Published in final edited form as:

J Child Psychol Psychiatry. 2020 January ; 61(1): 4–17. doi:10.1111/jcpp.13057.

Outcomes of 24- to 36-month-old children with ASD vary by ascertainment strategy: a systematic review and meta-analysis

Megan Micheletti^{1,2,3}, Courtney McCracken^{2,3}, John Constantino⁴, David Mandell⁵, Warren Jones^{1,2,3,6}, Ami Klin^{1,2,3,6}

¹Marcus Autism Center, Atlanta, GA

²Children's Healthcare of Atlanta, Atlanta, GA

³Department of Pediatrics, Emory University School of Medicine, Atlanta, GA

⁴Departments of Psychiatry and Pediatrics, and Intellectual and Developmental Disabilities Research Center, Washington University, St Louis, MO

⁵Center for Mental Health Policy and Services Research, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

⁶Emory Center for Translational Social Neuroscience, Atlanta, GA, USA

Abstract

Background: Despite widespread recommendations for early surveillance of risk for autism spectrum disorder (ASD), no research to date has shown that early surveillance leads to better clinical outcomes. Preliminary research has suggested that children with ASD ascertained via prospective follow-up have better outcomes than those ascertained via community referral. Because prospective studies include early surveillance, by comparing outcomes of children with ASD across ascertainment strategies, we may gain insight into the effects of early surveillance relative to its absence.

Methods: A systematic review was conducted to identify studies reporting outcomes of 24- to 36-month-olds with ASD ascertained via prospective follow-up, community referral, or universal screening. A meta-analysis using a random effects model was used to calculate overall effect size estimates for developmental level and symptom severity across ascertainment cohorts.

Results: 11 prospective, 10 community referral, and 8 universal screening studies were identified, reporting on 1,658 toddlers with ASD. We found no differences in outcomes between community referral and universal screening studies. Relative to both, prospective studies reported significantly higher developmental levels and lower symptom severities.

Conclusions: Outcomes of young children with ASD ascertained via prospective follow-up are better than those of children with ASD recruited via community referral or universal screening.

Correspondence: Ami Klin, Marcus Autism Center, 1920 Briarcliff Rd NE, Atlanta, GA 30329, USA; ami.klin@emory.edu.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of this article:

Conflict of interest statement: No conflicts declared.

The authors have declared that they have no competing or potential conflicts of interest.

Although we discuss why sampling bias is not likely the driving force behind these findings, we cannot rule out the possibility that sampling bias contributes to the observed differences; future studies should probe the effects of sociodemographic variables on clinical outcomes as a function of ascertainment strategy. This limitation notwithstanding, our results raise the possibility that prospective follow-up may confer a “surveillance effect” that contributes to improved developmental and diagnostic outcomes in children with ASD. Future research should test this hypothesis and determine the specific mechanism by which surveillance may improve outcomes.

Keywords

Autism spectrum disorder; toddlers; outcomes; sampling bias; surveillance; systematic review; meta-analysis.

Introduction

Early diagnosis and intervention optimize outcomes for children with autism spectrum disorder (ASD; Daniels, Halladay, Shih, Elder, & Dawson, 2014; Dawson, 2016; Dawson & Bernier, 2013; Dawson et al., 2010; Guthrie et al., 2016; Johnson, Myers, & American Academy of Pediatrics Council on Children With Disabilities, 2007; Reichow, Barton, Boyd, & Hume, 2012; Warren et al., 2011; Wetherby et al., 2014). The American Academy of Pediatrics and the US Centers for Disease Control and Prevention strongly recommend early surveillance of risk for ASD and related developmental disabilities (Daniel, Prue, Taylor, Thomas, & Scales, 2009; Johnson et al., 2007). In addition to sensitizing providers and families to typical child development, these guidelines are intended to prompt referral to diagnosis and early treatment when deviations from typical development are detected.

Despite these guidelines, no research to date has estimated the benefits of early surveillance on outcomes of young children with ASD. Two preliminary studies have suggested that children with ASD ascertained via prospective follow-up have better outcomes than children with ASD ascertained through community referral (Sacrey et al., 2017; Micheletti et al., under review). Children recruited prospectively – primarily infant siblings of children with ASD who are at high familial risk for the condition – are followed longitudinally before the emergence of symptoms. In contrast, children recruited via community referral are being evaluated for the first time, due to parent or provider concern, as part of participation in research. Prospective studies, in following and repeatedly assessing children’s development over time, incorporate early surveillance into their study design, whereas community referral studies do not. Thus, in comparing the outcomes of children with ASD ascertained via prospective follow-up vs. community referral, we may open a window into the effects of early surveillance relative to its absence.

The first study suggesting prospectively-followed children had better outcomes than community-referred children involved two large cohorts ($n = 86$ in each) of toddlers with ASD (Sacrey et al., 2017). The cohorts were matched on multiplex status (i.e., children in both groups had an older sibling with ASD), age of diagnosis, and ethnicity. Clinical comparisons at 36 months revealed lower symptomatology and higher adaptive skills in the prospective vs. community referral cohort. Differences in adaptive behavior were striking:

on the *Vineland Adaptive Behavior Scales* (Vineland; Sparrow, Balla, & Cicchetti, 1984), percentages of children in the prospective vs. community referral groups scoring within the adequate range were 75.0% vs. 20.5% in the communication domain, and 50.0% vs. 4.5% in the socialization domain.

The second study involved smaller prospective ($n = 25$) and community referral ($n = 22$) samples, and the prospective cohort of infant siblings had been assessed more frequently, with 9 data points prior to age 24 months. The groups were matched on multiplex status, age of diagnostic evaluation (between 24 – 36 months), and a range of potential confounds, including sex ratio, race/ethnicity, maternal education, pregnancy complications, birth complications, gestational age, and maternal age. Symptom levels were measured with the *Autism Diagnostic Observation Schedule* calibrated severity score (ADOS CSS), ranging from 1 – 10 (Lord, Luyster, Gotham, & Guthrie, 2012; Gotham, Pickles, & Lord, 2009). Developmental levels were measured with the *Mullen Scales of Early Learning* (Mullen, 1995). Mirroring results in the previous study, prospectively-followed children exhibited lower levels of ASD symptoms than community-referred children —CSS=6.0 (2.2) vs. 7.9 (1.9); higher levels of receptive and expressive developmental quotients —85.7 (25.7) vs. 44.8 (22.5), and 80.8 (20.0) vs. 51.1 (20.4); and higher nonverbal cognitive developmental quotients —100.7 (20.2) vs. 78.3 (22.0).

Unfortunately, we cannot dismantle these studies to investigate the mechanism by which prospectively-followed children have better outcomes. Given the dramatic differences in outcome, one might even surmise that these ascertainment approaches are sampling from opposite ends of the “true” distribution of ASD in the population at large. Differences in outcome could, therefore, be attributed to sampling bias, not the presence or absence of developmental surveillance.

To address this concern, we first provide an overview on the extent to which sampling biases may influence the reported outcomes of children with ASD. Next, we conduct a meta-analysis to investigate whether different ascertainment strategies lead to systematically different outcomes in children with ASD. If they do, and we find that prospectively-followed children have better outcomes than community-referred children across the literature, we may establish the foundation for future research investigating the mechanism by which this occurs. We also interpret our results in light of potential sampling biases. Finally, we integrate our findings with their implications for future research and public health action.

Potential Biasing Factors

It is a methodological truism that different sampling approaches can, and often do, introduce biases (Bornstein, Jager, & Putnick, 2013). As we are concerned with whether a bias may impact developmental and symptom level outcomes in toddlers with ASD, we focus our summary on the sampling biases most likely to impact these domains.

Clinical referral.—Possibly the most important factor differentiating prospective and community referral samples is that prospectively-followed children have diagnostic evaluations regardless of their symptom severity. In contrast, the ascertainment of children with ASD from the community depends on clinical referral, the timing of which is likely to

be influenced by severity of symptoms and community and family resources. In the general population, earlier diagnosis is associated with lower developmental levels, greater symptom severity, higher socioeconomic status, and greater parental concern about initial symptoms (Daniels & Mandell, 2014).

To examine whether community referral sampling systematically selects more disabled children, we can enlist the comparison of another sampling strategy: universal screening. In universal screening studies, cases are defined by a positive screen followed by diagnosis; thus, ascertainment is largely independent from referrals triggered by parent or provider concern. If we hypothesize that ascertainment via community referral is systematically selecting a more disabled distribution of children with ASD, we would expect children with ASD obtained via universal screening to have relatively more positive outcomes. The only study comparing outcomes across these ascertainment strategies found virtually no differences (Kleinman et al., 2008). Community referral ($N = 20$) and universal screening ($N = 117$) cohorts were matched on age (mean age 27.19 months (4.84) and 26.65 months (4.58), respectively), sex ratio (80% and 82%, respectively), diagnosis, and age at developmental evaluation. The study found virtually identical profiles across cohorts in developmental status (Mullen), adaptive behavior (Vineland), and ASD symptoms (ADOS; *Autism Diagnostic Interview – Revised*, ADI-R (Rutter, Le Couteur, & Lord, 2003); *Childhood Autism Rating Scale*, CARS (Chlebowski, Green, Barton, & Fein, 2010); and DSM-IV number of ASD symptoms). At least in this study, there did not appear to be any referral bias, raising the possibility that sampling via community referral is not systematically selecting a more impaired population of children with ASD. Additional studies are needed to further probe this potential confound: for example, within universal screening samples, one could ascertain whether children with and without a clinical referral prior to the screening procedure differ in terms of developmental or symptom levels.

Multiplex vs. simplex families.—Another potential source of bias is that most children participating in prospective studies are infant siblings, and thus come from multiplex families. Multiplex status is associated with known differences from simplex status in mechanisms of genetic transmission (Constantino, Zhang, Frazier, Abbacchi, & Law, 2010). Some studies report more than threefold rates of *de novo* mutations in simplex families compared with multiplex families (Marshall et al., 2008; Sebat et al., 2007), which in turn suggest that individuals from simplex families are more likely to develop ASD as a result of sporadic genetic and/or non-genetic causes. These kinds of biological differences may also affect clinical profiles of resulting samples, typically lowering average IQ (Constantino et al., 2010). Given the much larger representation of simplex families in cohorts of children with ASD ascertained via community referral compared to prospective follow-up, IQ distributions in community referral cohorts could be biased toward the lower ranges.

Findings to date, however, are not consistent with this conclusion. For example, in the largest study of simplex families (Lord et al., 2012), in which there was a vigorous effort to prioritize children more likely to have *de novo* copy number variations (the Simons Simplex Collection; 86.3% males and mean age around 8–9 years), the average verbal IQ of the sample was 79.3 (30.5) and nonverbal IQ was 86.1 (25.3). This average IQ is not lower than that reported for 8-year-olds in recent CDC autism surveillance publications (Christensen et

al., 2016). More directly, two studies compared clinical profiles of ASD individuals across simplex and multiplex families. The first study (Oerlemans, Hartman, Franke, Buitelaar, & Rommelse, 2016) matched simplex and multiplex children on sibship size, parental education, age, and sex. There were no differences on measures of verbal and performance IQ, or on measures of social cognition and executive function. The second study, by Sacrey and colleagues (Sacrey et al., 2017), is the same study that reported better outcomes for ASD children ascertained via prospective follow-up compared to community referral. The authors found no significant differences in ADOS or ADI-R scores, or in percentages of children scoring within the adequate range on the Vineland, across simplex vs. multiplex children.

In summary, although prospective and community referral cohorts differ, by definition, in percentage of children from simplex and multiplex families, this factor does not appear to affect children's developmental or symptom level profiles and, therefore, is likely not a strong biasing factor on the outcomes of young children with ASD by ascertainment.

Sex ratio.—Although the frequently reported male:female sex ratio in ASD is approximately 4:1, there is wide variability depending on ascertainment method (Halladay et al., 2015; Howe, Yatchmink, Viscidi, & Morrow, 2014), ranging from 1.96:1 in a large prospective study of infant siblings (Zwaigenbaum et al., 2012), thus multiplex families, to 6.9:1 in the Simons Simplex Collection (Frazier, Georgiades, Bishop, & Hardan, 2014). Among simplex and multiplex families in a national register (Constantino et al., 2010), sex ratio was 5.9:1 and 3.3:1, respectively; for reference, CDC population surveillance of 4-year-olds reports a 3.3:1 ratio (Christensen et al., 2016), and a universal screening study reports a 4.33:1 ratio (Reinhardt, Wetherby, Schatschneider, & Lord, 2015). Thus, in simplex and older samples, relative to multiplex and younger samples, the male:female ratio is larger. These data suggest that there may be a higher percentage of boys in community referral cohorts than prospective cohorts.

However, research to date suggests that this skewed sex ratio is unlikely to bias phenotypic differences in young children. In a large prospective study of infant siblings evaluated at 3 years of age, boys and girls with ASD had comparable developmental levels (Zwaigenbaum et al., 2012). And in a universal screening study of 24 – 36 month children, there were no significant differences on measures of cognitive and adaptive functioning, early social communication, or ASD symptoms between boys and girls (Reinhardt et al., 2015).

Race/ethnicity and socioeconomic status (SES).—In population-based studies of children with ASD, race/ethnicity and socioeconomic variables are typically associated with ASD children's outcomes, with children of non-Hispanic, White, well-educated mothers more likely to have more positive outcomes (Christensen et al., 2016; Fountain, Winter, & Bearman, 2012). Available data (Mandell et al., 2009) suggest that this association is mediated by age of diagnosis (later in minority/lower SES families) and access to early intervention services (reduced in minority/lower SES families), and resulting health disparities may be ameliorated via parental training and engagement (Daniels & Mandell, 2014; Donohue, Childs, Richards, & Robins, 2017; Herlihy et al., 2014).

The only published study to date comparing prospectively-followed vs. community-referred samples found no differences in race/ethnicity across groups and higher levels of parental education, social ability, and adaptive functioning in the prospective vs. community referral sample (Sacrey et al., 2017). As noted, the effects of SES on clinical outcome have been suggested to be mediated by age at diagnosis and access to early intervention (Mandell et al., 2009); however, children in the Sacrey et al. study, and in the present meta-analysis, were matched on age at diagnosis between 24 – 36 months. Thus, parental education, one measure of SES, has the most potential to bias outcomes in the following two ways.

The first is that family background may influence patterns of phenotypic expression in genetic syndromes (Finucane, Challman, Martin, & Ledbetter, 2016). In this line of research, a deleterious effect conferred by a genetic mutation would act upon family background, like IQ of parents, by lowering IQ of affected offspring. According to this rationale, there would be a strong correlation between high parental IQ, high SES/parental education, and high offspring IQ. However, this relationship is not straightforward. Genetic effects on intelligence are similar in high and low SES families, but genetic influences in low SES children are moderated by shared experiences, indicating an environment-environment interaction. In other words, shared experiences seem to suggest that low SES, or low parental education, is not inevitably tied to poorer outcomes (Hanscombe et al., 2012). That notwithstanding, to the best of our knowledge, this hypothesis has not been tested in ASD research, although this has been studied in regards to level of social disability (Constantino & Todd, 2005).

The second possibility is that families with higher SES/parental education may be more aware of and effective in attaining access to services. To date, these findings have been obtained from studies of older children (Daniels & Mandell, 2014; Mandell et al., 2010), and no studies have investigated access to services by ascertainment strategy for children with ASD under 36 months. One might suspect that families enrolled in prospective studies are made more aware of services, but there are currently no data comparing awareness and eventual utilization of services by ascertainment strategy.

In sum, sociodemographic factors, especially SES as indexed by parental education, are important variables to consider when examining developmental outcomes of children with ASD. However, given that participants in the present study are matched on age at diagnosis and there exists no data on access to services by ascertainment strategy – the leading mechanism by which race/ethnicity and SES have been suggested to act on outcomes – it remains unclear the extent to which sociodemographic factors systematically bias the phenotypic expression of ASD in young children.

Subtler forms of ASD, the broader autism phenotype, and optimal outcomes.

—Investigators may refer to “subtler” forms of ASD to describe children who exhibit less pronounced symptoms, which, in turn, are typically associated with higher developmental levels in young children. The term “broader autism phenotype” (BAP) refers to individuals who are biologically related to a person with ASD and who present with subthreshold ASD symptoms and associated developmental delays, without meeting full criteria for ASD (Ozonoff et al., 2011; Ozonoff et al., 2014). The term “optimal outcome” refers to children

with ASD who no longer meet diagnostic criteria for the condition, and who reach normal cognitive function (Eigsti et al., 2016; Fein et al., 2013; Sutter et al., 2007). These constructs may play a role in the observed phenotypic differences across ascertainment strategies.

First, it is possible that community referral samples are not capturing the subtle, subthreshold symptoms of ASD in their participants, because children with subtle symptoms are not referred for evaluation. Prospective samples, however, may too be excluding children with a more subtle symptom expression, specifically those who do not yet, or no longer, meet criteria for ASD in a 24 – 36 month diagnostic window. For example: a toddler who exhibits subthreshold symptoms (BAP) at outcome may appear more disabled one or two years later when social demands increase, thus later meeting criteria for ASD; or a toddler thought to be of concern early on but responded well to treatment (optimal outcome) may be characterized as a non-ASD child at outcome rather than a child with ASD who did very well. If so, in prospective studies there may too be an under-representation of children with subtler, improving, or age-dependent expressions of ASD (Ozonoff et al., 2018).

Second, given the multiple opportunities for developmental evaluation associated with longitudinal research, one could view prospective samples as more likely to capture subtle forms of ASD relative to community samples. However, in the vast majority of cases, diagnostic ascertainment in prospective studies is conducted blind from a child's risk status and/or cumulative observations. Thus, the diagnostic procedure in prospective and community referral samples is fairly comparable, and unlikely to bias clinical outcomes.

To date, no studies have attempted to probe these potential sources of sampling bias. In the first 3 years of life, ASD expression can be fluid and highly variable, and clinical diagnosis is a complicated, and, to some degree, a subjective process (Zwaigenbaum et al., 2015). Given the current framework of categorical boundaries defining ASD, we do not have any evidence on the extent, or direction, to which dynamic and evolving symptom expression may be biasing the outcomes of children with ASD across ascertainment strategies.

Objective

The goal of the following systematic review and meta-analysis is to examine the extent and direction to which ascertainment strategy may influence clinical outcomes of young children with ASD. We compare the developmental and symptom level outcomes of 24- to 36-month-old children with ASD ascertained across three different recruitment methods: prospective follow-up (PRO), community referral (COMM), and universal screening (UNI).

Methods

Systematic Review

This review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (Moher, Liberati, Tetzlaff, Altman, & PRISMA Group, 2009). A literature search was conducted using the electronic database PubMed for peer-reviewed, English-language journal articles published from 2001 up to and including May 15, 2018. Search terms were selected to capture a broad range of studies that could then be evaluated more extensively (Appendix S1 in the Supporting Information section for search terms). In

addition, manual searches of the reference lists of included studies, previous systematic reviews, and research group websites were completed.

Eligibility criteria were uniform across our three ascertainment cohorts. First, all studies had to include a minimum of five participants with a diagnosis of ASD, autism, or pervasive developmental disorder (PDD-NOS). The diagnostic procedure had to include both clinical judgment and the administration of a developmental evaluation. Next, participant recruitment strategy had to be clearly stated; the strategy itself was specific to each cohort. PRO participants could not be recruited based on developmental functioning, parental concern, community referral, or universal screening results. Participants had to be enrolled by 24 months of age and attend at least two research visits before receiving a diagnosis. COMM studies included participants that were referred by a parent or clinical provider to a hospital or clinic on the basis of developmental concern. Studies where referral was explicitly based on the results from a screening tool were not included. UNI participants were identified from a general population sample using screening tools intended to pick up infants and toddlers with ASD. Next, studies were limited to those whose participants had a rounded mean age between 22–38 months at the time of evaluation. This range was extended from 24–36 months to maximize eligible studies. Lastly, all studies had to include a measure of developmental functioning. No participant could be excluded on the basis of their developmental level. With regard to multiple publications on the same cohort of children, the article with the most relevant developmental measures or the largest sample size was included, in that order. As the first UNI study on autism meeting criteria for this meta-analysis was published in 2001, and there were no PRO studies on young children prior to this date, the search was limited to articles from 2001 to present. Two authors (MM and AK) independently searched the literature, reviewed and screened articles, and reached consensus on all included studies.

Data Extraction

Principal outcome measures were sample characteristics and developmental levels. Measures of ASD symptoms were of secondary interest, extracted where available. Journals were contacted if an article's supporting information was not available online. Two authors (MM and AK) extracted data and independently double-entered items. Intraobserver agreement was 99.5%. Any areas of discrepancy were resolved by reviewer discussion. The descriptive data and initial outcome measures extracted from included studies are available as supporting information (see Table S1).

Data Analysis

Analysis was performed using the Comprehensive Meta-Analysis Software v3 (Englewood, NJ). Effect size (ES) calculations were performed for each study/ascertainment cohort and consisted of means and standard deviations for continuous variables (e.g., developmental level) or counts and percentages for categorical variables (e.g., sex). For studies reporting more than one ASD cohort (i.e., males/females), ES estimates were obtained by pooling estimates over the different study subgroups. This resulted in each study contributing only one ES per outcome measure. A random effects model was used to pool effect sizes across studies within ascertainment cohort to get an overall ES estimate for each outcome of

interest by cohort. Resulting effects are presented as mean estimates with associated 95% confidence intervals (CI). Additional assumptions and calculation details can be found in Appendix S2.

We compared ES estimates across ascertainment cohorts by comparing the overall estimates and 95% CIs for each outcome of interest. Conservatively, non-overlapping CIs were indicative of a statistically significant difference between ascertainment cohorts. Additionally, the Q-test was used to formally determine if heterogeneity was present within and across ascertainment cohorts. A significant Q-test (p-value < 0.05) was indicative of differences in ES estimates within and across studies. The I^2 statistic was also calculated and used to describe the percent of variation due to heterogeneity rather than chance. A large I^2 value (>75%) indicates significant heterogeneity.

The robustness of results was assessed by conducting a sensitivity analysis to determine how sensitive the combined estimate was to any one study. This analysis was performed by repeatedly calculating the overall effect size after omitting one study per iteration, then comparing the new effect size with the overall effect size. To analyze the threat of publication bias to result validity, visual inspection of the funnel plots, Begg and Mazumdar's rank correlation test, and Egger's regression test were used. Begg and Mazumdar's rank correlation test examines whether there is a correlation between the study effect size and its standard error, which is driven primarily by sample size. A significant correlation indicates possible publication bias. Egger's test of the intercept also assesses whether or not the study effect size is associated with the standard error.

To assess the quality of each contributing study, we developed a checklist system based on established guidelines for conducting observational studies informed by the Strengthening the Reporting of Observational Studies in Epidemiology recommendations (Elm et al., 2007) (see Table S2). A score was created by summing the total number of quality criterion each study met, and scores could range from zero to 11 with higher scores indicative of higher quality publications.

Meta-regression was performed for the developmental level outcome to assess what study characteristics were associated with higher (or lower) developmental levels. The regression parameters included: average age at diagnosis, quality rating score, and ascertainment cohort. Each parameter was examined in a univariate and multivariable regression analysis with resulting estimates presented as slopes with standard errors. Statistical significance was assessed at the 0.05 level.

Results

Literature

Our initial search strategy yielded 1,130 articles (Figure 1). Once duplicates were removed, 1,041 articles were assessed for eligibility. After reading titles and abstracts, 754 articles were excluded. The remaining 287 articles were systematically checked against eligibility criteria in the following order: 5 or more participants with a diagnosis of ASD, autism, or PDD-NOS; clear ascertainment strategy; appropriate age range; administration of a

developmental measure; no truncation of developmental range; and not one of multiple publications on the same cohort. The first criterion an article did not meet was identified as its reason for exclusion. An additional 258 articles failed to meet eligibility criteria. The remaining 29 articles included $n = 11$ PRO, $n = 10$ COMM, and $n = 8$ UNI studies.

Description of studies by ascertainment strategy

The 11 PRO studies (Appendix S3.1) reported data on 463 participants. Estimated age was 32.31 months (95% CI 28.43 – 36.18) and male:female sex ratio was 2.67:1 (95% CI 1.91 – 3.73). Three studies (27%) reported race/ethnicity, and one study (9%) reported SES. Of those studies reporting race/ethnicity, samples averaged 67% White. Almost all participants ($n = 459$; 99%) were high-risk infant siblings. Age at study enrollment ranged from 3 to 24 months, and the number of study visits ranged from 2 to 17. Nine studies reported on the Mullen Early Learning Composite (ELC), a standard score of general development, resulting in an overall ELC estimate of 81.25 (95% CI 76.61 – 85.79). Mullen receptive and expressive language T scores ($M = 50$, $SD = 10$) were reported for seven studies, resulting in overall estimates of 37.93 (95% CI 33.29 – 42.56) and 40.39 (95% CI 37.71 – 43.08), respectively. Four studies reported ADOS Social Affect scores—a measure of social communication ranging from 0–20, with higher scores indicating higher severity and scores from 6–10 beginning to reflect concern (Hus, Gotham, & Lord, 2014)—resulting in an overall estimate of 12.20 (95% CI 10.74 – 13.65).

A total of 710 toddlers were identified across ten COMM studies (Appendix S3.2), with an estimated age of 29.88 months (95% CI 27.37 – 32.40) and 4.52:1 males to females (95% CI 3.72 – 5.50). Fewer females were identified in COMM than in PRO studies. Five studies (50%) reported race/ethnicity, three (30%) provided maternal education, and one (10%) gave a measure of SES. Of those studies reporting race/ethnicity, on average, samples were approximately 71% White; estimates of maternal education and SES could not be calculated because of differences in reporting. Four studies contributed to a developmental level estimate (three Mullen ELC, one Bayley Cognitive Composite Score) of 62.71 (95% CI 58.60 – 66.83), and three studies led to an ADOS Social Affect estimate of 16.14 (95% CI 15.55 – 16.73).

The eight UNI studies (Appendix S3.3) reported on 485 participants. Population samples ranged from approximately 800 to 20,000 individuals, with an average population N of 7,500 and an estimated ASD prevalence of 0.8%. Average age of the ASD sample was 27.51 months (95% CI 24.25 – 30.76) and sex ratio was 3.92:1 males to females (95% CI 3.02 – 5.08). Fewer females were identified in UNI than in PRO studies, and no significant differences were observed across UNI and COMM studies. The overall sex ratio estimate was significantly different across the three ascertainment strategies ($p = 0.028$), with the most females identified in PRO studies, followed by COMM and UNI. Three UNI studies (38%) reported race/ethnicity for their sample, resulting in a cohort average of 69% White. A Mullen ELC estimate of 60.91 (95% CI 56.12 – 65.71) and Mullen receptive and expressive language T score estimates of 22.80 (95% CI 21.00 – 24.60) and 25.81 (95% CI 25.04 – 26.58), respectively, were derived from six studies. Two studies contributed to the ADOS Social Affect estimate of 15.52 (95% CI 15.11 – 15.92).

Meta-analysis

Due to a lack of consistency in reporting measures across studies, we were able to analyze only developmental level and ADOS Social Affect scores in the meta-analysis, resulting in a final sample of $N = 20$ studies. Table 1 details the effect sizes derived for developmental level and ADOS Social Affect score by ascertainment cohort.

Nineteen studies contributed to the analysis of developmental level (9 PRO, 4 COMM, 6 UNI). Mixed effect regression analyses indicated that there was significant heterogeneity in the observed estimates of developmental level both within and across ascertainment groups, with significant Q statistics and within-cohort I^2 values ranging from 59% to 89% (Table 1). The overall I^2 value was 94%, indicating that heterogeneity was substantial and not due to chance alone. Developmental levels were significantly higher in PRO studies relative to both COMM and UNI studies (Figure 2). There were no differences in developmental levels between COMM and UNI studies.

Nine studies contributed to the analysis of ADOS Social Affect Total (4 PRO, 3 COMM, 2 UNI). Examination of heterogeneity demonstrated that there was less within-cohort heterogeneity in the UNI and COMM cohorts ($I^2 < 6\%$ for both); however, this is likely due to the small number of studies used in analysis (Table 1). When comparing estimates across all 3 cohorts, ADOS Social Affect estimates were not different across COMM and UNI studies, but were significantly lower in PRO studies (Figure 3).

Sensitivity analyses and publication bias

Results from the sensitivity analysis showed that the overall ES estimates for each cohort were not driven by any one study (Table 1). After removing one study at a time, the recomputed ES estimates were all within the 95% confidence intervals of the overall estimates. Additionally, in an examination of studies that required a derivation of developmental level from measure sub-domains, ES estimates when these studies were excluded were not significantly different from the overall estimates when included.

Examination of publication bias included funnel plots, Begg and Mazumdar Rank correlation test, and Egger's regression intercept test. Results from these figures and tests showed that there was evidence of a possible publication bias for both developmental level and ADOS Social Affect Total, with more favorable outcomes (e.g., higher developmental level or lower autism severity scores) being reported in studies with larger standard errors (see Figure S1, Figure S2, and Table S3). In addition, there was a statistically significant negative correlation between the study's quality rating and its reported developmental level ($r = -0.66$, 95% CI: $-0.86, -0.29$), demonstrating that higher quality studies reported worse outcomes.

Meta-regression

Too few studies reported race/ethnicity and SES to perform a meta-regression including these variables. However, we were able to explore the impact of study quality, age at diagnosis, and ascertainment cohort on developmental level (Table 2). Our univariate meta-regression model revealed that 64% of the variance observed in developmental levels across

cohorts could be attributed to ascertainment strategy. Quality rating and age at diagnosis were weakly associated with developmental level in the univariate analysis. Only ascertainment strategy remained a significant predictor of developmental level in our multivariable meta-regression model.

Discussion

Meta-analytic estimates of clinical presentation indicated significantly better functioning in PRO vs. COMM and UNI cohorts, with no differences between COMM and UNI cohorts. Young children with ASD recruited and diagnostically ascertained via prospective studies exhibited greater developmental skills and fewer ASD symptoms, with virtually no overlap in distributions with children recruited via community referral and universal screening. Our meta-regression results indicate that ascertainment strategy accounts for 64% of the variance in developmental levels in children in these studies. These findings are consistent with results of the two available case-control studies to date (Sacrey et al., 2017; Micheletti et al., under review) comparing clinical outcomes of prospectively-followed and community-referred children.

Our findings indicate that the referral aspect of case recruitment in community referral cohorts, an important potential bias in comparisons, does not systematically result in worse development or symptoms. Outcomes from community referral studies were not significantly different from those reported in universal screening studies, where children underwent diagnostic ascertainment without a clinical referral.

Study quality and age at diagnosis were weakly associated with developmental level in the univariate analysis. This was not surprising, given the restricted distribution in quality rating scores across our studies, and the fact that we constrained age of diagnosis (between 24 and 36 months) to standardize age at diagnosis for meta-analytic comparison. Our multivariate analysis revealed that ascertainment strategy was the best predictor of developmental level, above and beyond study quality or age at diagnosis.

Furthermore, prospective studies – generally reporting better clinical outcomes – tended to have higher standard errors than community referral or universal screening studies. This could reflect publication bias, such that studies with lower precision (more standard error) were biased to report better outcomes. More probable, however, this is due to differences in (a) sample size across ascertainment cohorts and (b) actual outcomes, reflecting true heterogeneity in the data (Sterne et al., 2011). Although we had a higher number of PRO studies, most had fewer participants than COMM or UNI studies, likely contributing to a higher standard error. It is also possible that PRO studies captured a more heterogeneous distribution of clinical outcomes, contributing to the observed differences in effect sizes across cohorts. Our random effects model helped to account for this heterogeneity; and even though PRO studies had a lower degree of precision, we were still able to identify significantly higher developmental levels in PRO vs. COMM and UNI studies.

The immediate implication of our results is that the clinical outcomes associated with prospectively-followed children are different from those of young children with ASD

ascertained via community referral or universal screening. Therefore, results from prospective, longitudinal infant sibling studies are not generalizable to children with ASD ascertained using other recruitment methods, or to the ASD population at large. The critical question that follows is whether prospective studies are biased towards sampling a “higher functioning” distribution of children with ASD or, alternatively, through some mechanism, prospective ascertainment may be contributing to better outcomes in children with ASD at 24 – 36 months.

As previously reviewed, although prospective and community referral cohorts may differ in terms of multiplex-simplex status and sex ratio, neither of these factors are likely to account for the large observed differences in ASD phenotypic expression across cohorts. In fact, the fluidity and complexity of the diagnostic process in prospective studies is likely to bias prospective cohorts towards exclusion of some of the highest-functioning, prospectively-followed children (Ozonoff et al., 2018). At-risk children who do very well by 36 months, or children who exhibit only subthreshold levels of ASD at that age, would not be included in a finalized ASD sample.

The question as to whether prospective and community referral cohorts systematically differ in demographics, particularly race/ethnicity and SES, cannot be addressed with the existing published data. Sociodemographic data are typically not reported in ASD research and clinical outcomes are often not examined as a function of sociodemographic variables. Although it remains possible that race or SES biased our results, differences in phenotypic profiles of children by race and SES are often attributed to later age at diagnosis and entry into early intervention (Mandell et al., 2009). Also, these health disparities have been identified in cohorts of children older than those in the current meta-analysis (Daniels & Mandell, 2014). We can estimate that because age at diagnosis did not explain a large portion of the variance in our results, differences in race/ethnicity and SES – which have been suggested to influence age at diagnosis – are not likely driving our findings. Furthermore, because diagnosis is usually the “ticket” to receiving early intervention services, it is unlikely that there are wide differences in early intervention receipt across cohorts at this point in time given comparable ages at diagnosis across the three ascertainment cohorts.

Another important difference between prospective and community referral cohorts is the potential benefits of the prospective follow-up on children’s clinical outcomes. In prospective studies, infants exposed to repeated assessments may learn from these novel experiences; parents may learn new ways to interact with their children by observing assessments conducted by trained professionals; or the very early detection of delays may result in enrollment in very early intervention, all of which could influence the child’s developmental trajectory (Szatmari et al., 2016). Some prospective studies may include parental and provider education, sensitization to normative developmental milestones, and new strategies to engage children (Guthrie et al., 2016), all of which could potentially attenuate cascading symptomatic effects while improving social and communicative engagement (Shultz, Klin, & Jones, 2018). Collectively, we term this beneficial effect on outcomes as the “surveillance effect”, and our results raise the possibility that developmental surveillance, and the aforementioned factors that accompany it, is the very mechanism by

which children with ASD enrolled in prospective studies have more positive outcomes. Additional research should investigate this hypothesis further, empirically exploring the effects of each one of the various discrete surveillance factors. The identification of the specific beneficial factor(s) could then be prioritized for intervention.

Limitations

This systematic review and meta-analysis rigorously synthesized the available evidence on the effects of sampling bias and ascertainment strategy on clinical outcomes in young children with ASD. While the findings identify significant differences in children's development and symptoms by ascertainment strategy, it is important to note some study limitations.

First, although we consider universal screening to be a population-based ascertainment strategy devoid of sampling biases, in reality, "universal" screening may not in fact be universal. Parents who are more concerned about their children may be more likely to participate in screening projects than those who are not. These concerns may reflect, proportionally, a child's level of developmental delay or disability. As a result, cohorts resulting from universal screening might tend to include more developmentally disabled children than a population-based sample. And yet, while this is possibly true, at present, we cannot estimate the effect of this bias. Data on families who refuse to participate in screening are typically not collected or reported. We know, however, that while there may be some self-selection for participation in universal screening studies, typically all parents presenting to a primary care practice are invited to participate in these studies. None of the universal screening reports we surveyed reported high refusal rates, and the very large sample sizes of these studies (typically in the 1,000's) may dilute some of the sampling bias introduced by parents refusing to participate in these studies. As more epidemiological surveillance data on toddlers begin to emerge, a greater effort should be invested in understanding factors facilitating or hindering participation in universal screening programs.

Similarly, although we contrasted universal screening and community referral sampling strategies, the clinical reality is that there may be overlap; community-referred children may have been previously screened and screened-positive children may be subsequently referred. Although our eligibility criteria for community referral studies prohibited inclusion of children screened then referred, it is possible for sampling overlap to occur in real-world clinical practice. For the purpose of this study, however, which was to compare the outcomes of children ascertained via different sampling strategies, we only included studies that had clearly distinct and non-overlapping recruitment approaches.

Second, prospective studies may have more opportunity to detect ASD symptoms compared to "one-shot" screeners or community referrals. Although this may be true, prospective studies typically keep clinicians blind to a child's risk status in an effort to maintain an unbiased diagnostic process. In other words, the developmental and diagnostic data reported by these studies typically do not reflect a *cumulative* set of observations; rather, diagnostic outcome is decided in evaluations similar to those conducted in universal screening or community referral studies (using the ADOS, the Mullen, and clinical judgment at a single point in time). That notwithstanding, it is possible that in some prospective cases, for whom

there is low clinician confidence in the differential diagnosis, the diagnostic procedure might include observations obtained in prior assessments. However, because clinicians' confidence in diagnosis is typically not provided in publications and, contrary to universal screening and community referral studies, prospective studies often assign a subthreshold quasi-diagnostic category of BAP, it is not known whether "higher functioning" cases of ASD are systematically included or excluded from the ASD sample in prospective studies. Although it is possible that prospective follow-up offers more opportunities to identify developmental concern, a more systematic discussion of this limitation generates alternatives that can only be elucidated empirically.

Third, it is possible that screening tools could fail to detect more subtle forms of ASD. This potential bias can be assessed by comparing developmental data across a sample of True Positives (screened positive and diagnosed positive) and False Negatives (screened negative and diagnosed positive). In a large study by Robins et al. (2014), a comparison of Mullen scores in True Positives vs. False Negatives revealed fairly comparable developmental levels: Visual Reception 29.64 (10.86) vs. 31.94 (12.99), Expressive Language 25.44 (8.63) vs. 25.44 (7.24), and Receptive Language 23.70 (8.97) vs. 26.17 (9.41), respectively. If the screening procedure were systematically excluding less severe cases of ASD, False Negatives should exhibit significantly higher developmental scores than True Positives. Unfortunately, to our knowledge, no other universal screening study has analyzed results of False Negatives. Therefore, although it is possible that screeners fail to detect more subtle forms of ASD, the data currently available do not substantiate this claim.

Fourth, many publications either did not include developmental and diagnostic data, or reported them in nonstandard formats. As a result, there were few published studies to include in our analyses. This was particularly true for our comparison of symptom severity, which was limited to the ADOS Social Affect domain. We were unable to compare severity across other domains, like Restricted and Repetitive Behaviors, or other measures, like the SRS (*Social Responsiveness Scale*; Constantino & Gruber, 2012), ADI-R, or CARS.

Fifth, to probe the possibility of beneficial surveillance effects, varying as a function of intensity of surveillance, it would have been helpful to enter number of study visits into a meta-regression on developmental outcome in prospectively-followed participants. However, most prospective studies provided only a range of study visits for their participants, with not all participants attending every visit; future studies should assess clinical outcomes by number of visits attended.

Lastly, we were unable to synthesize studies across demographic variables (sex, race/ethnicity, socioeconomic status, access to healthcare) or treatment variables (presence, timing, quality, quantity). These variables are typically not reported, or are not reported in standardized formats. Even in the small number of publications reporting them, they are by-and-large not analyzed to probe possible effects on clinical outcomes. While our review demonstrates that some potential confounds are unlikely to have strong effects in the current meta-analysis, the scarcity of research on the potential effects of sociodemographic variables on the early clinical expression of ASD is of great concern (Bornstein et al., 2013).

Research Recommendations

We recommend that investigators and journal editors adopt a more rigorous and standardized reporting format of clinical and sociodemographic variables in future ASD publications. Variables should be disaggregated as a function of the study's subgroups (e.g., ASD vs. non-ASD). Findings from all measures administered to participants should be reported, either in the main text, in supplementary information, or in an open-access repository. In regard to sociodemographic variables, studies should report on age of participants, sex, SES, parental education, and race/ethnicity following categories standardized by the National Institutes of Health. Studies should also offer a statement on the generalizability of their sample to the general population (Bornstein et al., 2013) and, when appropriate, an analysis of clinical outcomes relative to sociodemographic variables (American Academy of Pediatrics, 2000).

The effects of social determinants like race/ethnicity and SES on clinical outcomes will undoubtedly be complex, multifactorial, and multilevel (Newschaffer, 2017). Only with adequate reporting and consideration of sample characteristics, however, may we begin to disentangle the effects of sampling bias on ASD outcomes.

Lastly, additional research is required to directly test the “surveillance effect”, and the mechanisms by which surveillance may improve clinical outcomes of children with ASD. The impetus for this study was to systematize the various confounds associated with prospective surveillance studies and incentivize future research to address these confounds empirically. Future studies should invest in measuring potential mechanisms of change – like sensitizing parents to early normative developmental milestones (as advanced in public health campaigns), enhancing parent-child communicative interactions (as commonly adopted in parent-mediated early treatment), and early enrollment into early intervention – in longitudinal studies of ASD children ascertained via prospective surveillance vs. community referral and universal screening. As one example of a possible research design, following a transactional model of early development of social communication skills (Shultz, Klin, & Jones, 2018), it should be possible to assess the child's longitudinal trajectory as a function of change in daily parental interactions with the child resulting from parent-mediated treatment.

Conclusion

This is the first meta-analysis to investigate the effects of ascertainment strategy on observed outcomes in published studies of young children with ASD. Results suggest that the outcomes of young children with ASD who are diagnostically ascertained in prospective follow-up studies are significantly better than those of children ascertained in studies that recruit via community referral. Results also suggest that cohorts resulting from clinical referrals are not biased relative to the general population, as evidenced in comparisons with studies that recruit via population-based universal screening. Additional research is needed to probe sociodemographic effects on the clinical outcomes of young children with ASD. Our results raise the possibility that factors associated with early surveillance – such as parent engagement and education, and child monitoring and early assessment – may improve developmental and diagnostic outcomes. Future research should test this hypothesis and determine the specific mechanisms by which surveillance may improve outcomes. Such

mechanisms, if modifiable, could be promoted as ways to optimize the outcomes of all young children at risk for ASD (Daniels et al., 2014; Daniels & Mandell, 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by grants from the National Institute of Mental Health NIMH 2P50 MH100029–6. Additional support was provided by the Marcus Foundation, the J.B. Whitehead Foundation, the Children's Healthcare of Atlanta Foundation, and the Georgia Research Alliance. The authors would like to thank Aiden Ford and Sarah Markert for their helpful comments on the manuscript.

References

- American Academy of Pediatrics. (2000). Race/Ethnicity, Gender, Socioeconomic Status-Research Exploring Their Effects on Child Health: A Subject Review. *Pediatrics*, 105(6), 1349–1351. [PubMed: 10835080]
- Bornstein MH, Jager J, & Putnick DL (2013). Sampling in Developmental Science: Situations, Shortcomings, Solutions, and Standards. *Developmental Review: DR*, 33(4), 357–370. [PubMed: 25580049]
- Chlebowski C, Green JA, Barton ML, & Fein D (2010). Using the childhood autism rating scale to diagnose autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 40(7), 787–799. 10.1007/s10803-009-0926-x [PubMed: 20054630]
- Christensen DL, Baio J, Braun KVN, Bilder D, Charles J, Constantino JN, ... Centers for Disease Control and Prevention (CDC). (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR. Surveillance Summaries*, 65(3), 1–23. 10.15585/mmwr.ss6503a1
- Christensen DL, Bilder DA, Zahorodny W, Pettygrove S, Durkin MS, Fitzgerald RT, ... Yeargin-Allsopp M (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among 4-Year-Old Children in the Autism and Developmental Disabilities Monitoring Network. *Journal of Developmental & Behavioral Pediatrics*, 37(1), 1–8. 10.1097/DBP.0000000000000235 [PubMed: 26651088]
- Constantino JN, & Todd RD (2005). Intergenerational transmission of subthreshold autistic traits in the general population. *Biological Psychiatry*, 57(6), 655–660. 10.1016/j.biopsych.2004.12.014 [PubMed: 15780853]
- Constantino JN, Zhang Y, Frazier T, Abbacchi AM, & Law P (2010). Sibling recurrence and the genetic epidemiology of autism. *The American Journal of Psychiatry*, 167(11), 1349–1356. 10.1176/appi.ajp.2010.09101470 [PubMed: 20889652]
- Daniel KL, Prue C, Taylor MK, Thomas J, & Scales M (2009). 'Learn the signs. Act early': A campaign to help every child reach his or her full potential. *Public Health*, 123, e11–e16. 10.1016/j.puhe.2009.06.002
- Daniels AM, Halladay AK, Shih A, Elder LM, & Dawson G (2014). Approaches to Enhancing the Early Detection of Autism Spectrum Disorders: A Systematic Review of the Literature. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(2), 141–152. 10.1016/j.jaac.2013.11.002 [PubMed: 24472250]
- Daniels AM, & Mandell DS (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism*, 18(5), 583–597. 10.1177/1362361313480277 [PubMed: 23787411]
- Dawson G (2016). Why It's Important to Continue Universal Autism Screening While Research Fully Examines Its Impact. *JAMA Pediatrics*, 170(6), 527. 10.1001/jamapediatrics.2016.0163 [PubMed: 26882277]

- Dawson G, & Bernier R (2013). A quarter century of progress on the early detection and treatment of autism spectrum disorder. *Development and Psychopathology*, 25(4pt2), 1455–1472. 10.1017/S0954579413000710 [PubMed: 24342850]
- Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, ... Varley J (2010). Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics*, 125(1), e17–23. [PubMed: 19948568]
- Donohue MR, Childs AW, Richards M, & Robins DL (2017). Race influences parent report of concerns about symptoms of autism spectrum disorder. *Autism*, 136236131772203. 10.1177/1362361317722030
- Eigsti I-M, Stevens MC, Schultz RT, Barton M, Kelley E, Naigles L, ... Fein DA (2016). Language comprehension and brain function in individuals with an optimal outcome from autism. *NeuroImage. Clinical*, 10, 182–191. 10.1016/j.nicl.2015.11.014 [PubMed: 26862477]
- Elm E von, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, & STROBE Initiative. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*, 335(7624), 806–808. 10.1136/bmj.39335.541782.AD [PubMed: 17947786]
- Fein D, Barton M, Eigsti I-M, Kelley E, Naigles L, Schultz RT, ... Tyson K (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, 54(2), 195–205. 10.1111/jcpp.12037 [PubMed: 23320807]
- Finucane B, Challman TD, Martin CL, & Ledbetter DH (2016). Shift happens: family background influences clinical variability in genetic neurodevelopmental disorders. *Genetics in Medicine*, 18(4), 302–304. 10.1038/gim.2015.92 [PubMed: 26158229]
- Fountain C, Winter AS, & Bearman PS (2012). Six Developmental Trajectories Characterize Children With Autism. *PEDIATRICS*, 129(5), e1112–e1120. 10.1542/peds.2011-1601 [PubMed: 22473372]
- Frazier TW, Georgiades S, Bishop SL, & Hardan AY (2014). Behavioral and cognitive characteristics of females and males with autism in the Simons Simplex Collection. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53(3), 329–40.e1–3. 10.1016/j.jaac.2013.12.004 [PubMed: 24565360]
- Guthrie W, Wetherby AM, Woods JJ, Schatschneider C, Holland RD, Morgan L, & Lord C (2016). Earlier Intervention Is Better for Toddlers with ASD: Evidence from a Randomized Controlled Trial of the Early Social Interaction Project. Poster presented at the 2016 International Meeting for Autism Research, Baltimore, MD.
- Halladay AK, Bishop S, Constantino JN, Daniels AM, Koenig K, Palmer K, ... Szatmari P (2015). Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. *Molecular Autism*, 6(1), 36. 10.1186/s13229-015-0019-y [PubMed: 26075049]
- Hanscombe KB, Trzaskowski M, Haworth CMA, Davis OSP, Dale PS, & Plomin R (2012). Socioeconomic status (SES) and children's intelligence (IQ): in a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PloS One*, 7(2), e30320. 10.1371/journal.pone.0030320
- Herlihy LE, Brooks B, Dumont-Mathieu T, Barton ML, Fein D, Chen C-M, & Robins DL (2014). Standardized screening facilitates timely diagnosis of autism spectrum disorders in a diverse sample of low-risk toddlers. *Journal of Developmental and Behavioral Pediatrics: JDBP*, 35(2), 85–92. 10.1097/DBP.0000000000000014 [PubMed: 24509053]
- Howe YJ, Yatchmink Y, Viscidi EW, & Morrow EM (2014). Ascertainment and Gender in Autism Spectrum Disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(6), 698–700. 10.1016/j.jaac.2014.04.003 [PubMed: 24839890]
- Hus V, Gotham K, & Lord C (2014). Standardizing ADOS domain scores: separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*, 44(10), 2400–2412. 10.1007/s10803-012-1719-1 [PubMed: 23143131]
- Johnson CP, Myers SM, & American Academy of Pediatrics Council on Children With Disabilities. (2007). Identification and Evaluation of Children With Autism Spectrum Disorders. *PEDIATRICS*, 120(5), 1183–1215. 10.1542/peds.2007-2361 [PubMed: 17967920]

- Kleinman JM, Robins DL, Ventola PE, Pandey J, Boorstein HC, Esser EL, ... Fein D (2008). The modified checklist for autism in toddlers: a follow-up study investigating the early detection of autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38(5), 827–839. 10.1007/s10803-007-0450-9 [PubMed: 17882539]
- Lord C, Luyster RJ, Gotham K, & Guthrie W (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2): Toddler Module. (Second Edi)*. Torrance, CA: Western Psychological Services.
- Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, ... Risi S (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*, 69(3), 306–313. 10.1001/archgenpsychiatry.2011.148 [PubMed: 22065253]
- Mandell DS, Morales KH, Xie M, Lawer LJ, Stahmer AC, & Marcus SC (2010). Age of Diagnosis Among Medicaid-Enrolled Children With Autism, 2001–2004. *Psychiatric Services*, 61(8), 822–829. 10.1176/ps.2010.61.8.822 [PubMed: 20675842]
- Mandell DS, Wiggins LD, Carpenter LA, Daniels J, DiGuseppi C, Durkin MS, ... Kirby RS (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health*, 99(3), 493–498. 10.2105/AJPH.2007.131243 [PubMed: 19106426]
- Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, ... Scherer SW (2008). Structural Variation of Chromosomes in Autism Spectrum Disorder. *The American Journal of Human Genetics*, 82(2), 477–488. 10.1016/j.ajhg.2007.12.009 [PubMed: 18252227]
- Micheletti M, Shultz S, Klin A, Jones W, Hoffenberg S, Saulnier C, & Klaiman C (2018). The Effects of Prospectively Following High-Risk Infant Siblings with ASD: Observation Positively Alters Outcome.
- Moher D, Liberati A, Tetzlaff J, Altman DG, & Group TP (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, 6(7), e1000097. 10.1371/journal.pmed.1000097
- Mullen E (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Services, Inc.
- Newschaffer CJ (2017). Trends in Autism Spectrum Disorders: The Interaction of Time, Group-Level Socioeconomic Status, and Individual-Level Race/Ethnicity. *American Journal of Public Health*, 107(11), 1698–1699. 10.2105/AJPH.2017.304085 [PubMed: 29019763]
- Oerlemans AM, Hartman CA, Franke B, Buitelaar JK, & Rommelse NNJ (2016). Does the cognitive architecture of simplex and multiplex ASD families differ? *Journal of Autism and Developmental Disorders*, 46(2), 489–501. 10.1007/s10803-015-2572-9 [PubMed: 26340958]
- Ozonoff S, Young GS, Belding A, Hill M, Hill A, Hutman T, ... Iosif A-M (2014). The Broader Autism Phenotype in Infancy: When Does It Emerge? *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(4), 398–407.e2. 10.1016/j.jaac.2013.12.020 [PubMed: 24655649]
- Ozonoff S, Young GS, Brian J, Charman T, Shephard E, Solish A, & Zwaigenbaum L (2018). Diagnosis of Autism Spectrum Disorder After Age 5 in Children Evaluated Longitudinally Since Infancy. *Journal of the American Academy of Child & Adolescent Psychiatry*, 57(11), 849–857.e2. 10.1016/j.jaac.2018.06.022 [PubMed: 30392626]
- Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, ... Stone WL (2011). Recurrence risk for autism spectrum disorders: a baby siblings research consortium study. *Pediatrics*, 128(3), e1–e8. [PubMed: 21690117]
- Reichow B, Barton EE, Boyd BA, & Hume K (2012). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews*, (10). 10.1002/14651858.CD009260.pub2
- Reinhardt VP, Wetherby AM, Schatschneider C, & Lord C (2015). Examination of sex differences in a large sample of young children with autism spectrum disorder and typical development. *Journal of Autism and Developmental Disorders*, 45(3), 697–706. 10.1007/s10803-014-2223-6 [PubMed: 25189824]
- Rutter M, Le Couteur A, & Lord C (2003). *ADI-R: Autism Diagnostic Interview - Revised*. Los Angeles: Western Psychological Services.
- Sacrey L-AR, Zwaigenbaum L, Szatmari P, Bryson S, Georgiades S, Brian J, ... Elsabbagh M (2017). Brief Report: Characteristics of preschool children with ASD vary by ascertainment. *Journal of*

- Autism and Developmental Disorders, 47(5), 1542–1550. 10.1007/s10803-017-3062-z [PubMed: 28224344]
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, ... Wigler M (2007). Strong Association of De Novo Copy Number Mutations with Autism. *Science*, 316(5823), 445–449. 10.1126/science.1138659 [PubMed: 17363630]
- Shultz S, Klin A, & Jones W (2018). Neonatal Transitions in Social Behavior and Their Implications for Autism. *Trends in Cognitive Sciences*, 22(5), 452–469. 10.1016/j.tics.2018.02.012 [PubMed: 29609895]
- Sparrow SS, Balla D, & Cicchetti DV (1984). *Vineland Adaptive Behavior Scales*. (Second edi). Circle Pines, MN: American Guidance Services, Inc.
- Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, ... Higgins JPT (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343(7818), 1–8. 10.1136/bmj.d4002
- Sutera S, Pandey J, Esser EL, Rosenthal MA, Wilson LB, Barton M, ... Fein D (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37(1), 98–107. 10.1007/s10803-006-0340-6 [PubMed: 17206522]
- Szatmari P, Chawarska K, Dawson G, Georgiades S, Landa R, Lord C, ... Halladay A (2016). Prospective Longitudinal Studies of Infant Siblings of Children With Autism: Lessons Learned and Future Directions. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(3), 179–187. 10.1016/j.jaac.2015.12.014 [PubMed: 26903251]
- Warren Z, McPheeters ML, Sathe N, Foss-Feig JH, Glasser A, & Veenstra-Vanderweele J (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics*, 127(5), e1303–11. 10.1542/peds.2011-0426 [PubMed: 21464190]
- Wetherby AM, Guthrie W, Woods J, Schatschneider C, Holland RD, Morgan L, & Lord C (2014). Parent-implemented social intervention for toddlers with autism: an RCT. *Pediatrics*, 134(6), 1084–1093. 10.1542/peds.2014-0757 [PubMed: 25367544]
- Zwaigenbaum L, Bauman ML, Choueiri R, Fein D, Kasari C, Pierce K, ... Wetherby A (2015). Early Identification and Interventions for Autism Spectrum Disorder: Executive Summary. *PEDIATRICS*, 136(Supplement), S1–S9. 10.1542/peds.2014-3667B [PubMed: 26430167]
- Zwaigenbaum L, Bryson SE, Szatmari P, Brian J, Smith IM, Roberts W, ... Roncadin C (2012). Sex Differences in Children with Autism Spectrum Disorder Identified Within a High-Risk Infant Cohort. *Journal of Autism and Developmental Disorders*, 42(12), 2585–2596. 10.1007/s10803-012-1515-y [PubMed: 22453928]

Key points

- No research to date has shown that early surveillance leads to better outcomes in children with ASD
- This meta-analysis compares the clinical outcomes of children with ASD ascertained via 3 different recruitment strategies to determine if outcomes systematically vary in the presence or absence of surveillance
- Results suggest that prospectively-followed children—that necessarily receive surveillance—have significantly higher developmental levels and fewer symptoms than children ascertained via community referral or universal screening, who show no differences
- Improved reporting and consideration of sociodemographic variables are major challenges to be overcome in order to investigate the role of surveillance in the phenotypic expression of ASD
- Future research should identify the specific mechanisms by which prospectively-followed children have better clinical outcomes

Figure 1.
PRISMA flow diagram depicting study selection

Figure 2.

Forest plot of study-specific measures of developmental level. Mean refers to the study-specific mean developmental level, where developmental level is a standard score with a mean of 100 and standard deviation of 15. Circles represent the study-specific mean estimate with bar spanning the 95% confidence interval. Size of the dot represents the relative weight each study contributed to the overall cohort effect size. Squares represent the cohort-specific estimate from a random effects model. Ascertainment cohorts are prospective follow-up (PRO), community referral (COMM), and universal screening (UNI).

Figure 3.

Forest plot of study-specific measures of ASD severity measured by the ADOS Social Affect Total Score. Mean refers to the study-specific mean of ADOS Social Affect total scores, where a higher mean reflects greater social impairment. Circles are the study-specific mean estimate with bar spanning the 95% confidence interval. Size of the dot represents the relative weight each study contributed to the overall cohort effect size. Squares represent the cohort-specific estimate from a random effects model. Ascertainment cohorts are prospective follow-up (PRO), community referral (COMM), and universal screening (UNI).

Table 1.

Summary of measures of mixed effect analysis

Measure	Cohort ^a	Number of Studies	Estimate ^b	95% CI	Heterogeneity ^c (p-value)	I ² ^d	Sensitivity Analysis (range of ES)
Developmental Level (All)	PRO	9	81.25	(76.71, 85.79)	19.70 (0.01)	59.38	78.80, 82.58
	COMM	4	62.71	(58.60, 66.83)	28.46 (<0.01)	89.46	60.97, 65.09
	UNI	6	60.91	(56.12, 65.71)	42.61 (<0.01)	88.27	59.81, 62.58
	Overall	19	68.14	(65.57, 70.72)	47.44 (<0.01)	94.29	--
Developmental Level (No Estimates)	PRO	6	78.83	(73.73, 83.93)	14.02 (0.02)	64.33	77.48, 80.37
	COMM	3	61.76	(56.71, 66.81)	22.35 (<0.01)	91.05	59.48, 64.42
	UNI	2	64.47	(59.27, 69.67)	0.38 (0.54)	0.00	63.00, 66.29
	Overall	11	68.36	(65.40, 71.31)	24.88 (<0.01)	94.11	--
ADOS Social Affect Total ^e	PRO	4	12.20	(10.74, 13.65)	7.42 (0.06)	59.59	11.70, 12.77
	COMM	3	16.14	(15.55, 16.73)	2.13 (0.35)	5.90	15.89, 16.59
	UNI	2	15.52	(15.11, 15.92)	0.17 (0.68)	0.00	15.30, 15.50
	Overall	9	15.54	(15.22, 15.87)	24.25 (<0.01)	86.72	--

^aProspective follow-up (PRO), community referral (COMM), and universal screening (UNI) cohorts.

^bThe cohort-specific pooled estimate of the outcome measure obtained from the random effects model. Developmental level estimates are interpreted as mean standard scores with a mean of 100 and standard deviation of 15.

^cCochran's Q test assesses whether there is significant heterogeneity across estimates and justifies use of a random effects model. Tests whether the studies (or overall cohorts) have identical effects.

^dRepresents the percent of variation due to heterogeneity rather than chance.

^eStudies reporting only ADOS Communication Total and Social Total were added together to get a Social Affect Total Score. The standard deviation was estimated by taking the square root of the added variances. This calculation assumes independence between the variances and may actually be underestimating variance.

Table 2.

Summary meta-regression for developmental level at outcome

Covariate	N	Slope (SE)	P-value	R ²
Univariate				
Quality Rating	19	−11.1 (3.3)	<0.001	0.05
Age at Diagnosis	19	1.04 (0.37)	0.004	0.03
Cohort	19			0.64
<i>PRO</i>		18.6 (3.31)	<0.001	
<i>COMM</i>		reference		
<i>UNI</i>		−1.87 (3.28)	0.568	
Multivariate	19			0.58
Quality Rating		−3.06 (3.18)	0.337	
Age at Diagnosis		−0.11 (0.36)	0.761	
Cohort				
<i>PRO</i>		17.88 (3.95)	<0.001	
<i>COMM</i>		reference		
<i>UNI</i>		−1.34 (3.78)	0.717	