
AEPS-Social Communication	CCC-Communication	FEAS Developmental Questionnaire	PLS Total	Scales of Independent Behavior- Social Interaction	Test of Playfulness
AEPS-Social	CCC Social Interaction	GARS	PIA Total	Scales of Independent Behavior- Expressive	VB-MAPP
ADI-Communication	CELF-4	GMDS Language	PIA Understanding	Scales of Independent Behavior- Receptive	Vineland Communication
ADI-Social Interaction	Communicative Developmental Inventory	GMDS Social	PIA Social Reciprocity	Schlichting Test of Language Production	Vineland Expressive
ADOS Severity	CBRS Initiations	GFTA	PIA Non-Verbal Communication	Social Communication Checklist Social Engagement	Vineland Receptive
ADOS Social Affect	CBRS Affect	HKBABS Communication	PDD-Behavioral Inventory	Social Communication Checklist- Language	Vineland Socialization
ADOS Language	CBRS Joint Attention	HKBABS Social Communication	Preschool Kindergarten Behavior Scales Social Communication	Social Communication Checklist-Play	
ADOS Play	CBRS Interest	Joy and Fun Questionnaire	PPVT-Receptive	SCQ	
ADOS Social Interaction	CBRS Attention	PEP-R Overall Communication	PPVT-Word Count	SPA	
ADOS Total	CSBS Social Communication	PJAM Finding Faces	Merrill Palmer- R	SPACE	
ASQ Social-Emotional	CSBS Speech	PJAM IJA	Mullen Expressive Language	SRS Total	
ASQ Communication	CSBS-Symbolic	PJAM RJA	Mullen Receptive Language	SRS Cognition	
Autism Symptom Rating Scale	CSBS	PJAM Turn Taking	Mullen Verbal Combined	SRS Social Awareness	
ATEC-Language	Developmental Profile 3 Communication	PLS Expressive	PEP-R Expressive	SRS Social Communication	
ATEC-Social	Developmental Profile 3 Social	KTEA Oral Language	PEP-R Receptive	SRS Social Motivation	
BASC-Social Skills	EOWPVT	KTEA Comprehension		SSIS	
BASC-Functional Communication	ESCS Gestures				

Note. ABAS= Adaptive Behavior Assessment System, ABBLS= The Assessment of Basic Language and learning Skills, ABC= Aberrant Behavior Checklist, AEPS= Assessment, Evaluation and Programming System, ADI= Autism Diagnostic Interview, ADOS= Autism Diagnostic Observation Schedule, ASQ= Ages and Stages Questionnaire, ATEC= Autism Treatment Evaluation Checklist, BASC= Behavior Assessment System for Children, BOSCC= Brief Observation of Social Communication Change, CARS= Childhood Autism Rating Scales, CCC= Children's Communication Checklist, CBRS= Conner's Comprehensive Behavior Rating Scales, CSBS= Communication and Symbolic Behavior Scale, EOWPVT= Expressive One Word Picture Vocabulary Test, ESCS= Early Social Communication Scales, FEAS= Functional Emotional Assessment Scale, GARS= Gilliam Autism Rating Scale, GMDS= Griffiths Mental Development Scale, GFTA= Goldman-Fristoe Test of Articulation, HKBABS= Hong Kong Based Adaptive Behavior Scale, PEP-R= Psychoeducational Profile-Revised, PJAM= Precursors of Joint Attention Measure, PLS= Preschool Language Scales, KTEA= Kaufman Test of Educational Achievement, MCDI= Macarthur Communicative Development Inventories, PIA= Parent Interview for Autism, PDD= Pervasive-Developmental Disorder, PPVT=Peabody Picture Vocabulary Test, ROWPVT= Receptive One Word Picture Vocabulary Test, SCQ= Social Communication Questionnaire, SPA= Structured Play Assessment, SRS= Social Responsiveness Scale, SSIS= Social Skills Improvement System, SSRS= Social Skills Rating System, TOLD= Test of Language Development, TOPP= Test of Pretend Play, VB-MAPP= Verbal Behavior Milestones and Placement Program.

Appendix B- Less Commonly Used Outcomes

Figure B- 1 *Bubble Plots for CSBS- Social Communication Effect Sizes*

Length	Age	Year

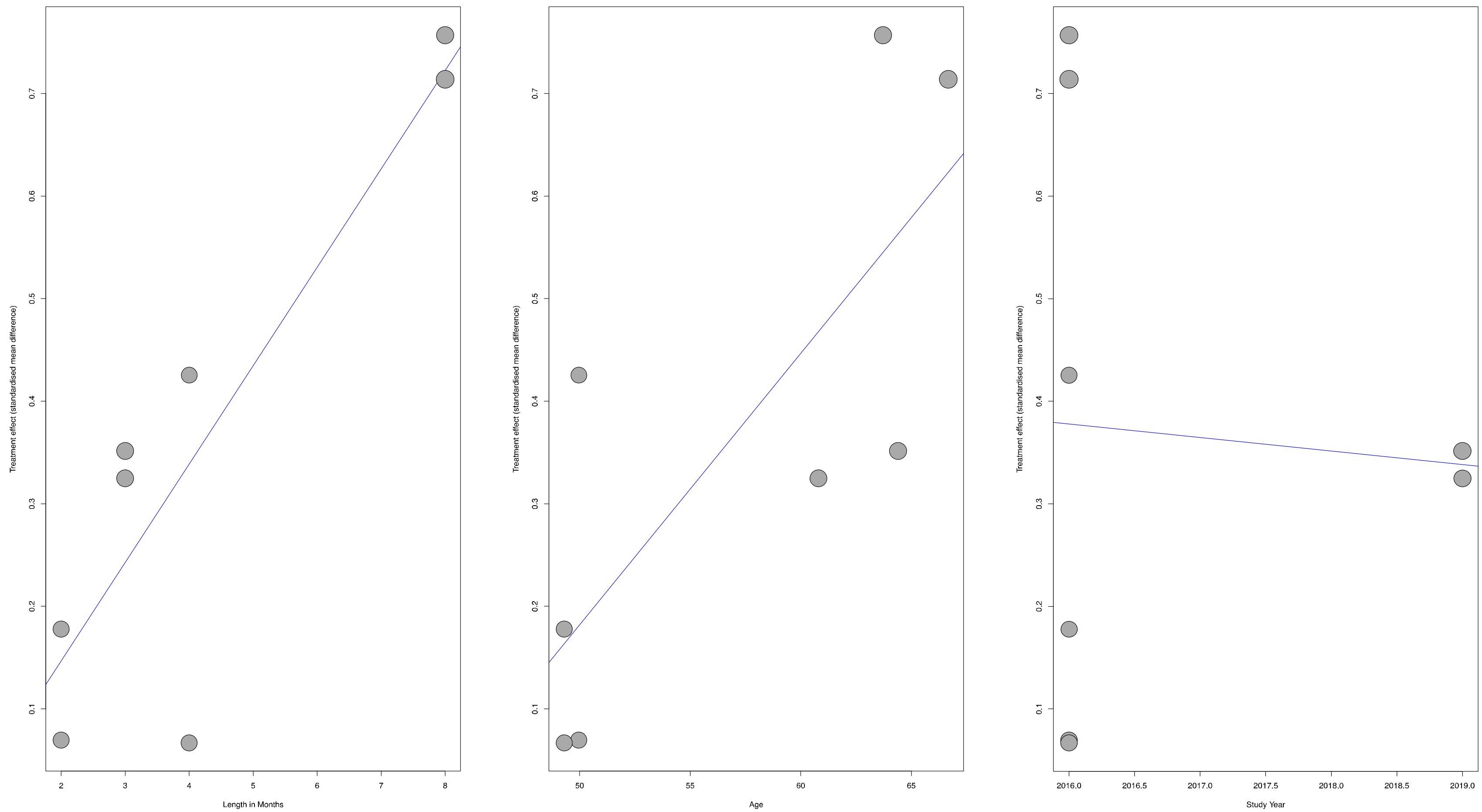


Figure B- 2: *Bubble Plots for EOWPVT Effect Sizes*

Length	Age	Year

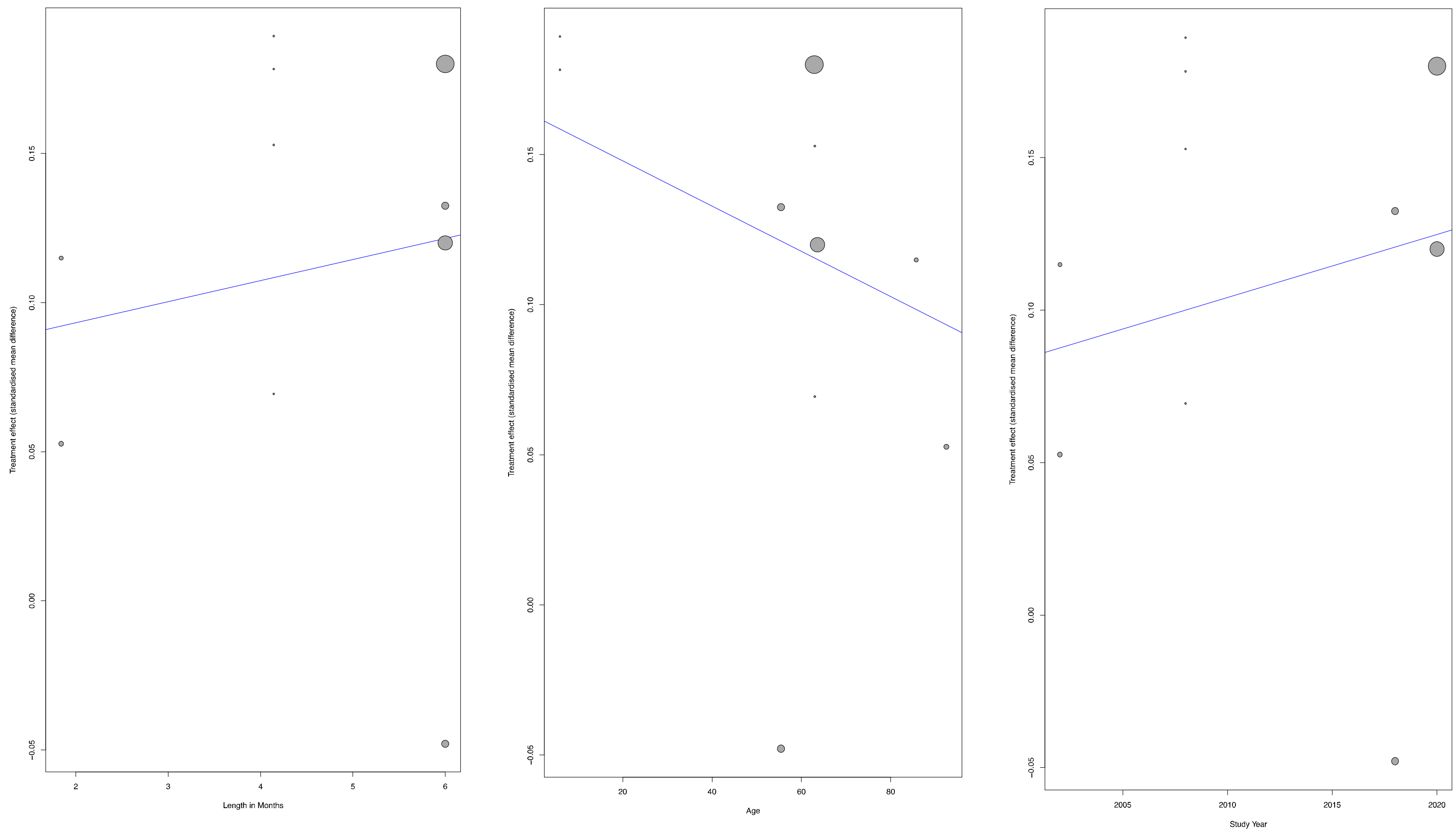


Figure B- 3: *Bubble Plots for FEAS Effect Sizes*

Length	Age	Year

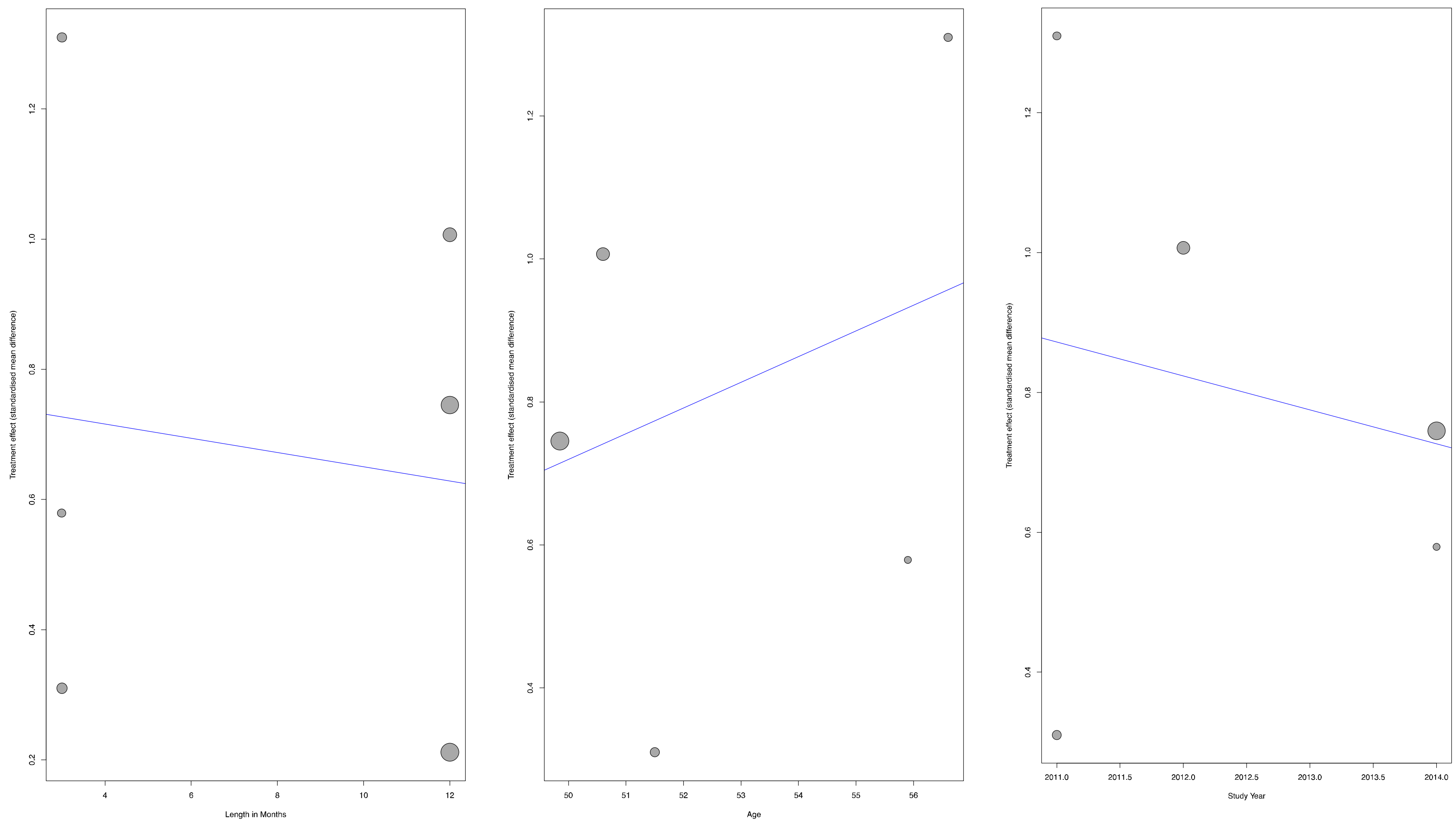


Figure B- 4: *Bubble Plots for GMDS Language Effect Sizes*

Length	Age	Year

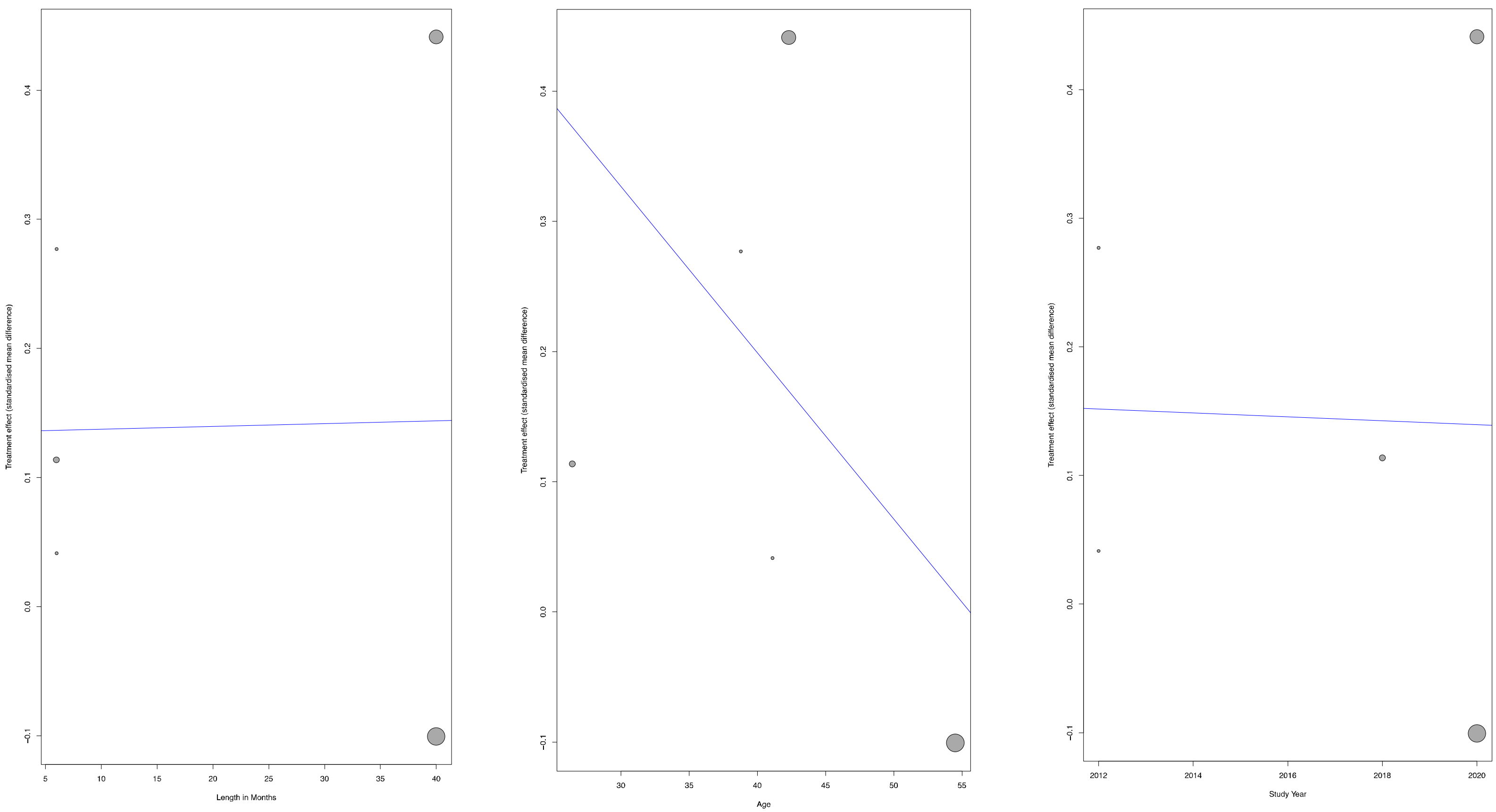


Figure B- 5: *Bubble Plots for GMDS Personal Social Effect Sizes*

Length	Age	Year

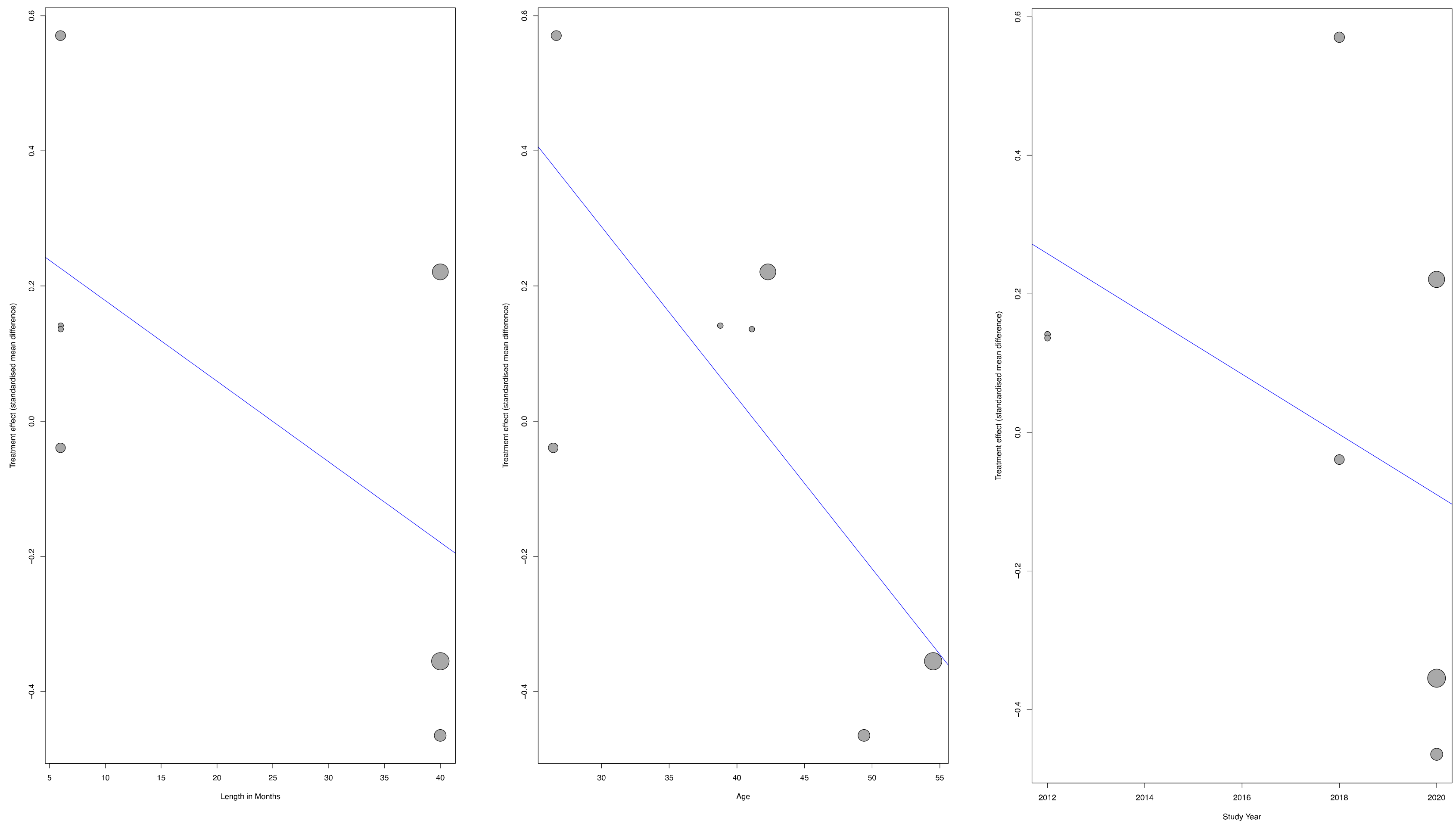


Figure B- 6: *Bubble Plots for MCDI Gestures Produced Effect Sizes*

Length	Age	Year

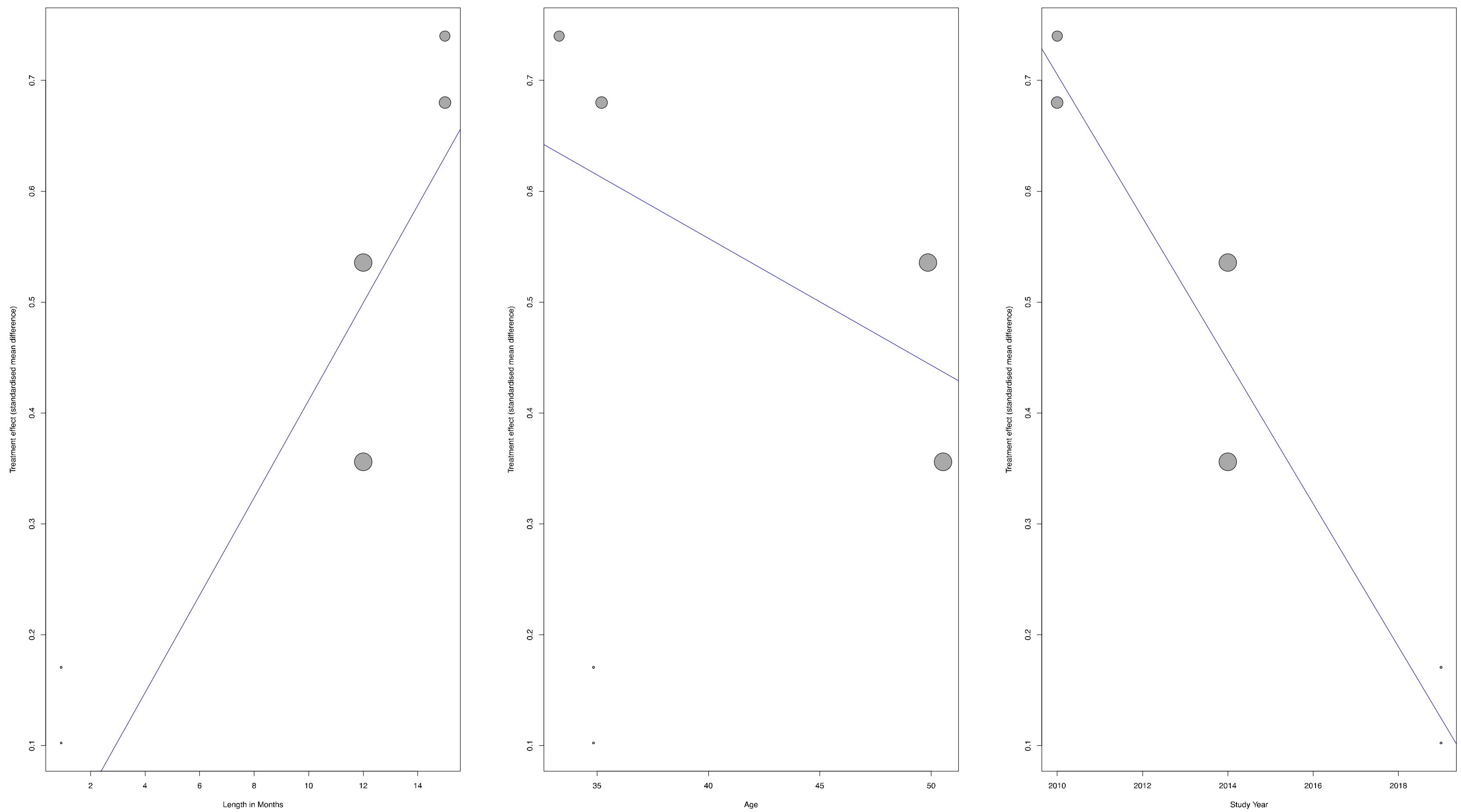


Figure B- 7: *Bubble Plots for Mullen Expressive and Receptive Combined Effect Sizes*

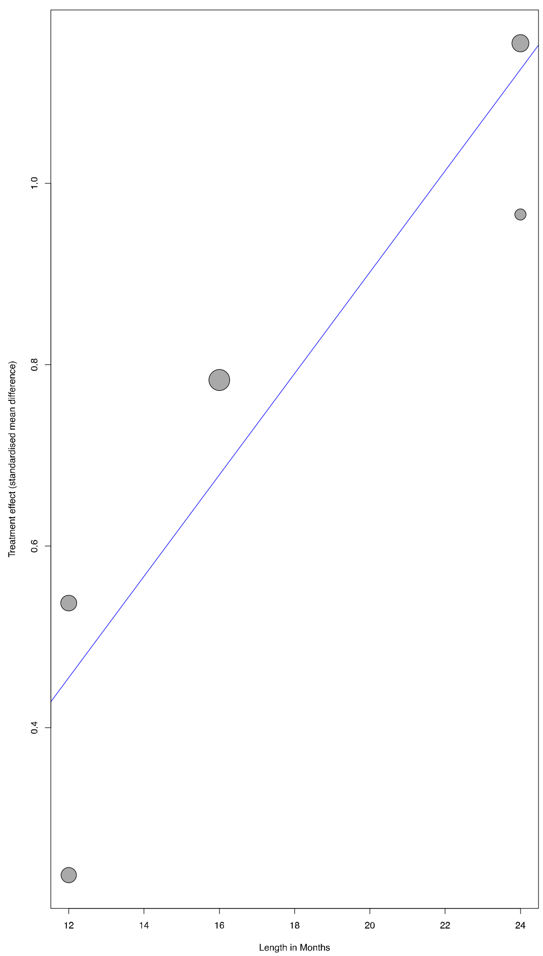
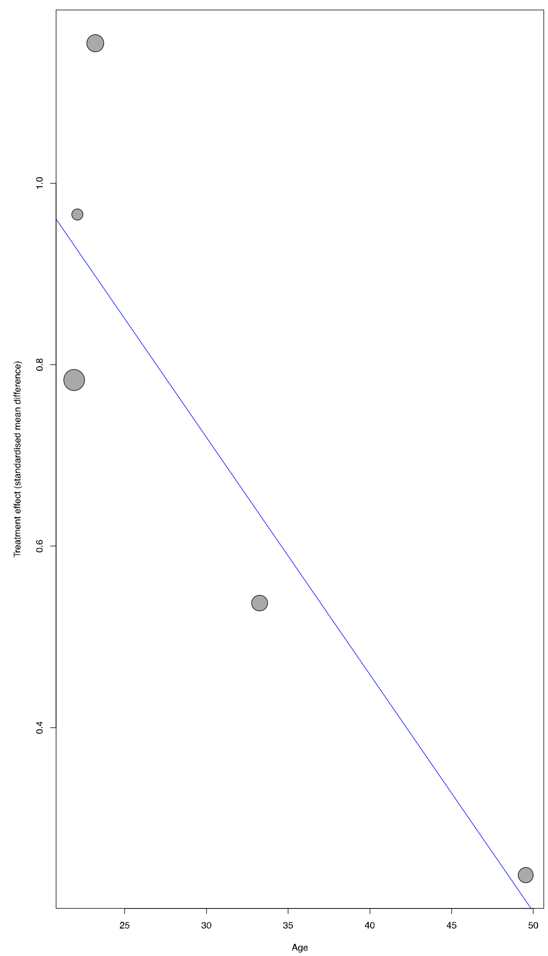
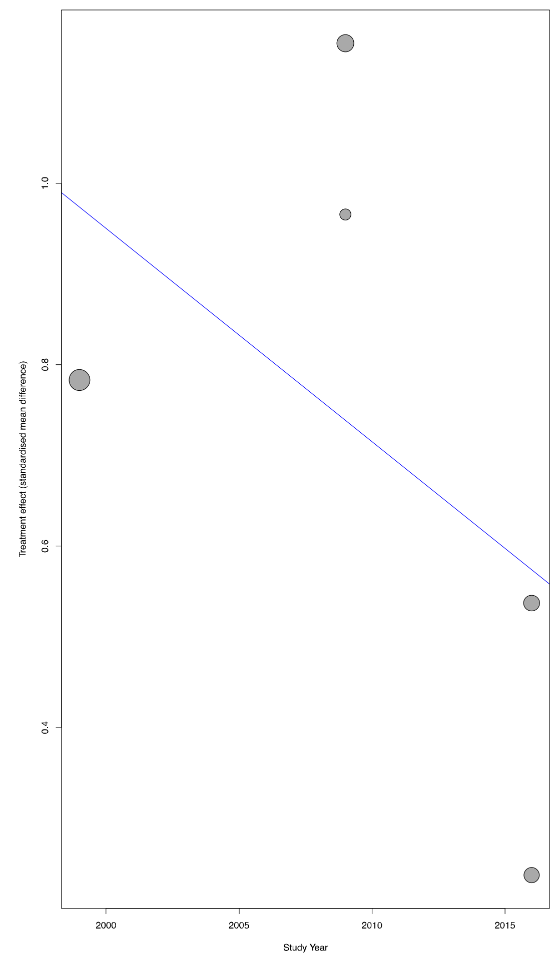
Length	Age	Year
		

Figure B- 8: *Bubble Plots for PEP-R Verbal Cognitive Effect Sizes*

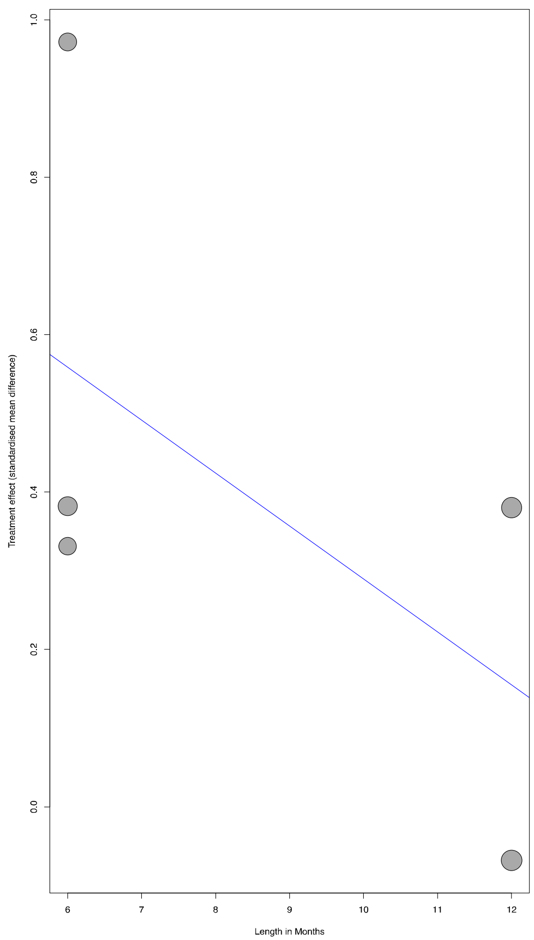
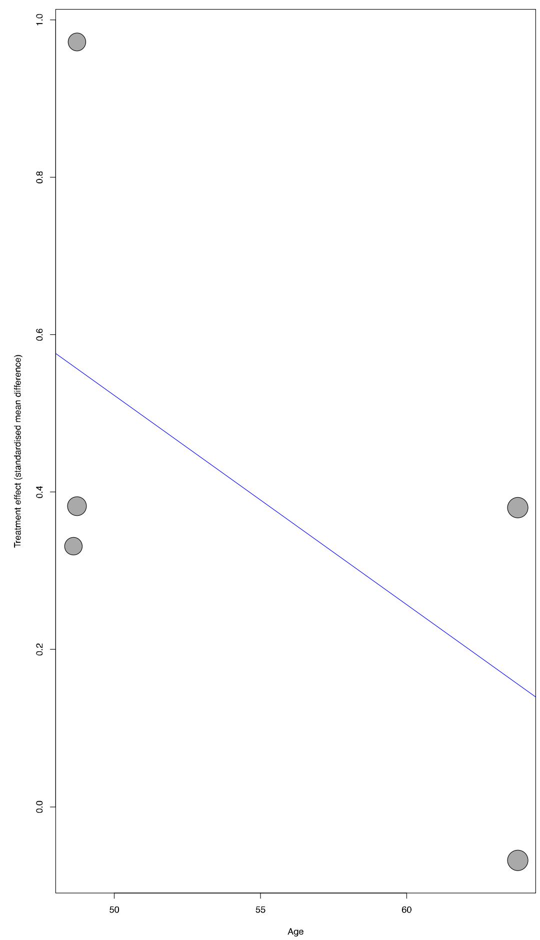
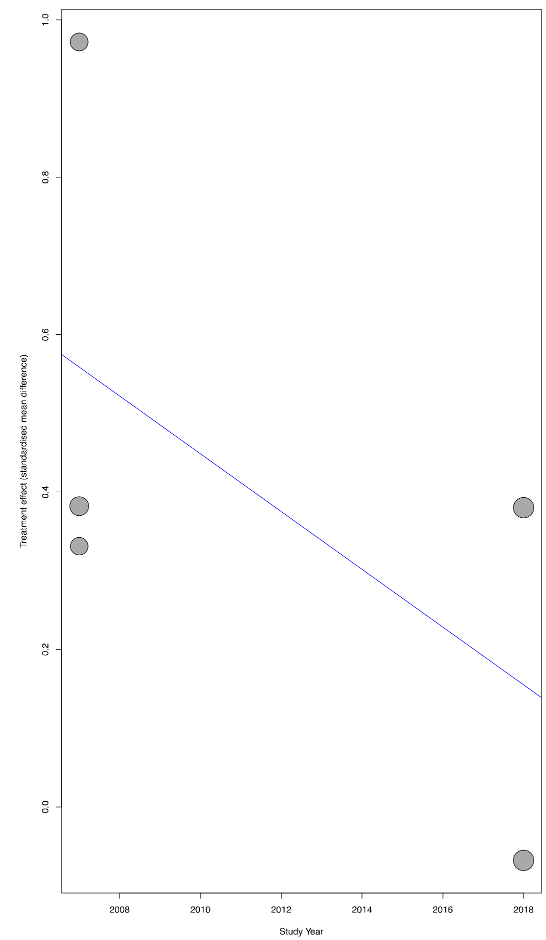
Length	Age	Year
		

Figure B- 9: *Bubble Plots for PIA Nonverbal Communication Effect Sizes*

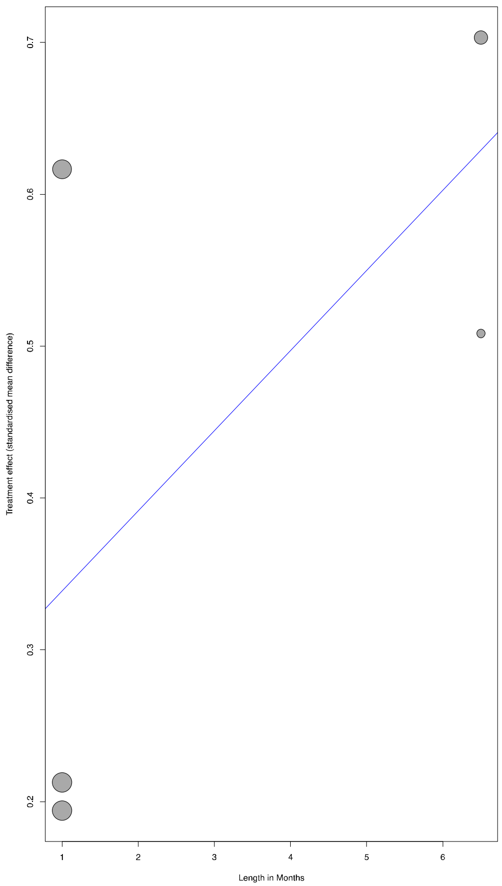
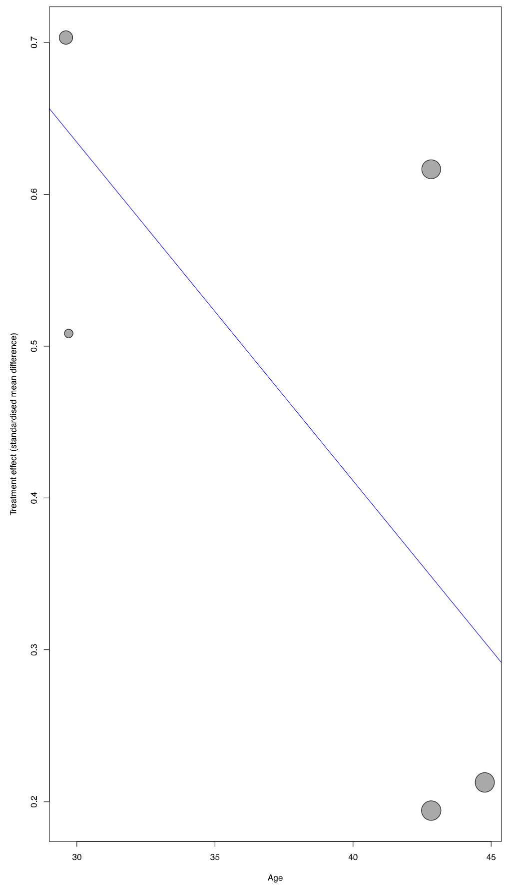
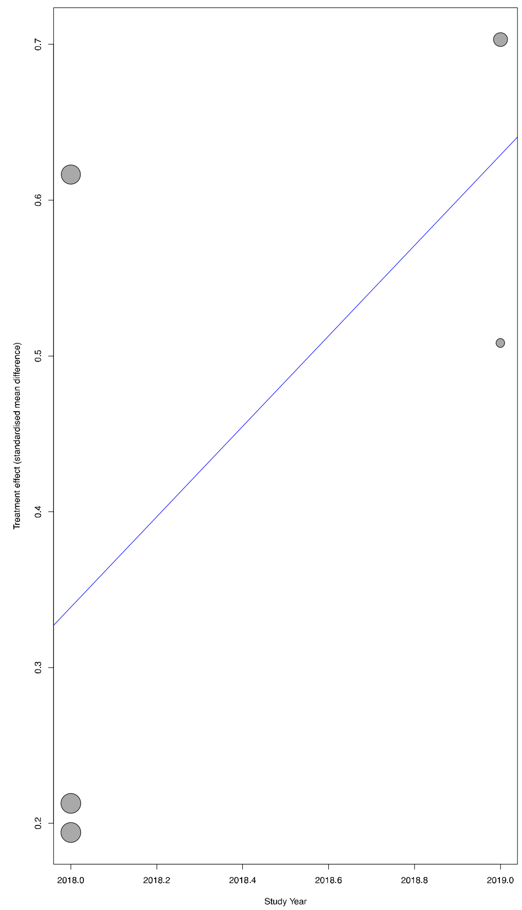
Length	Age	Year
		

Figure B- 10: *Bubble Plots for PIA Social Reciprocity Effect Sizes*

Length	Age	Year

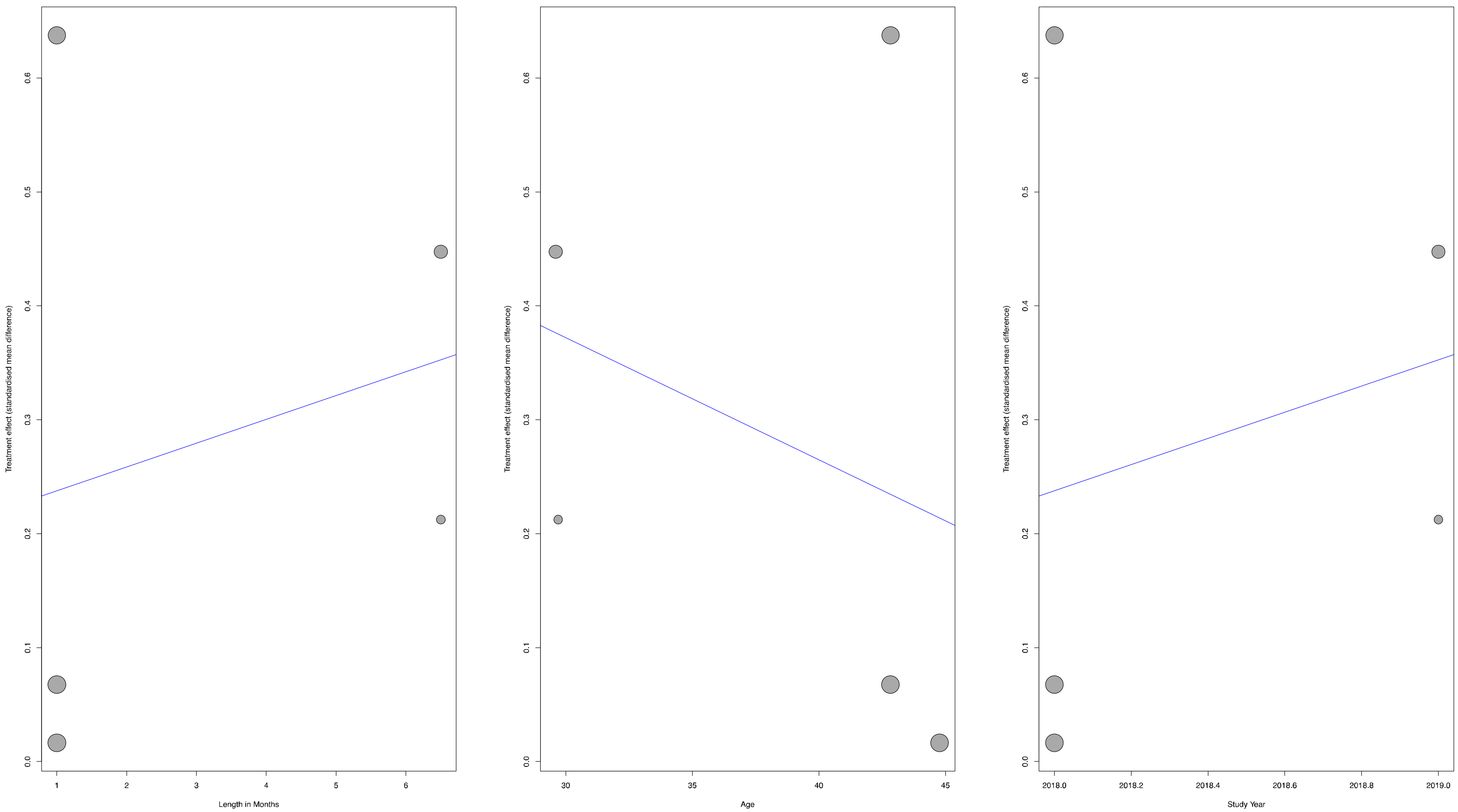


Figure B- 11: *Bubble Plots for PLS Expressive and Receptive Combined Effect Sizes*

Length	Age	Year

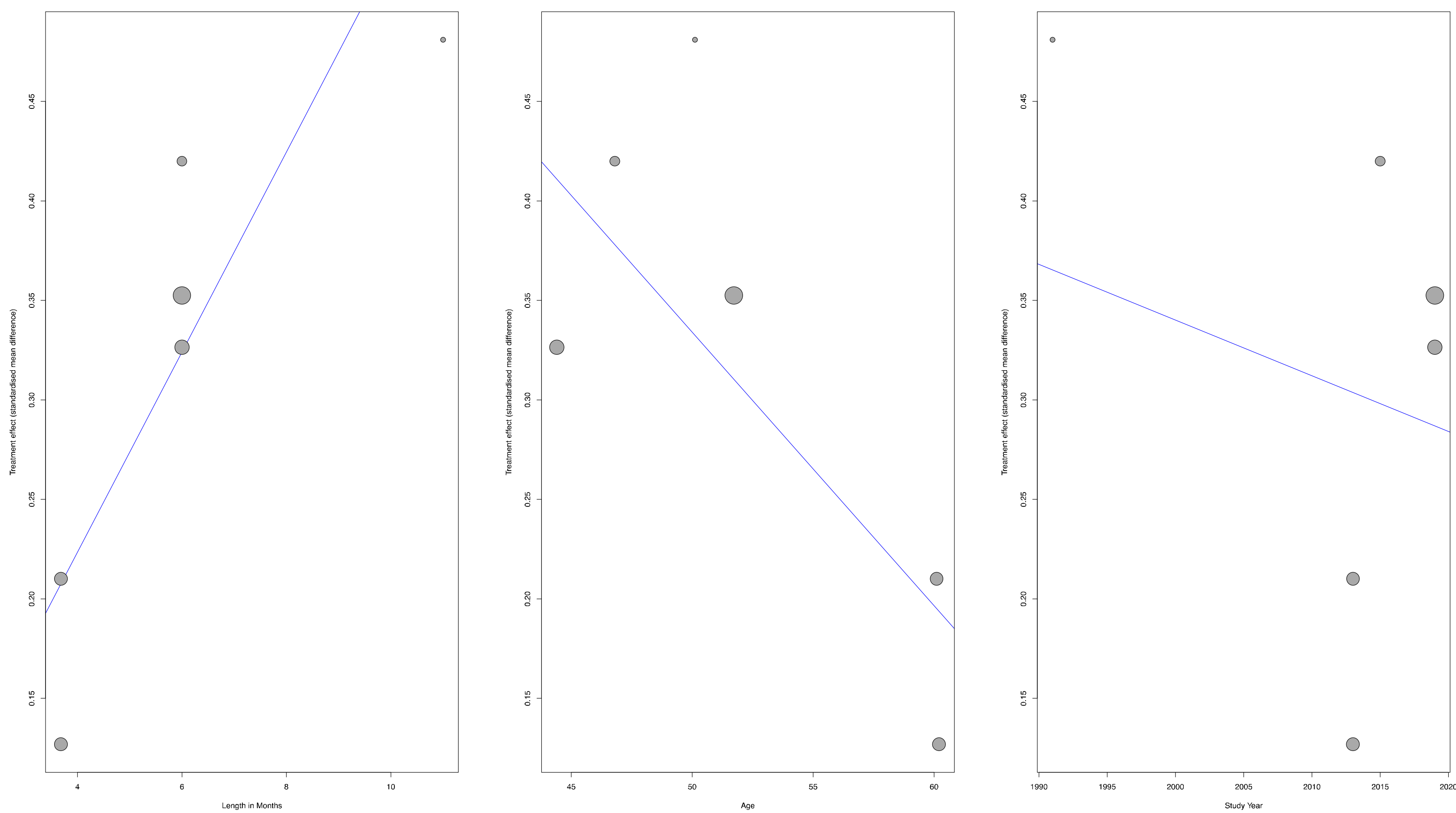


Figure B- 12: *Bubble Plots for Social Communication Questionnaire Effect Sizes*

Length	Age	Year

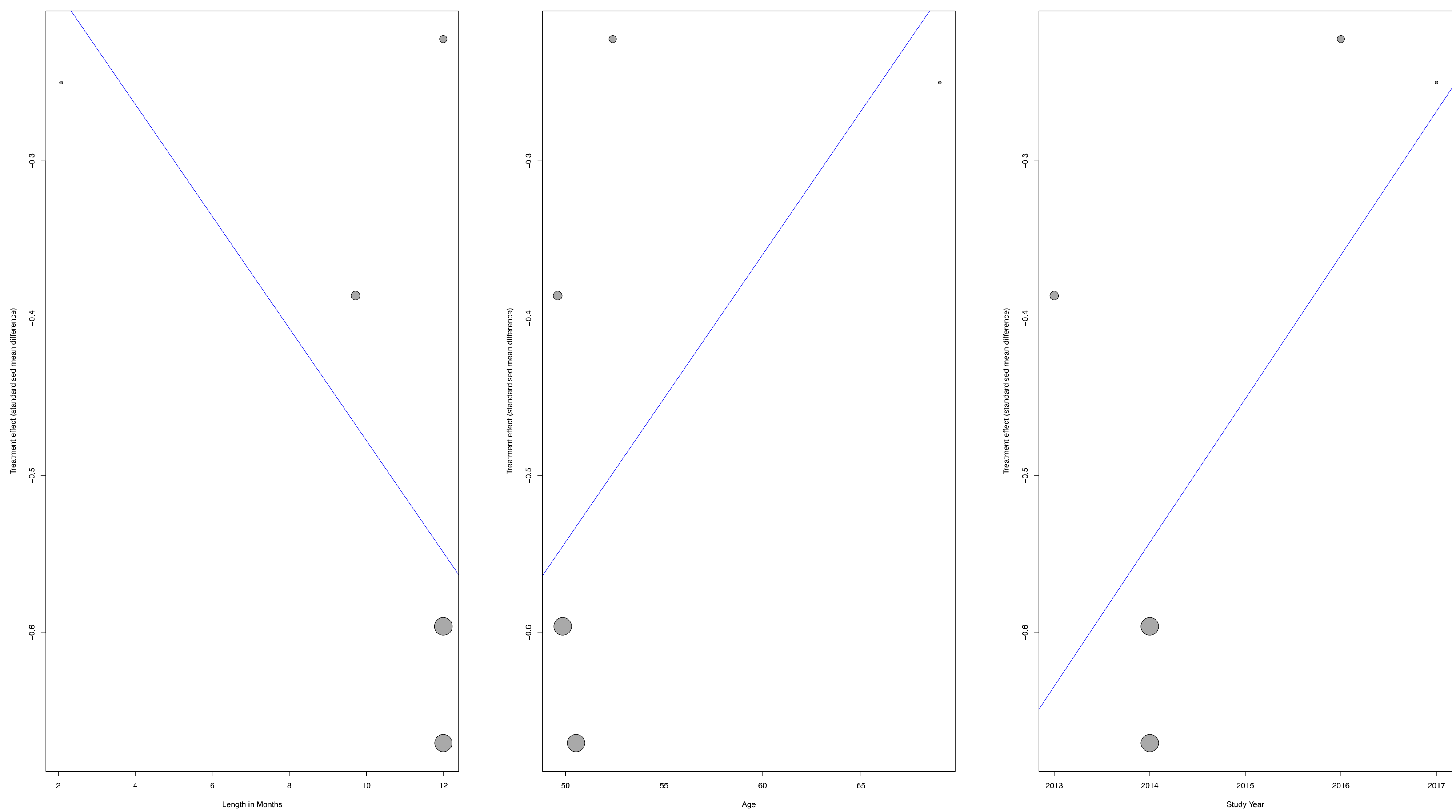


Figure B- 13: *Bubble Plots for SSIS Effect Sizes*

Length	Age	Year

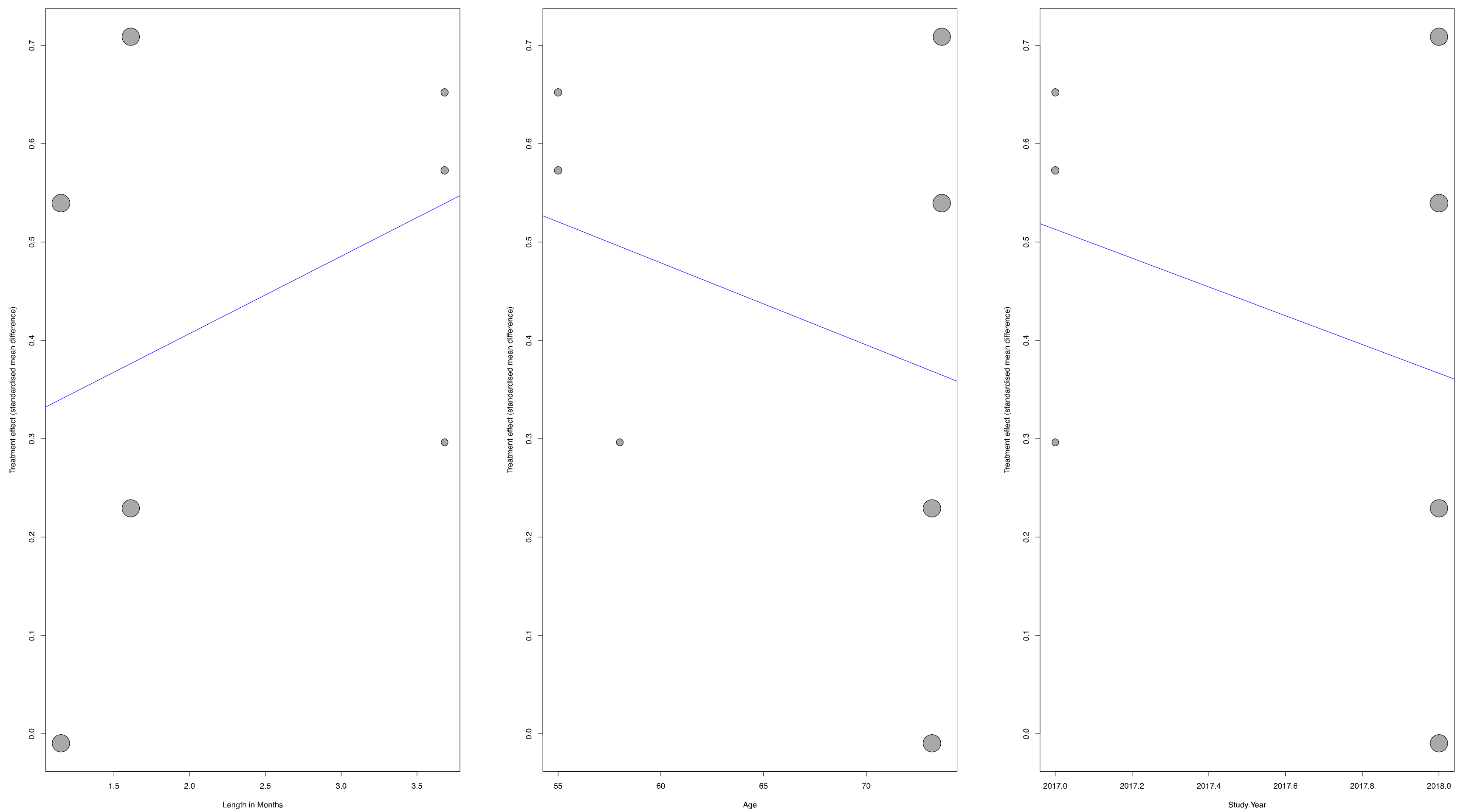
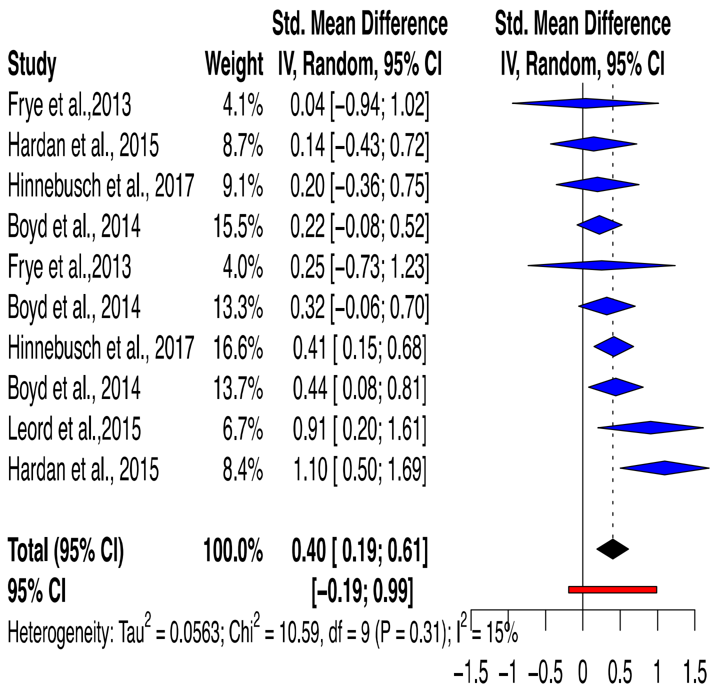


Figure B- 14: *Vineland Expressive Language Subgroup Effect Sizes*

Overall



Interventions Received

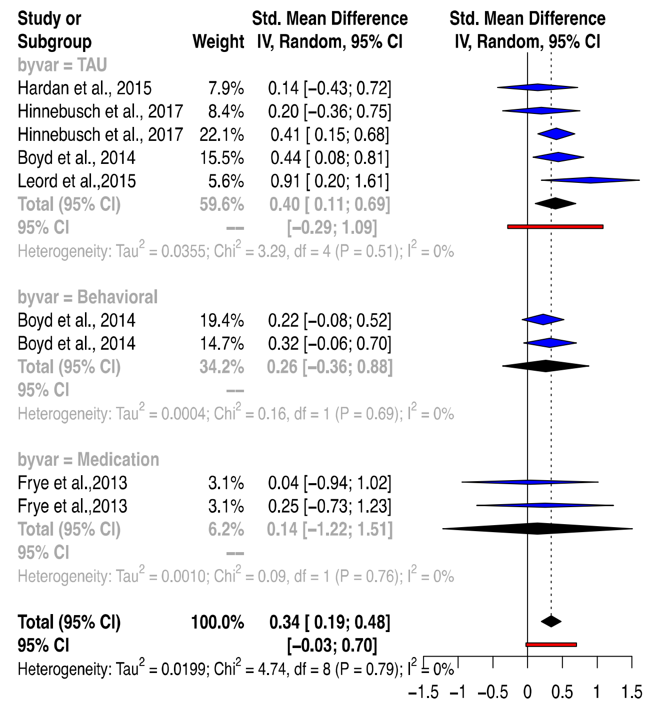


Figure B- 15: *Bubble Plots for Vineland Expressive Language Effect Sizes*

Length	Age	Year

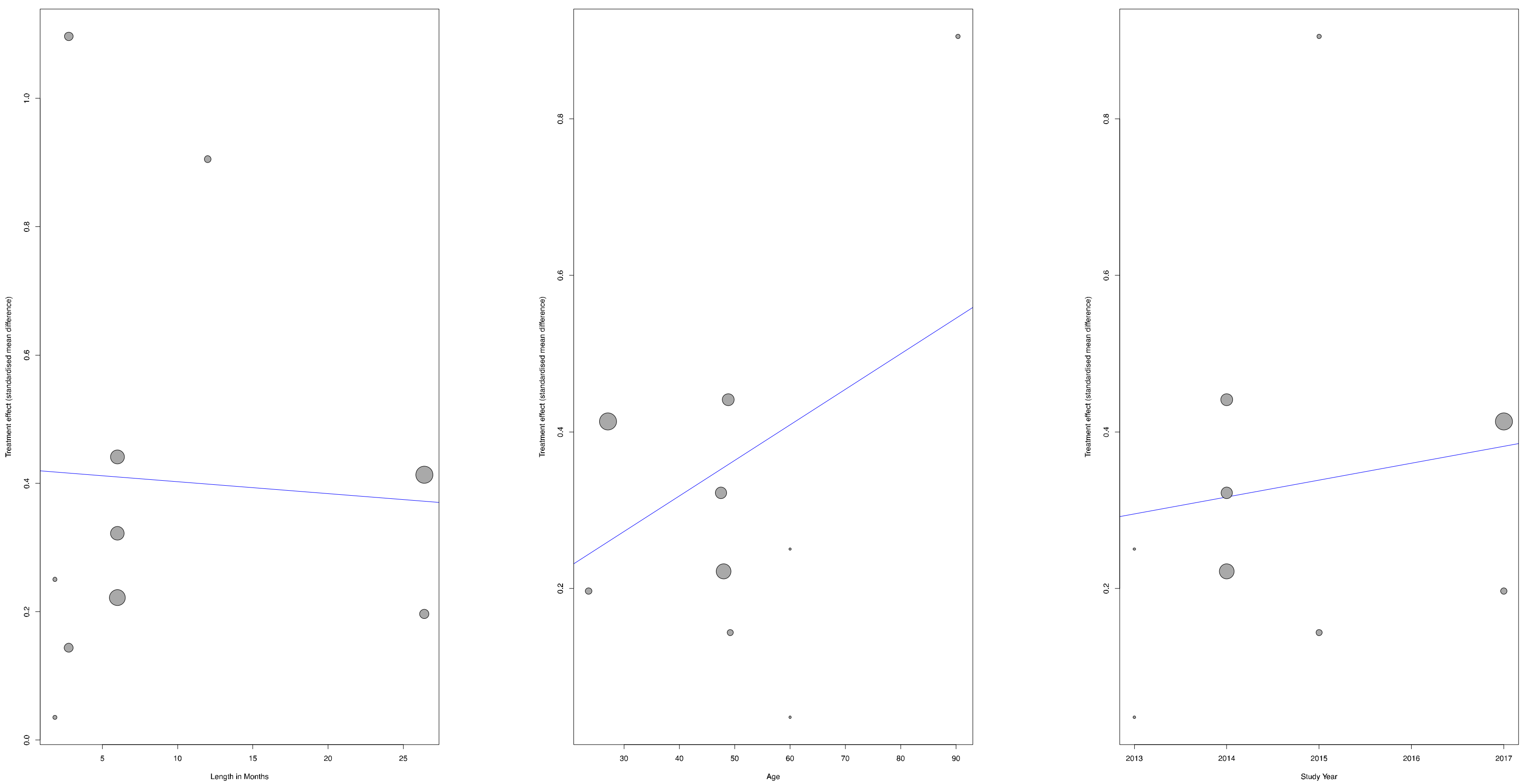
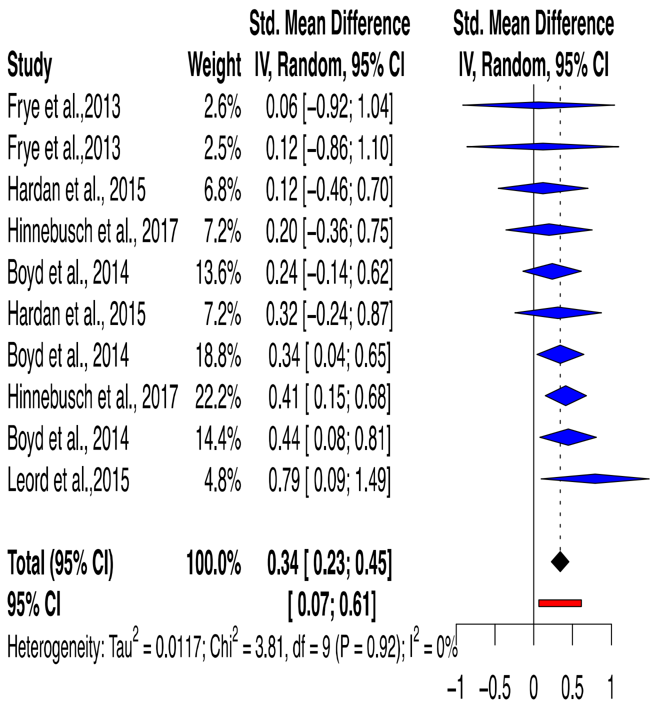


Figure B- 16: *Vineland Receptive Language Subgroup Effect Sizes*

Overall	Interventions Received



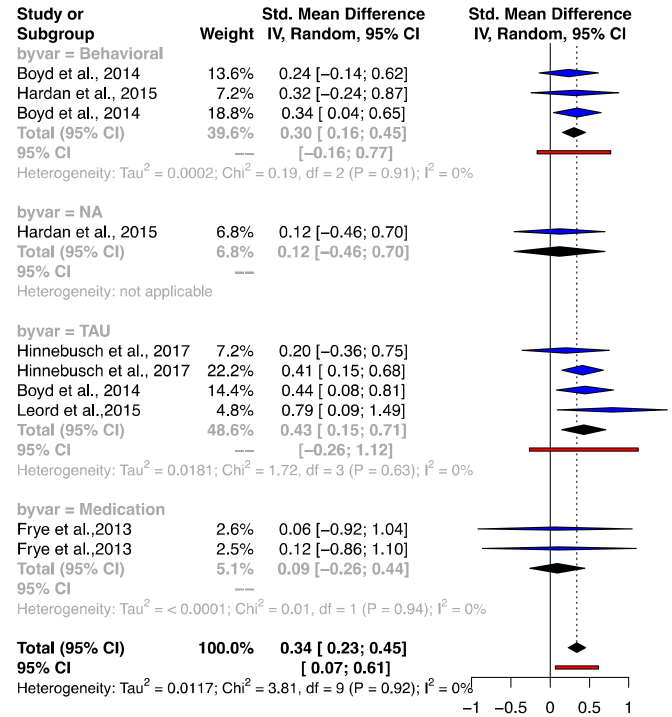
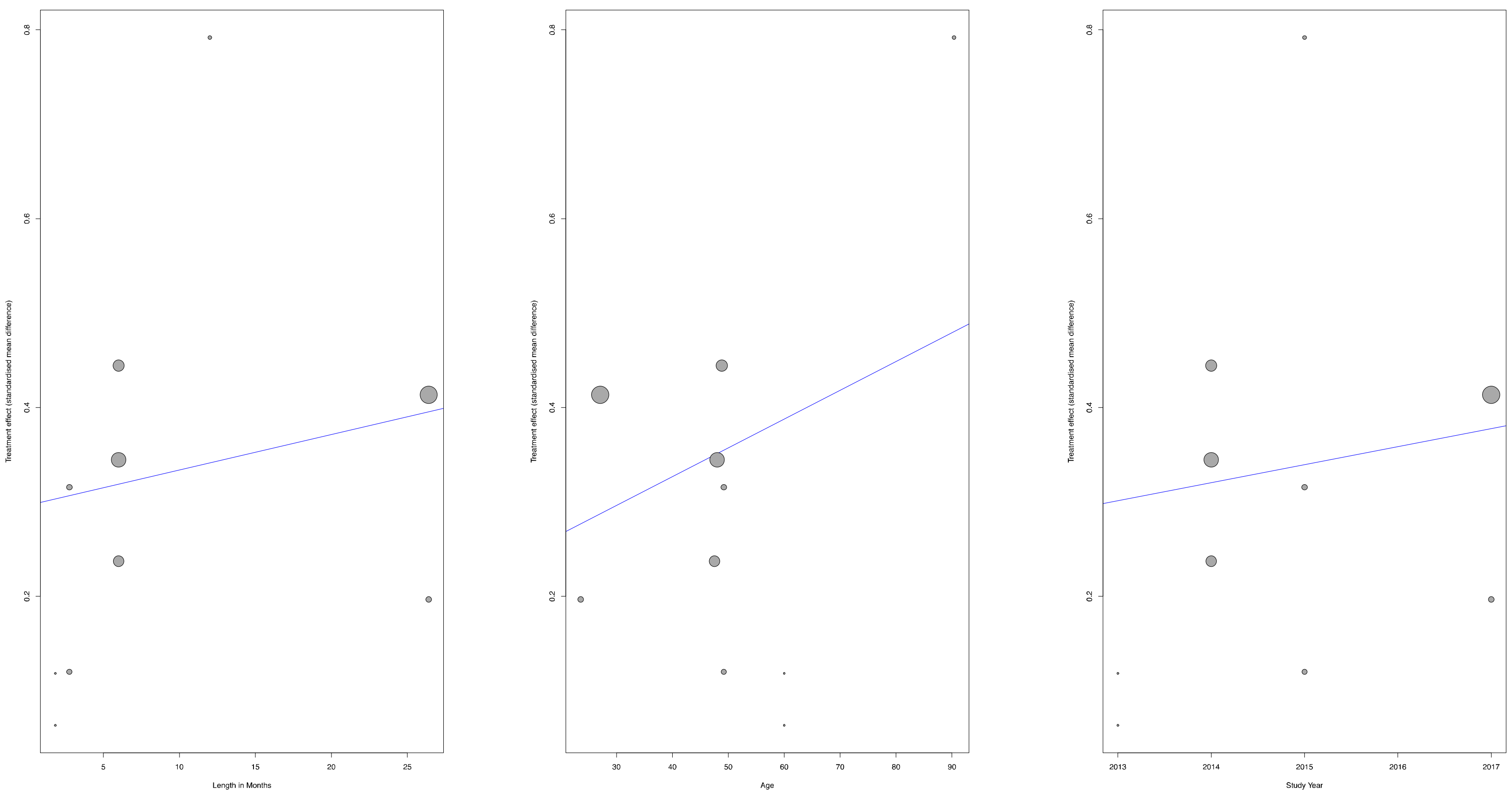


Figure B- 17: *Bubble Plots for Vineland Receptive Language Effect Sizes*

Length	Age	Year



References

*Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. (2009). Joint Engagement and the Emergence of Language in Children with Autism and Down Syndrome. *Journal of Autism and Developmental Disorders*, *39*(1), 84–96. https://doi.org/10.1007/s10803-008-0601-7

*Aldred, C., Green, J., & Adams, C. (2004). A new social communication intervention for children with autism: Pilot randomised controlled treatment study suggesting effectiveness. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *45*(8), 1420–1430. https://doi.org/10.1111/j.1469-7610.2004.00848.x

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: Author.

Anagnostou, E., Zwaigenbaum, L., Szatmari, P., Fombonne, E., Fernandez, B. A., Woodbury-Smith, M., ... & Scherer, S. W. (2014). Autism spectrum disorder: advances in evidence-based practice. *Cmaj*, *186*(7), 509-519. https://doi.org/10.1503/cmaj.121756

Anagnostou, E., Jones, N., Huerta, M., Halladay, A. K., Wang, P., Scahill, L., ... & Dawson, G. (2015). Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism*, *19*(5), 622-636. https://doi.org/10.1177/1362361314542955

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., ... & Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. Journal of consulting and clinical psychology, 75(4), 594. https://doi.org/10.1037/0022-006X.75.4.594

Anthony, C. J., DiPerna, J. C., & Lei, P.-W. (2016). Maximizing measurement efficiency of behavior rating scales using Item Response Theory: An example with the Social Skills Improvement System — Teacher Rating Scale. *Journal of School Psychology*, *55*, 57–69. https://doi.org/10.1016/j.jsp.2015.12.005

*Artoni, S., Bastiani, L., Buzzi, M. C., Buzzi, M., Curzio, O., Pelagatti, S., & Senette, C. (2018). Technology-enhanced ABA intervention in children with autism: A pilot study. *Universal Access in the Information Society*, *17*(1), 191–210. https://doi.org/10.1007/s10209-017-0536-x

*Azad, G., Reisinger, E., Xie, M., & Mandell, D. S. (2016). Parent and Teacher Concordance on the Social Responsiveness Scale for Children with Autism. *School Mental Health*, *8*(3), 368–376. https://doi.org/10.1007/s12310-015-9168-6

*Baghdadli, A., Michelon, C., Pernon, E., Picot, M.-C., Miot, S., Sonié, S., Rattaz, C., & Mottron, L. (2018). Adaptive trajectories and early risk factors in the autism spectrum: A 15-year prospective study. *Autism Research: Official Journal of the International Society for Autism Research*, *11*(11), 1455–1467. https://doi.org/10.1002/aur.2022

*Baker, J. K., Messinger, D. S., Lyons, K. K., & Grantz, C. J. (2010). A pilot study of maternal sensitivity in the context of emergent autism. *Journal of Autism and Developmental Disorders*, *40*(8), 988–999. https://doi.org/10.1007/s10803-010-0948-4

*Bang, J., & Nadig, A. (2015). Learning language in autism: Maternal linguistic input contributes to later vocabulary. *Autism Research: Official Journal of the International Society for Autism Research*, *8*(2), 214–223. https://doi.org/10.1002/aur.1440

*Barbaro, J., & Dissanayake, C. (2017). Diagnostic stability of autism spectrum disorder in toddlers prospectively identified in a community-based setting: Behavioural characteristics and predictors of change over time. *Autism: The International Journal of Research and Practice*, *21*(7), 830–840. https://doi.org/10.1177/1362361316654084

Barokova, M. D., La Valle, C., Hassan, S., Lee, C., Xu, M., McKechnie, R., ... & Tager-Flusberg, H. (2021). Eliciting language samples for analysis (ELSA): a new protocol for assessing expressive language and communication in autism. *Autism Research*, *14*(1), 112-126. https://doi.org/10.1002/aur.2380

*Bass, M. M., Duchowny, C. A., & Llabre, M. M. (2009). The effect of therapeutic horseback riding on social functioning in children with autism. *Journal of Autism and Developmental Disorders*, *39*(9), 1261–1267. https://doi.org/10.1007/s10803-009-0734-3

*Beaudoin, A. J., Sébire, G., & Couture, M. (2019). Parent-mediated intervention tends to improve parent-child engagement, and behavioral outcomes of toddlers with ASD-positive screening: A randomized crossover trial. *Research in Autism Spectrum Disorders*, *66*, 101416. https://doi.org/10.1016/j.rasd.2019.101416

*Belsito, K. M., Law, P. A., Kirk, K. S., Landa, R. J., & Zimmerman, A. W. (2001a). Lamotrigine therapy for autistic disorder: A randomized, double-blind, placebo-controlled trial. *Journal of Autism and Developmental Disorders*, *31*(2), 175–181. https://doi.org/10.1023/a:1010799115457

*Belsito, K. M., Law, P. A., Kirk, K. S., Landa, R. J., & Zimmerman, A. W. (2001b). Lamotrigine therapy for autistic disorder: A randomized, double-blind, placebo-controlled trial. *Journal of Autism and Developmental Disorders*, *31*(2), 175–181. https://doi.org/10.1023/a:1010799115457

*Ben Itzchak, E., Lahat, E., Burgin, R., & Zachor, A. D. (2008). Cognitive, behavior and intervention outcome in young children with autism. *Research in Developmental Disabilities*, *29*(5), 447–458. https://doi.org/10.1016/j.ridd.2007.08.003

*Ben Itzchak, E., & Zachor, D. A. (2009). Change in autism classification with early intervention: Predictors and outcomes. *Research in Autism Spectrum Disorders*, *3*(4), 967–976. https://doi.org/10.1016/j.rasd.2009.05.001

Ben-Itzchak, E., Watson, L. R., & Zachor, D. A. (2014). Cognitive ability is associated with different outcome trajectories in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *44*(9), 2221-2229. https://doi.org/10.1007/s10803-014-2091-0

*Bent, S., Dang, K., Widjaja, F., Lawton, B., Nazneen, N., & Hendren, R. L. (2017). Examining Clinics for Children with Autism: The Autism Translating To Treatment Study. *Journal of Alternative and Complementary Medicine (New York, N.Y.)*, *23*(5), 340–347. https://doi.org/10.1089/acm.2016.0301

*Berman, S., Ventola, P., & Gordon, I. (2018). Improvements in micro level indices of social communication following Pivotal Response Treatment (PRT). *Research in Autism Spectrum Disorders*, *51*, 56–65. https://doi.org/10.1016/j.rasd.2018.04.003

*Bieleninik, L., Geretsegger, M., Mössler, K., Assmus, J., Thompson, G., Gattino, G., Elefant, C., Gottfried, T., Igliozzi, R., Muratori, F., Suvini, F., Kim, J., Crawford, M. J., Odell-Miller, H., Oldfield, A., Casey, Ó., Finnemann, J., Carpente, J., Park, A.-L., … for the TIME-A Study Team. (2017). Effects of Improvisational Music Therapy vs Enhanced Standard Care on Symptom Severity Among Children With Autism Spectrum Disorder: The TIME-A Randomized Clinical Trial. *JAMA*, *318*(6), 525. https://doi.org/10.1001/jama.2017.9478

*Boccanfuso, L., Scarborough, S., Abramson, R. K., Hall, A. V., Wright, H. H., & O'Kane, J. M. (2017). A low-cost socially assistive robot and robot-assisted intervention for children with autism spectrum disorder: Field trials and lessons learned. *Autonomous Robots*, *41*(3), 637–655. https://doi.org/10.1007/s10514-016-9554-4

Bolte, E. E., & Diehl, J. J. (2013). Measurement tools and target symptoms/skills used to assess treatment response for individuals with autism spectrum disorder. *Journal of autism and developmental disorders*, *43*(11), 2491-2501. https://doi.org/10.1007/s10803-013-1798-7

Borenstein, M., & Higgins, J. P. (2013). Meta-analysis and subgroups. *Prevention science*, *14*(2), 134-143. https://doi.org/10.1007/s11121-013-0377-7

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

*Boyd, B. A., Hume, K., McBee, M. T., Alessandri, M., Gutierrez, A., Johnson, L., Sperry, L., & Odom, S. L. (2014). Comparative Efficacy of LEAP, TEACCH and Non-Model-Specific Special Education Programs for Preschoolers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *44*(2), 366–380. https://doi.org/10.1007/s10803-013-1877-9

*Boyd, B. A., Watson, L. R., Reszka, S. S., Sideris, J., Alessandri, M., Baranek, G. T., Crais, E. R., Donaldson, A., Gutierrez, A., Johnson, L., & Belardi, K. (2018). Efficacy of the ASAP Intervention for Preschoolers with ASD: A Cluster Randomized Controlled Trial. *Journal of Autism and Developmental Disorders*, *48*(9), 3144–3162. https://doi.org/10.1007/s10803-018-3584-z

*Brian, J. A., Smith, I. M., Zwaigenbaum, L., & Bryson, S. E. (2017). Cross-site randomized control trial of the Social ABCs caregiver-mediated intervention for toddlers with autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *10*(10), 1700–1711. https://doi.org/10.1002/aur.1818

*Bussu, G., Jones, E. J. H., Charman, T., Johnson, M. H., Buitelaar, J. K., & BASIS Team. (2018). Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis. *Journal of Autism and Developmental Disorders*, *48*(7), 2418–2433. https://doi.org/10.1007/s10803-018-3509-x

*Campbell, S. B., Mahoney, A. S., Northrup, J., Moore, E. L., Leezenbaum, N. B., & Brownell, C. A. (2018). Developmental Changes in Pretend Play from 22- to 34-Months in Younger Siblings of Children with Autism Spectrum Disorder. *Journal of Abnormal Child Psychology*, *46*(3), 639–654. https://doi.org/10.1007/s10802-017-0324-3

*Caplan, B. (2019). *Responsive Parenting, Gene-Environment Interactions, and Heterogeneous Social Development in Autism Spectrum Disorder* [UCLA]. https://escholarship.org/uc/item/5bq906nv

*Caputo, G., Ippolito, G., Mazzotta, M., Sentenza, L., Muzio, M. R., Salzano, S., & Conson, M. (2018). Effectiveness of a Multisystem Aquatic Therapy for Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *48*(6), 1945–1956. https://doi.org/10.1007/s10803-017-3456-y

*Cariveau, T., Shillingsburg, M. A., Alamoudi, A., Thompson, T., Bartlett, B., Gillespie, S., & Scahill, L. (2019). Brief Report: Feasibility and Preliminary Efficacy of a Behavioral Intervention for Minimally Verbal Girls with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *49*(5), 2203–2209. https://doi.org/10.1007/s10803-018-03872-3

Carter, A. S., Volkmar, F. R., Sparrow, S. S., Wang, J.-J., Lord, C., Dawson, G., Fombonne, E., Loveland, K., Mesibov, G., & Schopler, E. (1998). The Vineland Adaptive Behavior Scales: Supplementary Norms for Individuals with Autism. *Journal of Autism and Developmental Disorders*, *28*(4), 287–302. https://doi.org/10.1023/A:1026056518470

*Chang, Y.-C., Shire, S. Y., Shih, W., Gelfand, C., & Kasari, C. (2016). Preschool Deployment of Evidence-Based Social Communication Intervention: JASPER in the Classroom. *Journal of Autism and Developmental Disorders*, *46*(6), 2211–2223. https://doi.org/10.1007/s10803-016-2752-2

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., Drew, A. (1997). Infants with autism: An investigation of empathy, pretend play, joint attention and imitation. Developmental Psychology, 33, 781–789. https://doi.org/10.1037/0012-1649.33.5.781

Charman, T., & Howlin, P. (2003). Research into early intervention for children with autism and related disorders: Methodological and design issues. *Autism*. https://doi.org/10.1177/1362361303007002008

Charman, T. (2003). Why is joint attention a pivotal skill in autism?. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358(1430), 315-324. https://doi.org/10.1098/rstb.2002.1199

*Chawarska, K., Klin, A., Paul, R., Macari, S., & Volkmar, F. (2009). A prospective study of toddlers with ASD: Short-term diagnostic and cognitive outcomes. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *50*(10), 1235–1245. https://doi.org/10.1111/j.1469-7610.2009.02101.x

*Chenausky, K., Norton, A., Tager-Flusberg, H., & Schlaug, G. (2018). Behavioral predictors of improved speech output in minimally verbal children with autism. *Autism Research:*

Official Journal of the International Society for Autism Research, *11*(10), 1356–1365. https://doi.org/10.1002/aur.2006

*Chez, M. G., Buchanan, C. P., Aimonovitch, M. C., Becker, M., Schaefer, K., Black, C., & Komen, J. (2002). Double-blind, placebo-controlled study of L-carnosine supplementation in children with autistic spectrum disorders. *Journal of Child Neurology*, *17*(11), 833–837. https://doi.org/10.1177/08830738020170111501

*Chiang, C.-H., Chu, C.-L., & Lee, T.-C. (2016). Efficacy of caregiver-mediated joint engagement intervention for young children with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, *20*(2), 172–182. https://doi.org/10.1177/1362361315575725

*Chugani, D. C., Chugani, H. T., Wiznitzer, M., Parikh, S., Evans, P. A., Hansen, R. L., Nass, R., Janisse, J. J., Dixon-Thomas, P., Behen, M., Rothermel, R., Parker, J. S., Kumar, A., Muzik, O., Edwards, D. J., Hirtz, D., & Autism Center of Excellence Network. (2016). Efficacy of Low-Dose Buspirone for Restricted and Repetitive Behavior in Young Children with Autism Spectrum Disorder: A Randomized Trial. *The Journal of Pediatrics*, *170*, 45-53.e1-4. https://doi.org/10.1016/j.jpeds.2015.11.033

*Clark, M. L. E., Barbaro, J., & Dissanayake, C. (2017). Continuity and Change in Cognition and Autism Severity from Toddlerhood to School Age. *Journal of Autism and Developmental Disorders*, *47*(2), 328–339. https://doi.org/10.1007/s10803-016-2954-7

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

*Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental and*

Behavioral Pediatrics: JDBP, *27*(2 Suppl), S145-155. https://doi.org/10.1097/00004703-200604002-00013

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale: SRS-2*. Torrance, CA: Western Psychological Services.

*Corbett, B. A., Shickman, K., & Ferrer, E. (2008). Brief report: The effects of Tomatis sound therapy on language in children with autism. *Journal of Autism and Developmental Disorders*, *38*(3), 562–566. https://doi.org/10.1007/s10803-007-0413-1

Crank, J. E., Sandbank, M., Dunham, K., Crowley, S., Bottema-Beutel, K., Feldman, J., & Woynaroski, T. G. (2021). Understanding the Effects of Naturalistic Developmental Behavioral Interventions: A Project AIM Meta-analysis. *Autism Research*, *14*(4), 817–834. https://doi.org/10.1002/aur.2471

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. Developmental Psychology, 40, 271-283. https://doi.org/10.1037/0012-1649.40.2.271

*Dawson, G., Sun, J. M., Baker, J., Carpenter, K., Compton, S., Deaver, M., Franz, L., Heilbron, N., Herold, B., Horrigan, J., Howard, J., Kosinski, A., Major, S., Murias, M., Page, K., Prasad, V. K., Sabatos-DeVito, M., Sanfilippo, F., Sikich, L., … Kurtzberg, J. (2020). A Phase II Randomized Clinical Trial of the Safety and Efficacy of Intravenous Umbilical Cord Blood Infusion for Treatment of Children with Autism Spectrum Disorder. *The Journal of Pediatrics*, *222*, 164-173.e5. https://doi.org/10.1016/j.jpeds.2020.03.011

*Dean, O. M., Gray, K. M., Villagonzalo, K.-A., Dodd, S., Mohebbi, M., Vick, T., Tonge, B. J., & Berk, M. (2017). A randomised, double blind, placebo-controlled trial of a fixed dose of

N-acetyl cysteine in children with autistic disorder. *The Australian and New Zealand Journal of Psychiatry*, *51*(3), 241–249. https://doi.org/10.1177/0004867416652735

De Giacomo, A., & Fombonne, E. (1998). Parental recognition of developmental abnormalities in autism. *European child & adolescent psychiatry*, *7*(3), 131-136. https://doi.org/10.1007/s007870050058

*D'Elia, L., Valeri, G., Sonnino, F., Fontana, I., Mammone, A., & Vicari, S. (2014). A longitudinal study of the teacch program in different settings: The potential benefits of low intensity intervention in preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*(3), 615–626. https://doi.org/10.1007/s10803-013-1911-y

*Dereu, M., Roeyers, H., Raymaekers, R., Meirsschaut, M., & Warreyn, P. (2012). How useful are screening instruments for toddlers to predict outcome at age 4? General development, language skills, and symptom severity in children with a false positive screen for autism spectrum disorder. *European Child & Adolescent Psychiatry*, *21*(10), 541–551. https://doi.org/10.1007/s00787-012-0280-y

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, *7*(3), 177-188. https://doi.org/10.1016/0197-2456(86)90046-2

*Doussard-Roosevelt, J. A., Joe, C. M., Bazhenova, O. V., & Porges, S. W. (2003). Mother-child interaction in autistic and nonautistic children: Characteristics of maternal approach behaviors and child social responses. *Development and Psychopathology*, *15*(2), 277–295. https://doi.org/10.1017/s0954579403000154

*Duifhuis, E. A., den Boer, J. C., Doornbos, A., Buitelaar, J. K., Oosterling, I. J., & Klip, H. (2017). The Effect of Pivotal Response Treatment in Children with Autism Spectrum

Disorders: A Non-randomized Study with a Blinded Outcome Measure. *Journal of Autism and Developmental Disorders*, *47*(2), 231–242. https://doi.org/10.1007/s10803-016-2916-0

Dunn, L. M., & Dunn, L. M. (1997). *PPVT-III: Peabody picture vocabulary test*. American Guidance Service

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. https://doi.org/10.1111/j.0006-341X.2000.00455.x

*Duvekot, J., van der Ende, J., Verhulst, F. C., & Greaves-Lord, K. (2018). Examining bidirectional effects between the autism spectrum disorder (ASD) core symptom domains and anxiety in children with ASD. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *59*(3), 277–284. https://doi.org/10.1111/jcpp.12829

*Eapen, V., Crnčec, R., & Walter, A. (2013). Clinical outcomes of an early intervention program for preschool children with Autism Spectrum Disorder in a community group setting. *BMC Pediatrics*, *13*(1), 3. https://doi.org/10.1186/1471-2431-13-3

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of life research*, *16*(1), 5-18. https://doi.org/10.1007/s11136-007-9198-0

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

*Eikeseth, S., Klintwall, L., Jahr, E., & Karlsson, P. (2012). Outcome for children with autism receiving early and intensive behavioral intervention in mainstream preschool and kindergarten settings. *Research in Autism Spectrum Disorders*, *6*(2), 829–835. https://doi.org/10.1016/j.rasd.2011.09.002

*Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4- to 7-year-old children with autism. A 1-year comparison controlled study. *Behavior Modification*, *26*(1), 49–68. https://doi.org/10.1177/0145445502026001004

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. Journal of Clinical Child & Adolescent Psychology, 38(3), 439-450. https://doi.org/10.1080/15374410902851739

*Emberti Gialloreti, L., Benvenuto, A., Battan, B., Benassi, F., & Curatolo, P. (2016). Can biological components predict short-term evolution in Autism Spectrum Disorders? A proof-of-concept study. *Italian Journal of Pediatrics*, *42*. https://doi.org/10.1186/s13052-016-0281-4

*Fava, L., Strauss, K., Valeri, G., D'Elia, L., Arima, S., & Vicari, S. (2011). The effectiveness of a cross-setting complementary staff- and parent-mediated early intensive behavioral intervention for young children with ASD. *Research in Autism Spectrum Disorders*, *5*(4), 1479–1492. https://doi.org/10.1016/j.rasd.2011.02.009

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD: Paul H. Brookes Publishing Company.

*Fletcher-Watson, S., Petrou, A., Scott-Barrett, J., Dicks, P., Graham, C., O'Hare, A., Pain, H., & McConachie, H. (2016). A trial of an iPad™ intervention targeting social communication skills in children with autism. *Autism: The International Journal of Research and Practice*, *20*(7), 771–782. https://doi.org/10.1177/1362361315605624

*Fox, M. (2019). *Using the Brief Observation of Social Communication Change (BOSCC) to Evaluate Change in Parent-child Interaction Following Pivotal Response Treatment (PRT)* [Psy.D., Palo Alto University]. https://search.proquest.com/docview/2320973974/abstract/568878DD0E864697PQ/1

*Freeman, B. J., Del'Homme, M., Guthrie, D., & Zhang, F. (1999). Vineland Adaptive Behavior Scale scores as a function of age and initial IQ in 210 autistic children. *Journal of Autism and Developmental Disorders*, *29*(5), 379–384. https://doi.org/10.1023/a:1023078827457

French, L., & Kennedy, E. M. (2018). Annual Research Review: Early intervention for infants and young children with, or at-risk of, autism spectrum disorder: a systematic review. Journal of Child Psychology and psychiatry, 59(4), 444-456. https://doi.org/10.1111/jcpp.12828

*Frye, R. E., DeLatorre, R., Taylor, H. B., Slattery, J., Melnyk, S., Chowdhury, N., & James, S. J. (2013). Metabolic effects of sapropterin treatment in autism spectrum disorder: A preliminary study. *Translational Psychiatry*, *3*, e237. https://doi.org/10.1038/tp.2013.14

*Frye, R. E., Slattery, J., Delhey, L., Furgerson, B., Strickland, T., Tippett, M., Sailey, A., Wynne, R., Rose, S., Melnyk, S., Jill James, S., Sequeira, J. M., & Quadros, E. V. (2018). Folinic acid improves verbal communication in children with autism and language impairment: A randomized double-blind placebo-controlled trial. *Molecular Psychiatry*, *23*(2), 247–256. https://doi.org/10.1038/mp.2016.168

Fuller, E. A., & Kaiser, A. P. (2020). The Effects of Early Intervention on Social Communication Outcomes for Children with Autism Spectrum Disorder: A Meta-analysis.

Journal of Autism and Developmental Disorders, *50*(5), 1683–1700. https://doi.org/10.1007/s10803-019-03927-z

*Furukawa, K., Okuno, H., Mohri, I., Nakanishi, M., Eyberg, S. M., & Sakai, S. (2018). Effectiveness of Child-Directed Interaction Training for Young Japanese Children With Autism Spectrum Disorders. *Child & Family Behavior Therapy*, *40*(2), 166–186. https://doi.org/10.1080/07317107.2018.1477344

*Gammer, I., Bedford, R., Elsabbagh, M., Garwood, H., Pasco, G., Tucker, L., Volein, A., Johnson, M. H., & Charman, T. (2015). Behavioural markers for autism in infancy: Scores on the Autism Observational Scale for Infants in a prospective study of at-risk siblings. *Infant Behavior & Development*, *38*, 107–115. https://doi.org/10.1016/j.infbeh.2014.12.017

Gardner, M. F. (1990). *Expressive One-word Picture Vocabulary Test, Revised: EO-WPVT-R*. Academic Therapy Publications.

*Gengoux, G. W., Berquist, K. L., Salzman, E., Schapp, S., Phillips, J. M., Frazier, T. W., Minjarez, M. B., & Hardan, A. Y. (2015). Pivotal Response Treatment Parent Training for Autism: Findings from a 3-Month Follow-Up Evaluation. *Journal of Autism and Developmental Disorders*, *45*(9), 2889–2898. https://doi.org/10.1007/s10803-015-2452-3

*Gengoux, G. W., Schapp, S., Burton, S., Ardel, C. M., Libove, R. A., Baldi, G., Berquist, K. L., Phillips, J. M., & Hardan, A. Y. (2019). Effects of a parent-implemented Developmental Reciprocity Treatment Program for children with autism spectrum disorder. *Autism: The International Journal of Research and Practice*, *23*(3), 713–725. https://doi.org/10.1177/1362361318775538

*Gerlach-McDonald, B. (2017). *Developmental trajectories of adaptive behavior in Autism Spectrum Disorder and Down syndrome* (Doctoral dissertation, Colorado State University).

Geschwind, D. H. (2008). Autism: many genes, common pathways?. Cell, 135(3), 391-395. https://doi.org/10.1016/j.cell.2008.10.016

*Ginn, N. C., Clionsky, L. N., Eyberg, S. M., Warner-Metzger, C., & Abner, J.-P. (2017a). Child-Directed Interaction Training for Young Children With Autism Spectrum Disorders: Parent and Child Outcomes. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *46*(1), 101–109. https://doi.org/10.1080/15374416.2015.1015135

*Ginn, N. C., Clionsky, L. N., Eyberg, S. M., Warner-Metzger, C., & Abner, J.-P. (2017b). Child-Directed Interaction Training for Young Children With Autism Spectrum Disorders: Parent and Child Outcomes. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *46*(1), 101–109. https://doi.org/10.1080/15374416.2015.1015135

*Goods, K. S., Ishijima, E., Chang, Y.-C., & Kasari, C. (2013). Preschool Based JASPER Intervention in Minimally Verbal Children with Autism: Pilot RCT. *Journal of Autism and Developmental Disorders*, *43*(5), 1050–1056. https://doi.org/10.1007/s10803-012-1644-3

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*(5), 693-705. https://doi.org/10.1007/s10803-008-0674-3

Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, *130*(5), e1278-e1284. https://doi.org/10.1542/peds.2011-3668

*Granpeesheh, D., Tarbox, J., Dixon, D. R., Wilke, A. E., Allen, M. S., & Bradstreet, J. J. (2010). Randomized trial of hyperbaric oxygen therapy for children with autism. *Research in Autism Spectrum Disorders*, *4*(2), 268–275. https://doi.org/10.1016/j.rasd.2009.09.014

*Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., Howlin, P., Le Couteur, A., Leadbitter, K., Hudry, K., Byford, S., Barrett, B., Temple, K., Macdonald, W., & Pickles, A. (2010). Parent-mediated communication-focused treatment in children with autism (PACT): A randomised controlled trial. *Lancet*, *375*(9732), 2152–2160. https://doi.org/10.1016/S0140-6736(10)60587-9

Greenspan, S. I., DeGangi, G., & Wieder, S. (2001). *The Functional Emotional Assessment Scale (FEAS): For infancy & early childhood*. Interdisciplinary Council on Development & Learning Disorders.

Gresham, F., & Elliott, S. N. (2007). *Social skills improvement system (SSIS) rating scales*. San Antonio, TX: Pearson Education Inc.

Griffiths, R. (1970). The abilities of young children. A study in mental measurement.

Grzadzinski, R., Carr, T., Colombi, C., McGuire, K., Dufek, S., Pickles, A., & Lord, C. (2016). Measuring changes in social communication behaviors: preliminary development of the Brief Observation of Social Communication Change (BOSCC). Journal of autism and developmental disorders, 46(7), 2464-2479. https://doi.org/10.1007/s10803-016-2782-9

*Grzadzinski, R., Carr, T., Colombi, C., McGuire, K., Dufek, S., Pickles, A., & Lord, C. (2016). Measuring Changes in Social Communication Behaviors: Preliminary Development of the Brief Observation of Social Communication Change (BOSCC). *Journal of Autism and Developmental Disorders*, *46*(7), 2464–2479. https://doi.org/10.1007/s10803-016-2782-9

Grzadzinski, R., Janvier, D., & Kim, S. H. (2020). Recent Developments in Treatment Outcome Measures for Young Children With Autism Spectrum Disorder (ASD). *Seminars in Pediatric Neurology*, *34*, 100806. https://doi.org/10.1016/j.spen.2020.100806

Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. Journal of Child Psychology and Psychiatry, 54(5), 582-590. https://doi.org/10.1111/jcpp.12008

*Haglund, N., Dahlgren, S., Råstam, M., Gustafsson, P., & Källén, K. (2020). Improvement of Autism Symptoms After Comprehensive Intensive Early Interventions in Community Settings. *Journal of the American Psychiatric Nurses Association*, 1078390320915257. https://doi.org/10.1177/1078390320915257

*Hahn, L. J., Brady, N. C., & Versaci, T. (2019). Communicative Use of Triadic Eye Gaze in Children With Down Syndrome, Autism Spectrum Disorder, and Other Intellectual and Developmental Disabilities. *American Journal of Speech-Language Pathology*, *28*(4), 1509–1522. https://doi.org/10.1044/2019_AJSLP-18-0155

*Hale, C. M., & Tager-Flusberg, H. (2005). Social communication in children with autism: The relationship between theory of mind and discourse development. *Autism: The International Journal of Research and Practice*, *9*(2), 157–178. https://doi.org/10.1177/1362361305051395

*Handen, B. L., Johnson, C. R., Butter, E. M., Lecavalier, L., Scahill, L., Aman, M. G., McDougle, C. J., Arnold, L. E., Swiezy, N. B., Sukhodolsky, D. G., Mulick, J. A., White, S. W., Bearss, K., Hollway, J. A., Stigler, K. A., Dziura, J., Yu, S., Sacco, K., & Vitiello, B. (2013). Use of a Direct Observational Measure in a Trial of Risperidone and Parent

Training in Children with Pervasive Developmental Disorders. *Journal of Developmental and Physical Disabilities*, *25*(3), 355–371. https://doi.org/10.1007/s10882-012-9316-y

*Hardan, A. Y., Fung, L. K., Libove, R. A., Obukhanych, T. V., Nair, S., Herzenberg, L. A., Frazier, T. W., & Tirouvanziam, R. (2012). A randomized controlled pilot trial of oral N-acetylcysteine in children with autism. *Biological Psychiatry*, *71*(11), 956–961. https://doi.org/10.1016/j.biopsych.2012.01.014

*Hardan, A. Y., Gengoux, G. W., Berquist, K. L., Libove, R. A., Ardel, C. M., Phillips, J., Frazier, T. W., & Minjarez, M. B. (2015). A randomized controlled trial of Pivotal Response Treatment Group for parents of children with autism. *Journal of Child Psychology and Psychiatry*, *56*(8), 884–892. https://doi.org/10.1111/jcpp.12354

Harrer, M., Cuijpers, P., Furukawa, T. & Ebert, D. D. (2019). dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'. R package version 0.0.9000. URL http://dmetar.protectlab.org/.

Harrer, M., Cuijpers, P., Furukawa, T.A, & Ebert, D. D. (2019). Doing Meta-Analysis in R: A Hands-on Guide. https://10.5281/zenodo.2551803.

*Harris, S. L., Handleman, J. S., Gordon, R., Kristoff, B., & Fuentes, F. (1991). Changes in cognitive and language functioning of preschool children with autism. *Journal of Autism and Developmental Disorders*, *21*(3), 281–290. https://doi.org/10.1007/BF02207325

*Hazlett, H. C., Poe, M. D., Lightbody, A. A., Styner, M., MacFall, J. R., Reiss, A. L., & Piven, J. (2012). Trajectories of early brain volume development in fragile X syndrome and autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(9), 921–933. https://doi.org/10.1016/j.jaac.2012.07.003

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis*. Academic press.

*Hendren, R. L., James, S. J., Widjaja, F., Lawton, B., Rosenblatt, A., & Bent, S. (2016). Randomized, Placebo-Controlled Trial of Methyl B12 for Children with Autism. *Journal of Child and Adolescent Psychopharmacology*, *26*(9), 774–783. https://doi.org/10.1089/cap.2015.0159

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560. https://doi.org/10.1136/bmj.327.7414.557

*Hinnebusch, A. J., Miller, L. E., & Fein, D. A. (2017). Autism Spectrum Disorders and Low Mental Age: Diagnostic Stability and Developmental Outcomes in Early Childhood. *Journal of Autism and Developmental Disorders*, *47*(12), 3967–3982. https://doi.org/10.1007/s10803-017-3278-y

*Ho Yan, L. (2018). *Evaluation of a home-based, transdisciplinary intervention for autism spectrum disorder in Hong Kong* (1994018041; 2017-54456-028) [Psy.D., Alliant International University]. PsycINFO. https://search.proquest.com/docview/1994018041?accountid=14512

*Holzinger, D., Laister, D., Vivanti, G., Barbaresi, W. J., & Fellinger, J. (2019). Feasibility and Outcomes of the Early Start Denver Model Implemented with Low Intensity in a Community Setting in Austria. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *40*(5), 354–363. https://doi.org/10.1097/DBP.0000000000000675

*Howlin, P., Gordon, R. K., Pasco, G., Wade, A., & Charman, T. (2007). The effectiveness of Picture Exchange Communication System (PECS) training for teachers of children with autism: A pragmatic, group randomised controlled trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *48*(5), 473–481. https://doi.org/10.1111/j.1469-7610.2006.01707.x

*Ibañez, L. V., Kobak, K., Swanson, A., Wallace, L., Warren, Z., & Stone, W. L. (2018). Enhancing interactions during daily routines: A randomized controlled trial of a web-based tutorial for parents of young children with ASD. *Autism Research: Official Journal of the International Society for Autism Research*, *11*(4), 667–678. https://doi.org/10.1002/aur.1919

*Ingersoll, B. (2012). Brief Report: Effect of a Focused Imitation Intervention on Social Functioning in Children with Autism. *Journal of Autism and Developmental Disorders*, *42*(8), 1768–1773. https://doi.org/10.1007/s10803-011-1423-6

*Ingersoll, B. R., & Wainer, A. L. (2013). Pilot study of a school-based parent training program for preschoolers with ASD. *Autism: The International Journal of Research and Practice*, *17*(4), 434–448. https://doi.org/10.1177/1362361311427155

*Ingersoll, B., Wainer, A. L., Berger, N. I., Pickard, K. E., & Bonter, N. (2016). Comparison of a Self-Directed and Therapist-Assisted Telehealth Parent-Mediated Intervention for Children with ASD: A Pilot RCT. *Journal of Autism and Developmental Disorders*, *46*(7), 2275–2284. https://doi.org/10.1007/s10803-016-2755-z

*Jones, R. M., Carberry, C., Hamo, A., & Lord, C. (2017). Placebo-like response in absence of treatment in children with Autism. *Autism Research: Official Journal of the International Society for Autism Research*, *10*(9), 1567–1572. https://doi.org/10.1002/aur.1798

*Kaale, A., Fagerland, M. W., Martinsen, E. W., & Smith, L. (2014). Preschool-based social communication treatment for children with autism: 12-month follow-up of a randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(2), 188–198. https://doi.org/10.1016/j.jaac.2013.09.019

*Kaale, A., Smith, L., & Sponheim, E. (2012). A randomized controlled trial of preschool-based joint attention intervention for children with autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *53*(1), 97–105. https://doi.org/10.1111/j.1469-7610.2011.02450.x

*Kamps, D., Thiemann-Bourque, K., Heitzman-Powell, L., Schwartz, I., Rosenberg, N., Mason, R., & Cox, S. (2015). A Comprehensive Peer Network Intervention to Improve Social Communication of Children with Autism Spectrum Disorders: A Randomized Trial in Kindergarten and First Grade. *Journal of Autism and Developmental Disorders*, *45*(6), 1809–1824. https://doi.org/10.1007/s10803-014-2340-2

Kanner, L. (1943). Autistic disturbances of affective contact. Nervous child, 2(3), 217-250.

Kasari, C. (2002). Assessing Change in Early Intervention Programs for Children with Autism. *Journal of Autism and Developmental Disorders*, *32*(5), 447–461. https://doi.org/10.1023/A:1020546006971

*Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(6), 611–620. https://doi.org/10.1111/j.1469-7610.2005.01567.x

*Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. Journal of Consulting and Clinical Psychology, 76, 125-137. doi:10.1037/0022-006x.76.1.125

*Kasari, C., Gulsrud, A. C., Wong, C., Kwon, S., & Locke, J. (2010). Randomized Controlled Caregiver Mediated Joint Engagement Intervention for Toddlers with Autism. *Journal of*

Autism and Developmental Disorders, *40*(9), 1045–1056. https://doi.org/10.1007/s10803-010-0955-5

*Kasari, C., Gulsrud, A., Paparella, T., Hellemann, G., & Berry, K. (2015). Randomized comparative efficacy study of parent-mediated interventions for toddlers with autism. *Journal of Consulting and Clinical Psychology*, *83*(3), 554–563. https://doi.org/10.1037/a0039080

*Kasari, C., Kaiser, A., Goods, K., Nietfeld, J., Mathy, P., Landa, R., Murphy, S., & Almirall, D. (2014). Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(6), 635–646. https://doi.org/10.1016/j.jaac.2014.01.019

*Kasari, C., Lawton, K., Shih, W., Barker, T. V., Landa, R., Lord, C., Orlich, F., King, B., Wetherby, A., & Senturk, D. (2014). Caregiver-Mediated Intervention for Low-Resourced Preschoolers With Autism: An RCT. *Pediatrics*, *134*(1), e72–e79. https://doi.org/10.1542/peds.2013-3229

*Kim, S. H., Grzadzinski, R., Martinez, K., & Lord, C. (2019). Measuring treatment response in children with autism spectrum disorder: Applications of the Brief Observation of Social Communication Change to the Autism Diagnostic Observation Schedule. *Autism: The International Journal of Research and Practice*, *23*(5), 1176–1185. https://doi.org/10.1177/1362361318793253

*Kitzerow, J., Teufel, K., Jensen, K., Wilker, C., & Freitag, C. M. (2020). Case-control study of the low intensive autism-specific early behavioral intervention A-FFIP: Outcome after one year. *Zeitschrift Fur Kinder- Und Jugendpsychiatrie Und Psychotherapie*, *48*(2), 103–112. https://doi.org/10.1024/1422-4917/a000661

*Kitzerow, J., Teufel, K., Wilker, C., & Freitag, C. M. (2016). Using the brief observation of social communication change (BOSCC) to measure autism-specific development. *Autism Research: Official Journal of the International Society for Autism Research*, *9*(9), 940–950. https://doi.org/10.1002/aur.1588

*Klaiman, C., Huffman, L., Masaki, L., & Elliott, G. R. (2013). Tetrahydrobiopterin as a treatment for autism spectrum disorders: A double-blind, placebo-controlled trial. *Journal of Child and Adolescent Psychopharmacology*, *23*(5), 320–328. https://doi.org/10.1089/cap.2012.0127

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in medicine*, *22*(17), 2693-2710. https://doi.org/10.1002/sim.1482

*Kovshoff, H., Hastings, R. P., & Remington, B. (2011). Two-year outcomes for children with autism after the cessation of early intensive behavioral intervention. *Behavior Modification*, *35*(5), 427–450. https://doi.org/10.1177/0145445511405513

*Krusch, D. A. (2004). *Effects of repeated secretin administration on a subset of children with pervasive developmental disorder* [Ph.D., University of Rochester]. https://search.proquest.com/docview/305153671/abstract/954D7B6159054396PQ/1

Kuhfeld, M., & Sturm, A. (2017). An Examination of the Precision of the Autism Diagnostic Observation Schedule Using Item Response Theory. *Psychological Assessment*. PsycINFO. https://doi.org/10.1037/pas0000512

*LaGasse, A. B. (2014). Effects of a music therapy group intervention on enhancing social skills in children with autism. *Journal of Music Therapy*, *51*(3), 250–275. https://doi.org/10.1093/jmt/thu012

*Landa, R. J., Holman, K. C., O'Neill, A. H., & Stuart, E. A. (2011). Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *52*(1), 13–21. https://doi.org/10.1111/j.1469-7610.2010.02288.x

*Lawson, L. P., Joshi, R., Barbaro, J., & Dissanayake, C. (2018). Gender Differences During Toddlerhood in Autism Spectrum Disorder: A Prospective Community-Based Longitudinal Follow-Up Study. *Journal of Autism and Developmental Disorders*, *48*(8), 2619–2628. https://doi.org/10.1007/s10803-018-3516-y

*Leaf, J. B., Leaf, J. A., Milne, C., Taubman, M., Oppenheim-Leaf, M., Torres, N., Townley-Cochran, D., Leaf, R., McEachin, J., Yoder, P., & Autism Partnership Foundation. (2017). An Evaluation of a Behaviorally Based Social Skills Group for Individuals Diagnosed with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(2), 243–259. https://doi.org/10.1007/s10803-016-2949-4

*Lee, R. W. Y., Corley, M. J., Pang, A., Arakaki, G., Abbott, L., Nishimoto, M., Miyamoto, R., Lee, E., Yamamoto, S., Maunakea, A. K., Lum-Jones, A., & Wong, M. (2018). A modified ketogenic gluten-free diet with MCT improves behavior in children with autism spectrum disorder. *Physiology & Behavior*, *188*, 205–211. https://doi.org/10.1016/j.physbeh.2018.02.006

*Lei, J., Sukhodolsky, D. G., Abdullahi, S. M., Braconnier, M. L., & Ventola, P. (2017). Brief report: Reduced anxiety following Pivotal Response Treatment in young children with

Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *43–44*, 1–7. https://doi.org/10.1016/j.rasd.2017.09.002

*Lemonnier, E., Degrez, C., Phelep, M., Tyzio, R., Josse, F., Grandgeorge, M., Hadjikhani, N., & Ben-Ari, Y. (2012). A randomised controlled trial of bumetanide in the treatment of autism in children. *Translational Psychiatry*, *2*(12), e202. https://doi.org/10.1038/tp.2012.124

*Leonard, H. C., Bedford, R., Pickles, A., & Hill, E. L. (2015). Predicting the rate of language development from early motor skills in at-risk infants who develop autism spectrum disorder. *Research in Autism Spectrum Disorders*, *13–14*, 15–24. https://doi.org/10.1016/j.rasd.2014.12.012

*Lerna, A., Esposito, D., Conson, M., & Massagli, A. (2014). Long-term effects of PECS on social–communicative skills of children with autism spectrum disorders: A follow-up study. *International Journal of Language & Communication Disorders*, *49*(4), 478–485. https://doi.org/10.1111/1460-6984.12079

*Lerna, A., Esposito, D., Conson, M., Russo, L., & Massagli, A. (2012a). Social-communicative effects of the Picture Exchange Communication System (PECS) in autism spectrum disorders. *International Journal of Language & Communication Disorders*, *47*(5), 609–617. https://doi.org/10.1111/j.1460-6984.2012.00172.x

*Lerna, A., Esposito, D., Conson, M., Russo, L., & Massagli, A. (2012b). Social-communicative effects of the Picture Exchange Communication System (PECS) in autism spectrum disorders. *International Journal of Language & Communication Disorders*, *47*(5), 609–617. https://doi.org/10.1111/j.1460-6984.2012.00172.x

*Levy, S. E., Souders, M. C., Wray, J., Jawad, A. F., Gallagher, P. R., Coplan, J., Belchic, J. K., Gerdes, M., Mitchell, R., & Mulberg, A. E. (2003). Children with autistic spectrum disorders. I: Comparison of placebo and single dose of human synthetic secretin. *Archives of Disease in Childhood*, *88*(8), 731–736. https://doi.org/10.1136/adc.88.8.731

*Liao, S.-T., Hwang, Y.-S., Chen, Y.-J., Lee, P., Chen, S.-J., & Lin, L.-Y. (2014). Home-based DIR/Floortime intervention program for preschool children with autism spectrum disorders: Preliminary findings. *Physical & Occupational Therapy in Pediatrics*, *34*(4), 356–367. https://doi.org/10.3109/01942638.2014.918074

*Lightdale, J. R., Hayer, C., Duer, A., Lind-White, C., Jenkins, S., Siegel, B., Elliott, G. R., & Heyman, M. B. (2001b). Effects of intravenous secretin on language and behavior of children with autism and gastrointestinal symptoms: A single-blinded, open-label pilot study. *Pediatrics*, *108*(5), E90. https://doi.org/10.1542/peds.108.5.e90

*Lopata, C., McDonald, C. A., Thomeer, M. L., Donnelly, J. P., Jordan, A. K., & Rodgers, J. D. (2018). Pilot Trial of a Comprehensive Summer Psychosocial Treatment for High-Functioning Young Children with ASD. *Journal of Developmental and Physical Disabilities*, *30*(4), 439–454. https://doi.org/10.1007/s10882-018-9595-z

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ... & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, *30*(3), 205-223. https://doi.org/10.1023/A:1005592401947

Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., Dawson, G., Durand, V. M., Grossman, L., Guthrie, D., Harris, S., Kasari, C., Marcus, L., Murphy, S., Odom, S., Pickles, A., Scahill, L., Shaw, E., Siegel, B., … Yoder, P. (2005). Challenges in Evaluating Psychosocial Interventions for Autistic Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(6), 695–708. https://doi.org/10.1007/s10803-005-0017-6

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule, second edition: ADOS-2. Los Angeles: Western Psychological Services.

Lord, C., Bishop, S., & Anderson, D. (2015). Developmental trajectories as autism phenotypes. *American Journal of Medical Genetics Part C-Seminars in Medical Genetics*, *169*(2), 198–208. https://doi.org/10.1002/ajmg.c.31440

*Lotfizadeh, A. D., Kazemi, E., Pompa-Craven, P., & Eldevik, S. (2020). Moderate Effects of Low-Intensity Behavioral Intervention. *Behavior Modification*, *44*(1), 92–113. https://doi.org/10.1177/0145445518796204

*Mahoney, G., & Perales, F. (2003). Using Relationship-Focused Intervention to Enhance the Social—Emotional Functioning of Young Children with Autism Spectrum Disorders. *Topics in Early Childhood Special Education*, *23*(2), 74–86. https://doi.org/10.1177/02711214030230020301

*Mankad, D., Dupuis, A., Smile, S., Roberts, W., Brian, J., Lui, T., Genore, L., Zaghloul, D., Iaboni, A., Marcon, P. M. A., & Anagnostou, E. (2015). A randomized, placebo controlled trial of omega-3 fatty acids in the treatment of young children with autism. *Molecular Autism*, *6*, 18. https://doi.org/10.1186/s13229-015-0010-7

*Mazahery, H., Conlon, C. A., Beck, K. L., Mugridge, O., Kruger, M. C., Stonehouse, W., Camargo, C. A., Meyer, B. J., Tsang, B., Jones, B., & von Hurst, P. R. (2019). A Randomised-Controlled Trial of Vitamin D and Omega-3 Long Chain Polyunsaturated

Fatty Acids in the Treatment of Core Symptoms of Autism Spectrum Disorder in Children. *Journal of Autism and Developmental Disorders*, *49*(5), 1778–1794. https://doi.org/10.1007/s10803-018-3860-y

Mazurek, M. O., Carlson, C., Baker-Ericzén, M., Butter, E., Norris, M., Barr, C., & Kanne, S. (2020). The Autism Impact Measure (AIM): Examination of Sensitivity to Change. *Autism Research*, *13*(11), 1867-1879. https://doi.org/10.1002/aur.2397

*McCalla, M. K. (2015). *Enhancing the social interactions of preschool children with autism: The effectiveness of a peer-mediated intervention* (1690656904; 2015-99141-070). PsycINFO. https://search.proquest.com/docview/1690656904?accountid=14512

McConachie H, Parr JR, Glod M, Hanratty J, Livingstone N, Oono IP, *et al.* Systematic review of tools to measure outcomes for young children with autism spectrum disorder. *Health Technol Assess* 2015;19(41). https://doi.org/10.3310/hta19410

*Melamed, I. R., Heffron, M., Testori, A., & Lipe, K. (2018). A pilot study of high-dose intravenous immunoglobulin 5% for autism: Impact on autism spectrum and markers of neuroinflammation. *Autism Research: Official Journal of the International Society for Autism Research*, *11*(3), 421–433. https://doi.org/10.1002/aur.1906

*Messinger, D. S., Young, G. S., Webb, S. J., Ozonoff, S., Bryson, S. E., Carter, A., Carver, L., Charman, T., Chawarska, K., Curtin, S., Dobkins, K., Hertz-Picciotto, I., Hutman, T., Iverson, J. M., Landa, R., Nelson, C. A., Stone, W. L., Tager-Flusberg, H., & Zwaigenbaum, L. (2015). Early sex differences are not autism-specific: A Baby Siblings Research Consortium (BSRC) study. *Molecular Autism*, *6*, 32. https://doi.org/10.1186/s13229-015-0027-y

Mokkink, L.B., Terwee, C.B., Patrick, D.L. *et al.* The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* **19,** 539–549 (2010). https://doi.org/10.1007/s11136-010-9606-8

*Mouga, S., Correia, B. R., Café, C., Duque, F., & Oliveira, G. (2020). Language Predictors in Autism Spectrum Disorder: Insights from Neurodevelopmental Profile in a Longitudinal Perspective. *Journal of Abnormal Child Psychology*, *48*(1), 149–161. https://doi.org/10.1007/s10802-019-00578-7

Mullen, E. M. (1995). *Mullen scales of early learning* (pp. 58-64). Circle Pines, MN: AGS.

Mundy, P., Kasari, C., & Sigman, M. (1992). Nonverbal communication, affective sharing, and intersubjectivity. Infant Behavior and Development, 15(3), 377-381.

*Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 115–128. https://doi.org/10.1007/BF02206861

Mundy, P., Sigman, M., Kasari, C. (1994). Joint attention, developmental level, and symptom presentation in autism. Development and Psychopathology, 6(389–401), 1383–1398. https://doi.org/10.1016/0163-6383(92)80006-G

*Murias, M., Major, S., Compton, S., Buttinger, J., Sun, J. M., Kurtzberg, J., & Dawson, G. (2018). Electrophysiological Biomarkers Predict Clinical Improvement in an Open-Label Trial Assessing Efficacy of Autologous Umbilical Cord Blood for Treatment of Autism. *Stem Cells Translational Medicine*, *7*(11), 783–791. https://doi.org/10.1002/sctm.18-0090

Nahmias, A. S., Pellecchia, M., Stahmer, A. C., & Mandell, D. S. (2019). Effectiveness of community-based early intervention for children with autism spectrum disorder: A meta-analysis. *Journal of Child Psychology and Psychiatry*, *60*(11), 1200–1209. https://doi.org/10.1111/jcpp.13073

*Nordahl-Hansen, A., Fletcher-Watson, S., McConachie, H., & Kaale, A. (2016). Relations between specific and global outcome measures in a social-communication intervention for children with autism spectrum disorder. *Research in Autism Spectrum Disorders*, *29–30*, 19–29. https://doi.org/10.1016/j.rasd.2016.05.005

*O'Donnell, K. (2012). *Examining joint attention and its effect on skill acquisition in children with autism using a group drumming therapy approach* (1041000943; 2012-99130-555). PsycINFO. https://search.proquest.com/docview/1041000943?accountid=14512

Oono, I. P., Honey, E. J., & McConachie, H. (2013). Parent-mediated early intervention for young children with autism spectrum disorders (ASD). *The Cochrane Database of Systematic Reviews*, *4*, CD009774. https://doi.org/10.1002/14651858.CD009774.pub2

Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. Development and Psychopathology, 14. https://doi.org/10.1017/s0954579402002031.

*Oosterling, I., Visser, J., Swinkels, S., Rommelse, N., Donders, R., Woudenberg, T., Roos, S., van der Gaag, R. J., & Buitelaar, J. (2010). Randomized Controlled Trial of the Focus Parent Training for Toddlers with Autism: 1-Year Outcome. *Journal of Autism and Developmental Disorders*, *40*(12), 1447–1458. https://doi.org/10.1007/s10803-010-1004-0

*Owley, T., McMahon, W., Cook, E. H., Laulhere, T., South, M., Mays, L. Z., Shernoff, E. S., Lainhart, J., Modahl, C. B., Corsello, C., Ozonoff, S., Risi, S., Lord, C., Leventhal, B. L., & Filipek, P. A. (2001). Multisite, double-blind, placebo-controlled trial of porcine secretin in autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(11), 1293–1299. https://doi.org/10.1097/00004583-200111000-00009

*Paelt, S. V. der, Warreyn, P., & Roeyers, H. (2016). Effect of community interventions on social-communicative abilities of preschoolers with autism spectrum disorder. *Developmental Neurorehabilitation*, *19*(3), 162–174. https://doi.org/10.3109/17518423.2014.933983

*Pajareya, K., & Nopmaneejumruslers, K. (2011). A pilot randomized controlled trial of DIR/Floortime™ parent training intervention for pre-school children with autistic spectrum disorders. *Autism: The International Journal of Research and Practice*, *15*(5), 563–577. https://doi.org/10.1177/1362361310386502

*Pajareya, K., & Nopmaneejumruslers, K. (2012). A one-year prospective follow-up study of a DIR/Floortime parent training intervention for pre-school children with autistic spectrum disorders. *Journal of the Medical Association of Thailand = Chotmaihet Thangphaet*, *95*(9), 1184–1193.

*Panerai, S., Suraniti, G. S., Catania, V., Zingale, M., Ferri, R., Raggi, A., Trubia, G., & Elia, M. (2020). Early results from a combined low-intensive psychoeducational intervention for preschoolers with autism spectrum disorder. *Disability and Rehabilitation*, *42*(9), 1275–1283. https://doi.org/10.1080/09638288.2018.1522553

Tanya Paparella, K. S. G. (2011). The emergence of nonverbal joint attention and requesting skills in young children with autism. *Journal of Communication Disorders*, *44*(6), 569–583. https://doi.org/10.1016/j.jcomdis.2011.08.002.

*Parladé, M. V., Weinstein, A., Garcia, D., Rowley, A. M., Ginn, N. C., & Jent, J. F. (2020). Parent-Child Interaction Therapy for children with autism spectrum disorder and a matched case-control sample. *Autism: The International Journal of Research and Practice*, *24*(1), 160–176. https://doi.org/10.1177/1362361319855851

*Parr, J. R., Gray, L., Wigham, S., McConachie, H., & Couteur, A. L. (2015). Measuring the relationship between the parental Broader Autism Phenotype, parent–child interaction, and children's progress following parent mediated intervention. *Research in Autism Spectrum Disorders*, *20*, 24–30. https://doi.org/10.1016/j.rasd.2015.07.006

*Parsons, D., Cordier, R., Lee, H., Falkmer, T., & Vaz, S. (2019). A Randomised Controlled Trial of an Information Communication Technology Delivered Intervention for Children with Autism Spectrum Disorder Living in Regional Australia. *Journal of Autism and Developmental Disorders*, *49*(2), 569–581. https://doi.org/10.1007/s10803-018-3734-3

*Paul, R., Campbell, D., Gilbert, K., & Tsiouri, I. (2013). Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(2), 418–431. https://doi.org/10.1007/s10803-012-1583-z

*Paynter, J., Trembath, D., & Lane, A. (2018). Differential outcome subgroups in children with autism spectrum disorder attending early intervention. *Journal of Intellectual Disability Research: JIDR*, *62*(7), 650–659. https://doi.org/10.1111/jir.12504

*Pellicano, E. (2012). Do autistic symptoms persist across time? Evidence of substantial change in symptomatology over a 3-year period in cognitively able children with autism. *American Journal on Intellectual and Developmental Disabilities*, *117*(2), 156–166. https://doi.org/10.1352/1944-7558-117.2.156

Pender, R., Fearon, P., Heron, J., & Mandy, W. (2020). The longitudinal heterogeneity of autistic traits: A systematic review. *Research in Autism Spectrum Disorders*, *79*, 101671. https://doi.org/10.1016/j.rasd.2020.101671

*Perrin, N. L. (2012). *A Mixed Methods Investigation of the Kansas Autism Waiver's Implementation: The First Three Years*. Undefined. /paper/A-Mixed-Methods-Investigation-of-the-Kansas-Autism-Perrin/49481a7cadcfe27654cd5a3da9b8e841844f0f78

*Perryman, T. Y., Carter, A. S., Messinger, D. S., Stone, W. L., Ivanescu, A. E., & Yoder, P. J. (2013). Brief Report: Parental Child-Directed Speech as a Predictor of Receptive Language in Children with Autism Symptomatology. *Journal of Autism and Developmental Disorders*, *43*(8), 1983–1987. https://doi.org/10.1007/s10803-012-1725-3

Pickles, A., Harris, V., Green, J., Aldred, C., McConachie, H., Slonims, V., ... & PACT Consortium. (2015). Treatment mechanism in the MRC preschool autism communication trial: Implications for study design and parent-focussed therapy for children. *Journal of Child Psychology and Psychiatry*, *56*(2), 162-170. https://doi.org/10.1111/jcpp.12291

*Pickles, A., Couteur, A. L., Leadbitter, K., Salomone, E., Cole-Fletcher, R., Tobin, H., Gammer, I., Lowry, J., Vamvakas, G., Byford, S., Aldred, C., Slonims, V., McConachie, H., Howlin, P., Parr, J. R., Charman, T., & Green, J. (2016). Parent-mediated social communication therapy for young children with autism (PACT): Long-term follow-up of a randomised controlled trial. *The Lancet*, *388*(10059), 2501–2509. https://doi.org/10.1016/S0140-6736(16)31229-6

*Pijl, M. K., Rommelse, N. N., Hendriks, M., De Korte, M. W., Buitelaar, J. K., & Oosterling, I. J. (2018). Does the Brief Observation of Social Communication Change help moving forward in measuring change in early autism intervention studies? *Autism: The*

International Journal of Research and Practice, *22*(2), 216–226. https://doi.org/10.1177/1362361316669235

*Pittala, E. T., Saint-Georges-Chaumet, Y., Favrot, C., Tanet, A., Cohen, D., & Saint-Georges, C. (2018). Clinical outcomes of interactive, intensive and individual (3i) play therapy for children with ASD: a two-year follow-up study. *Bmc Pediatrics*, *18*, 165. https://doi.org/10.1186/s12887-018-1126-7

*Poslawsky, I. E., Naber, F. B., Bakermans-Kranenburg, M. J., van Daalen, E., van Engeland, H., & van IJzendoorn, M. H. (2015). Video-feedback Intervention to promote Positive Parenting adapted to Autism (VIPP-AUTI): A randomized controlled trial. *Autism*, *19*(5), 588–603. https://doi.org/10.1177/1362361314537124

*Potter, L. A., Scholze, D. A., Biag, H. M. B., Schneider, A., Chen, Y., Nguyen, D. V., Rajaratnam, A., Rivera, S. M., Dwyer, P. S., Tassone, F., Al Olaby, R. R., Choudhary, N. S., Salcedo-Arellano, M. J., & Hagerman, R. J. (2019). A Randomized Controlled Trial of Sertraline in Young Children With Autism Spectrum Disorder. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00810

*Rahman, A., Divan, G., Hamdani, S. U., Vajaratkar, V., Taylor, C., Leadbitter, K., Aldred, C., Minhas, A., Cardozo, P., Emsley, R., Patel, V., & Green, J. (2016). Effectiveness of the parent-mediated intervention for children with autism spectrum disorder in south Asia in India and Pakistan (PASS): A randomised controlled trial. *The Lancet. Psychiatry*, *3*(2), 128–136. https://doi.org/10.1016/S2215-0366(15)00388-0

*Ray-Subramanian, C. E., & Weismer, S. E. (2012). Receptive and Expressive Language as Predictors of Restricted and Repetitive Behaviors in Young Children with Autism

Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *42*(10), 2113–2120. https://doi.org/10.1007/s10803-012-1463-6

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

*Reed, P., Osborne, L. A., & Corness, M. (2010). Effectiveness of special nursery provision for children with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, *14*(1), 67–82. https://doi.org/10.1177/1362361309340030

*Remington, B., Hastings, R. P., Kovshoff, H., degli Espinosa, F., Jahr, E., Brown, T., Alsford, P., Lemaic, M., & Ward, N. (2007). Early intensive behavioral intervention: Outcomes for children with autism and their parents after two years. *American Journal of Mental Retardation: AJMR*, *112*(6), 418–438. https://doi.org/10.1352/0895-8017(2007)112[418:EIBIOF]2.0.CO;2

Reynell, J. K., & Gruber, C. P. (1997). *Reynell developmental language scales*. Western Psychological Services.

Richler, J., Huerta, M., Bishop, S. L., & Lord, C. (2010). Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. Development and psychopathology, 22(1), 55-69. https://doi.org/10.1017/S0954579409990265.

*Rickards, A. L., Walstab, J. E., Wright-Rossi, R. A., Simpson, J., & Reddihough, D. S. (2009). One-year follow-up of the outcome of a randomized controlled trial of a home-based intervention programme for children with autism and developmental delay and their families. *Child: Care, Health and Development*, *35*(5), 593–602. https://doi.org/10.1111/j.1365-2214.2009.00953.x

*Roberts, J., Williams, K., Carter, M., Evans, D., Parmenter, T., Silove, N., Clark, T., & Warren, A. (2011). A randomised controlled trial of two early intervention programs for young children with autism: Centre-based with parent program and home-based. *Research in Autism Spectrum Disorders*, *5*(4), 1553–1566. https://doi.org/10.1016/j.rasd.2011.03.001

Robinson, E. B., St Pourcain, B., Anttila, V., Kosmicki, J. A., Bulik-Sullivan, B., Grove, J., ... & Martin, J. (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nature genetics, 48(5), 552. https://doi.org/10.1038/ng.3529

Rogers, S. J., & Vismara, L. A. (2008). Evidence-based comprehensive treatments for early autism. Journal of Clinical Child & Adolescent Psychology, 37(1), 8-38. https://doi.org/10.1080/15374410701817808

Rothstein, Hannah R, Alexander J Sutton, and Michael Borenstein. 2006. Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments. John Wiley & Sons.

*Ruiz, A. A. (n.d.). *Motivational Behavioral Group Treatment for Social Deficits in Preschool Children with Autism: Impact on Talkativeness with Typically Developing Peers* [Psy.D., Palo Alto University]. Retrieved May 30, 2021, from https://www.proquest.com/docview/2322786702/abstract/A776FC5D9802493FPQ/1

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Rutter, M., Le Couteur, A., & Lord, C. (2003). Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*, *29*(2003), 30.

Ruzzo, E. K., Pérez-Cano, L., Jung, J. Y., Wang, L. K., Kashef-Haghighi, D., Hartl, C., ... & Wall, D. P. (2019). Inherited and de novo genetic risk for autism impacts shared networks. *Cell*, *178*(4), 850-866. https://doi.org/10.1016/j.cell.2019.07.015

*Sainato, D. M., Morrison, R. S., Jung, S., Axe, J., & Nixon, P. A. (2015). A Comprehensive Inclusion Program for Kindergarten Children with Autism Spectrum Disorder. *Journal of Early Intervention*, *37*(3), 208–225. https://doi.org/10.1177/1053815115613836

*Salomone, E., Shephard, E., Milosavljevic, B., Johnson, M. H., Charman, T., & BASIS Team. (2018). Adaptive Behaviour and Cognitive Skills: Stability and Change from 7 Months to 7 Years in Siblings at High Familial Risk of Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *48*(9), 2901–2911. https://doi.org/10.1007/s10803-018-3554-5

*Salt, J., Shemilt, J., Sellars, V., Boyd, S., Coulson, T., & McCool, S. (2002). The Scottish Centre for autism preschool treatment programme. II: The results of a controlled treatment outcome study. *Autism: The International Journal of Research and Practice*, *6*(1), 33–46. https://doi.org/10.1177/1362361302006001004

Sandbank, M., Bottema-Beutel, K., Crowley, S., Cassidy, M., Dunham, K., Feldman, J. I., ... & Woynaroski, T. G. (2020). Project AIM: Autism intervention meta-analysis for studies of young children. *Psychological Bulletin*, *146*(1), 1. https://doi.org/10.1037/bul0000215

Sandbank, M., Chow, J., Bottema-Beutel, K., & Woynaroski, T. (2021). Evaluating evidence-based practice in light of the boundedness and proximity of outcomes: Capturing the scope of change. *Autism Research*, *n/a*(n/a). https://doi.org/10.1002/aur.2527

*Sandler, A. D., Sutton, K. A., DeWeese, J., Girardi, M. A., Sheppard, V., & Bodfish, J. W. (1999). Lack of benefit of a single dose of synthetic human secretin in the treatment of autism and pervasive developmental disorder. *The New England Journal of Medicine*, *341*(24), 1801–1806. https://doi.org/10.1056/NEJM199912093412404

*Scahill, L., Bearss, K., Lecavalier, L., Smith, T., Swiezy, N., Aman, M. G., Sukhodolsky, D. G., McCracken, C., Minshawi, N., Turner, K., Levato, L., Saulnier, C., Dziura, J., & Johnson, C. (2016). Effect of Parent Training on Adaptive Behavior in Children With Autism Spectrum Disorder and Disruptive Behavior: Results of a Randomized Trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*(7), 602-609.e3. https://doi.org/10.1016/j.jaac.2016.05.001

*Scahill, L., McDougle, C. J., Aman, M. G., Johnson, C., Handen, B., Bearss, K., Dziura, J., Butter, E., Swiezy, N. G., Arnold, L. E., Stigler, K. A., Sukhodolsky, D. D., Lecavalier, L., Pozdol, S. L., Nikolov, R., Hollway, J. A., Korzekwa, P., Gavaletz, A., Kohn, A. E., … Research Units on Pediatric Psychopharmacology Autism Network. (2012). Effects of risperidone and parent training on adaptive functioning in children with pervasive developmental disorders and serious behavioral problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(2), 136–146. https://doi.org/10.1016/j.jaac.2011.11.010

*Schertz, H. H., Odom, S. L., Baggett, K. M., & Sideris, J. H. (2013). Effects of Joint Attention Mediated Learning for toddlers with autism spectrum disorders: An initial randomized controlled study. *Early Childhood Research Quarterly*, *28*(2), 249–258. https://doi.org/10.1016/j.ecresq.2012.06.006

*Schertz, H. H., Odom, S. L., Baggett, K. M., & Sideris, J. H. (2018). Mediating Parent Learning to Promote Social Communication for Toddlers with Autism: Effects from a Randomized Controlled Trial. *Journal of Autism and Developmental Disorders*, *48*(3), 853–867. https://doi.org/10.1007/s10803-017-3386-8

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of autism and developmental disorders*, *10*(1), 91-103. https://doi.org/10.1007/BF02408436

*Schreibman, L. E., & Stahmer, A. C. (2014). A Randomized Trial Comparison of the Effects of Verbal and Pictorial Naturalistic Communication Strategies on Spoken Language for Young Children with Autism. *Journal of Autism and Developmental Disorders*, *44*(5), 1244–1251. https://doi.org/10.1007/s10803-013-1972-y

Schwarzer, Guido, James R Carpenter, and Gerta Rücker. 2015. Meta-Analysis with R. Springer.

*Scudder, A., Wong, C., Ober, N., Hoffman, M., Toscolani, J., & Handen, B. L. (2019). Parent–child interaction therapy (PCIT) in young children with autism spectrum disorder. *Child & Family Behavior Therapy*, *41*(4), 201–220. https://doi.org/10.1080/07317107.2019.1659542

Seibert, J. M., Hogan, A. E., & Mundy, P. C. (1982). Assessing interactional competencies: The early social-communication scales. *Infant mental health Journal*, *3*(4), 244-258. https://doi.org/10.1002/1097-0355(198224)3:4<244::AID-IMHJ2280030406>3.0.CO;2-R

*Shade-Monuteaux, D. M. (2003). *An innovative approach for children with autism spectrum disorders: A preliminary outcome evaluation* (620260354; 2003-95012-118). PsycINFO. https://search.proquest.com/docview/620260354?accountid=14512

Shih, W., Shire, S., Chang, Y. C., & Kasari, C. (2021). Joint engagement is a potential mechanism leading to increased initiations of joint attention and downstream effects on language: JASPER early intervention for children with ASD. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.13405

*Shire, S. Y., Baker Worthman, L., Shih, W., & Kasari, C. (2020). Comparison of Face-to-Face and Remote Support for Interventionists Learning to Deliver JASPER Intervention with Children Who have Autism. *Journal of Behavioral Education*, *29*(2), 317–338. https://doi.org/10.1007/s10864-020-09376-4

*Shire, S. Y., Chang, Y.-C., Shih, W., Bracaglia, S., Kodjoe, M., & Kasari, C. (2017). Hybrid implementation model of community-partnered early intervention for toddlers with autism: A randomized trial. *Journal of Child Psychology and Psychiatry*, *58*(5), 612–622. https://doi.org/10.1111/jcpp.12672

*Shire, S. Y., Shih, W., Chang, Y.-C., Bracaglia, S., Kodjoe, M., & Kasari, C. (2019). Sustained Community Implementation of JASPER Intervention with Toddlers with Autism. *Journal of Autism and Developmental Disorders*, *49*(5), 1863–1875. https://doi.org/10.1007/s10803-018-03875-0

*Shire, S. Y., Shih, W., & Kasari, C. (2018). Brief Report: Caregiver Strategy Implementation-Advancing Spoken Communication in Children Who are Minimally Verbal. *Journal of Autism and Developmental Disorders*, *48*(4), 1228–1234. https://doi.org/10.1007/s10803-017-3454-0

Sidik, K., & Jonkman, J. N. (2005). Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(2), 367-384. https://doi.org/10.1111/j.1467-9876.2005.00489.x

Sidik, K., & Jonkman, J. N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in medicine*, *26*(9), 1964-1981. https://doi.org/10.1002/sim.2688

*Siller, M., Hutman, T., & Sigman, M. (2013). A parent-mediated intervention to increase responsive parental behaviors and child communication in children with ASD: A randomized clinical trial. *Journal of Autism and Developmental Disorders*, *43*(3), 540–555. https://doi.org/10.1007/s10803-012-1584-y

*Silva, L. M. T., Schalock, M., Gabrielsen, K. R., Budden, S. S., Buenrostro, M., & Horton, G. (2015). Early Intervention with a Parent-Delivered Massage Protocol Directed at Tactile Abnormalities Decreases Severity of Autism and Improves Child-to-Parent Interactions: A Replication Study. *Autism Research and Treatment*, *2015*, 904585. https://doi.org/10.1155/2015/904585

Smith, T., & Iadarola, S. (2015). Evidence base update for autism spectrum disorder. Journal of Clinical Child & Adolescent Psychology, 44(6), 897-922. https://doi.org/10.1080/15374416.2015.1077448

*Sokmun, S., Singh, S. J., & Vandort, S. (2018). The Impact of Hanen More Than Words Programme on Parents of Children with ASD in Malaysia. *Jurnal Sains Kesihatan Malaysia (Malaysian Journal of Health Sciences)*, *15*(2), Article 2. https://ejournal.ukm.my/jskm/article/view/14882

*Solomon, M., Iosif, A.-M., Reinhardt, V. P., Libero, L. E., Nordahl, C. W., Ozonoff, S., Rogers, S. J., & Amaral, D. G. (2018). What will my child's future hold? Phenotypes of intellectual development in 2-8-year-olds with autism spectrum disorder. *Autism Research: Official Journal of the International Society for Autism Research*, *11*(1), 121–132. https://doi.org/10.1002/aur.1884

*Solomon, R., Van Egeren, L. A., Mahoney, G., Quon Huber, M. S., & Zimmerman, P. (2014). PLAY Project Home Consultation Intervention Program for Young Children With Autism

Spectrum Disorders: A Randomized Controlled Trial. *Journal of Developmental and Behavioral Pediatrics*, *35*(8), 475–485. https://doi.org/10.1097/DBP.0000000000000096

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

*Srinivasan, S. M., Park, I. K., Neelly, L. B., & Bhat, A. N. (2015). A comparison of the effects of rhythm and robotic interventions on repetitive behaviors and affective states of children with Autism Spectrum Disorder (ASD). *Research in Autism Spectrum Disorders*, *18*, 51–63. https://doi.org/10.1016/j.rasd.2015.07.004

*Stadnick, N. A., Stahmer, A., & Brookman-Frazee, L. (2015). Preliminary Effectiveness of Project ImPACT: A Parent-Mediated Intervention for Children with Autism Spectrum Disorder Delivered in a Community Program. *Journal of Autism and Developmental Disorders*, *45*(7), 2092–2104. https://doi.org/10.1007/s10803-015-2376-y

Sterne, J. A., Egger, M., Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). Publication bias in meta-analysis.

Stone, W. L., & Hogan, K. L. (1993). A structured parent interview for identifying young children with autism. *Journal of Autism and Developmental Disorders*, *23*(4), 639-652. https://doi.org/10.1007/BF01046106

*Strain, P. S., & Bovey, E. H. (2011). Randomized, Controlled Trial of the LEAP Model of Early Intervention for Young Children With Autism Spectrum Disorders. *Topics in Early Childhood Special Education*, *31*(3), 133–154. https://doi.org/10.1177/0271121411408740

Study Quality Assessment Tools. (2019). Retrieved from https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools.

Sturm, A., Kuhfeld, M., Kasari, C., & McCracken, J. T. (2017). Development and validation of an item response theory-based social responsiveness scale short form. *Journal of Child Psychology and Psychiatry*. PsycINFO. https://doi.org/10.1111/jcpp.12731

*Suma, K., Adamson, L. B., Bakeman, R., Robins, D. L., & Abrams, D. N. (2016). After Early Autism Diagnosis: Changes in Intervention and Parent-Child Interaction. *Journal of Autism and Developmental Disorders*, *46*(8), 2720–2733. https://doi.org/10.1007/s10803-016-2808-3

*Sutera, S., Pandey, J., Esser, E. L., Rosenthal, M. A., Wilson, L. B., Barton, M., Green, J., Hodgson, S., Robins, D. L., Dumont-Mathieu, T., & Fein, D. (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *37*(1), 98–107. https://doi.org/10.1007/s10803-006-0340-6

*Swensen, L. D. (2007). Exploring the effects of maternal input on the language of children with autism. *Doctoral Dissertations*, 1–270.

*Szatmari, P., Georgiades, S., Duku, E., Bennett, T. A., Bryson, S., Fombonne, E., Mirenda, P., Roberts, W., Smith, I. M., Vaillancourt, T., Volden, J., Waddell, C., Zwaigenbaum, L., Elsabbagh, M., Thompson, A., & Pathways in ASD Study Team. (2015). Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA Psychiatry*, *72*(3), 276–283. https://doi.org/10.1001/jamapsychiatry.2014.2463

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: The neglected end of the spectrum. Autism research, 6(6), 468-478. https://doi.org/10.1002/aur.1329

*Taliou, A., Zintzaras, E., Lykouras, L., & Francis, K. (2013). An open-label pilot study of a formulation containing the anti-inflammatory flavonoid luteolin and its effects on behavior in children with autism spectrum disorders. *Clinical Therapeutics*, *35*(5), 592–602. https://doi.org/10.1016/j.clinthera.2013.04.006

Tavernor, L., Barron, E., Rodgers, J., & McConachie, H. (2013). Finding out what matters: validity of quality of life measurement in young people with ASD. *Child: care, health and development*, *39*(4), 592-601. https://doi.org/10.1111/j.1365-2214.2012.01377.x

*Thompson, G. A., McFerran, K. S., & Gold, C. (2014). Family-centred music therapy to promote social engagement in young children with severe autism spectrum disorder: A randomized controlled study. *Child: Care, Health and Development*, *40*(6), 840–852. https://doi.org/10.1111/cch.12121

*Thurm, A., Manwaring, S. S., Swineford, L., & Farmer, C. (2015). Longitudinal study of symptom severity and language in minimally verbal children with autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *56*(1), 97–104. https://doi.org/10.1111/jcpp.12285

*Tilmont Pittala, E., Saint-Georges-Chaumet, Y., Favrot, C., Tanet, A., Cohen, D., & Saint-Georges, C. (2018). Clinical outcomes of interactive, intensive and individual (3i) play therapy for children with ASD: A two-year follow-up study. *BMC Pediatrics*, *18*. https://doi.org/10.1186/s12887-018-1126-7

*Tonge, B., Brereton, A., Kiomall, M., Mackinnon, A., & Rinehart, N. J. (2014). A randomised group comparison controlled trial of “preschoolers with autism”: A parent education and skills training intervention for young children with autistic disorder. *Autism: The*

International Journal of Research and Practice, *18*(2), 166–177. https://doi.org/10.1177/1362361312458186

*Trembath, D., Westerveld, M. F., Teppala, S., Thirumanickam, A., Sulek, R., Rose, V., Tucker, M., Paynter, J., Hetzroni, O., Keen, D., & Vivanti, G. (2019). Profiles of vocalization change in children with autism receiving early intervention. *Autism Research*, *12*(5), 830–842. https://doi.org/10.1002/aur.2075

*Tsang, S. K. M., Shek, D. T. L., Lam, L. L., Tang, F. L. Y., & Cheung, P. M. P. (2007). Brief report: Application of the TEACCH program on Chinese pre-school children with autism--Does culture make a difference? *Journal of Autism and Developmental Disorders*, *37*(2), 390–396. https://doi.org/10.1007/s10803-006-0199-6

*Turner-Brown, L., Hume, K., Boyd, B. A., & Kainz, K. (2019). Preliminary Efficacy of Family Implemented TEACCH for Toddlers: Effects on Parents and Their Toddlers with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *49*(7), 2685–2698. https://doi.org/10.1007/s10803-016-2812-7

Ungerer, J. A., & Sigman, M. (1981). Symbolic play and language comprehension in autistic children. *Journal of the American Academy of Child Psychiatry*, *20*(2), 318-337. https://doi.org/10.1016/S0002-7138(09)60992-4

University of North Carolina (Chapel Hill). Department of Psychiatry. Division TEACCH, & Schopler, E. (1990). *Psychoeducational profile revised (PEP-R)*. Pro-ed.

*Valeri, G., Casula, L., Menghini, D., Amendola, F. A., Napoli, E., Pasqualetti, P., & Vicari, S. (2020). Cooperative parent-mediated therapy for Italian preschool children with autism spectrum disorder: A randomized controlled trial. *European Child & Adolescent Psychiatry*, *29*(7), 935–946. https://doi.org/10.1007/s00787-019-01395-5

Veenstra-VanderWeele, J., Christian, S. L., & Cook, Jr, E. H. (2004). Autism as a paradigmatic complex genetic disorder. Annu. Rev. Genomics Hum. Genet., 5, 379-405.https://doi.org/10.1146/annurev.genom.5.061903.180050

*Venker, C. E., Ray-Subramanian, C. E., Bolt, D. M., & Weismer, S. E. (2014). Trajectories of Autism Severity in Early Childhood. *Journal of Autism and Developmental Disorders*, *44*(3), 546–563. https://doi.org/10.1007/s10803-013-1903-y

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*(3), 261-293.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, **36**(3), 1–48. https://www.jstatsoft.org/v036/i03.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research synthesis methods*, *1*(2), 112-125.

*Vietze, P., & Lax, L. E. (2020). Early Intervention ABA for Toddlers with ASD: Effect of Age and Amount. *Current Psychology*, *39*(4), 1234–1244. https://doi.org/10.1007/s12144-018-9812-z

*Vinen, Z., Clark, M., Paynter, J., & Dissanayake, C. (2018). School Age Outcomes of Children with Autism Spectrum Disorder Who Received Community-Based Early Interventions. *Journal of Autism and Developmental Disorders*, *48*(5), 1673–1683. https://doi.org/10.1007/s10803-017-3414-8

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes. Clinical psychology review, 30(4), 387-399.

*Vivanti, G., Dissanayake, C., & Victorian ASELCC Team. (2016). Outcome for Children Receiving the Early Start Denver Model Before and After 48 Months. *Journal of Autism and Developmental Disorders*, *46*(7), 2441–2449. https://doi.org/10.1007/s10803-016-2777-6

*Vivanti, G., Paynter, J., Duncan, E., Fothergill, H., Dissanayake, C., Rogers, S. J., & the Victorian ASELCC Team. (2014). Effectiveness and Feasibility of the Early Start Denver Model Implemented in a Group-Based Community Childcare Setting. *Journal of Autism and Developmental Disorders*, *44*(12), 3140–3153. https://doi.org/10.1007/s10803-014-2168-9

*Waizbard-Bartov, E., Ferrer, E., Young, G. S., Heath, B., Rogers, S., Wu Nordahl, C., Solomon, M., & Amaral, D. G. (2021). Trajectories of Autism Symptom Severity Change During Early Childhood. *Journal of Autism and Developmental Disorders*, *51*(1), 227–242. https://doi.org/10.1007/s10803-020-04526-z

*Wang, J.-G., Cai, K.-L., Liu, Z.-M., Herold, F., Zou, L., Zhu, L.-N., Xiong, X., & Chen, A.-G. (2020). Effects of Mini-Basketball Training Program on Executive Functions and Core Symptoms among Preschool Children with Autism Spectrum Disorders. *Brain Sciences*, *10*(5). https://doi.org/10.3390/brainsci10050263

Wei, T., Chesnut, S. R., Barnard-Brak, L., & Richman, D. (2015). Psychometric Analysis of the Social Communication Questionnaire Using an Item-Response Theory Framework: Implications for the Use of the Lifetime and Current Forms. *Journal of Psychopathology and Behavioral Assessment*, *37*(3), 469–480. https://doi.org/10.1007/s10862-014-9468-4

*Welterlin, A., Turner-Brown, L. M., Harris, S., Mesibov, G., & Delmolino, L. (2012). The home TEACCHing program for toddlers with autism. *Journal of Autism and Developmental Disorders*, *42*(9), 1827–1835. https://doi.org/10.1007/s10803-011-1419-2

Wetherby, A. M., & Prizant, B. M. (2002). *Communication and symbolic behavior scales: Developmental profile*. Paul H Brookes Publishing Co.

*Whitehouse, A. J. O., Granich, J., Alvares, G., Busacca, M., Cooper, M. N., Dass, A., Duong, T., Harper, R., Marshall, W., Richdale, A., Rodwell, T., Trembath, D., Vellanki, P., Moore, D. W., & Anderson, A. (2017). A randomised controlled trial of an iPad-based application to complement early behavioural intervention in Autism Spectrum Disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *58*(9), 1042–1052. https://doi.org/10.1111/jcpp.12752

*Wild, D. L. (2012). The Impact of Canine Assistance for Children with Autism and the Family Unit. In *ProQuest LLC*. ProQuest LLC.

Williams, S. K., Scahill, L., Vitiello, B., Aman, M. G., Arnold, L. E., McDougle, C. J., ... & Sparrow, S. (2006). Risperidone and adaptive behavior in children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(4), 431-439. https://doi.org/10.1097/01.chi.0000196423.80717.32

Williams, J. H. (2008). Self–other relations in social development and autism: multiple roles for mirror neurons and other brain bases. Autism Research, 1(2), 73-90. https://doi.org/10.1002/aur.15

*Wong, C. S. (2013). A play and joint attention intervention for teachers of young children with autism: A randomized controlled pilot study. *Autism*, *17*(3), 340–357. https://doi.org/10.1177/1362361312474723

*Wong, V. C.-N., & Sun, J.-G. (2010). Randomized controlled trial of acupuncture versus sham acupuncture in autism spectrum disorder. *Journal of Alternative and Complementary Medicine (New York, N.Y.)*, *16*(5), 545–553. https://doi.org/10.1089/acm.2007.0768

*Woo, C. C., Donnelly, J. H., Steinberg-Epstein, R., & Leon, M. (2015). Environmental enrichment as a therapy for autism: A clinical trial replication and extension. *Behavioral Neuroscience*, *129*(4), 412–422. https://doi.org/10.1037/bne0000068

*Wood, S., Christian, M. P., & Sampson, A. (2018). Audit of outcomes following a community-based early intensive behaviour intervention program for children with autism in Australia. *Australian Journal of Psychology*, *70*(3), 217–224. https://doi.org/10.1111/ajpy.12193

*Xu, Y., Yang, J., Yao, J., Chen, J., Zhuang, X., Wang, W., Zhang, X., & Lee, G. T. (2018). A Pilot Study of a Culturally Adapted Early Intervention for Young Children With Autism Spectrum Disorders in China. *Journal of Early Intervention*, *40*(1), 52–68. https://doi.org/10.1177/1053815117748408

*Yoder, P., & Stone, W. L. (2006). A randomized comparison of the effect of two prelinguistic communication interventions on the acquisition of spoken communication in preschoolers with ASD. *Journal of Speech, Language, and Hearing Research: JSLHR*, *49*(4), 698–711. https://doi.org/10.1044/1092-4388(2006/051)

*Yoder, P., Watson, L. R., & Lambert, W. (2015). Value-added predictors of expressive and receptive language growth in initially nonverbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *45*(5), 1254–1270. https://doi.org/10.1007/s10803-014-2286-4

*Yun, S.-S., Choi, J., Park, S.-K., Bong, G.-Y., & Yoo, H. (2017). Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system.

Autism Research: Official Journal of the International Society for Autism Research, *10*(7), 1306–1323. https://doi.org/10.1002/aur.1778

*Zachor, D. A., Ben-Itzchak, E., Rabinovich, A.-L., & Lahat, E. (2007a). Change in Autism Core Symptoms with Intervention. *Research in Autism Spectrum Disorders*, *1*(4), 304–317. https://doi.org/10.1016/j.rasd.2006.12.001

*Zachor, D. A., Ben-Itzchak, E., Rabinovich, A.-L., & Lahat, E. (2007b). Change in autism core symptoms with intervention. *Research in Autism Spectrum Disorders*, *1*(4), 304–317. https://doi.org/10.1016/j.rasd.2006.12.001

*Zachor, D. A., Vardi, S., Baron-Eitan, S., Brodai-Meir, I., Ginossar, N., & Ben-Itzchak, E. (2017). The effectiveness of an outdoor adventure programme for young children with autism spectrum disorder: A controlled study. *Developmental Medicine and Child Neurology*, *59*(5), 550–556. https://doi.org/10.1111/dmcn.13337

*Zhao, M., & Chen, S. (2018). The Effects of Structured Physical Activity Program on Social Interaction and Communication for Children with Autism. *BioMed Research International*, *2018*, 1825046. https://doi.org/10.1155/2018/1825046

*Zhou, B., Xu, Q., Li, H., Zhang, Y., Wang, Y., Rogers, S. J., & Xu, X. (2018). Effects of Parent-Implemented Early Start Denver Model Intervention on Chinese Toddlers with Autism Spectrum Disorder: A Non-Randomized Controlled Trial. *Autism Research: Official Journal of the International Society for Autism Research*, *11*(4), 654–666. https://doi.org/10.1002/aur.1917

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). Preschool language scales–fifth edition (PLS-5). *Bloomington, MN: Pearson*.

*Zlomke, K. R., Jeter, K., & Murphy, J. (2017). Open-trial pilot of Parent-Child Interaction Therapy for children with Autism Spectrum Disorder. *Child & Family Behavior Therapy*, *39*(1), 1–18. https://doi.org/10.1080/07317107.2016.1267999

Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., Mailloux, Z., Roley, S. S., Wagner, S., Fein, D., Pierce, K., Buie, T., Davis, P. A., Newschaffer, C., Robins, D., Wetherby, A., Stone, W. L., Yirmiya, N., Estes, A., … Natowicz, M. R. (2015). Early intervention for children with autism spectrum disorder under 3 years of age: Recommendations for practice and research. *Pediatrics*, *136*(4Suppl 1), S60–S81. PsycINFO. https://doi.org/10.1542/peds.2014-3667E