Onset Patterns in Autism: Variation Across Informants, Methods, and Timing

Drs. Sally Ozonoff, PhD,
Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California – Davis

Devon Gangi, PhD,
Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California – Davis

Elise P. Hanzel, PhD,
Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California – Davis

Ms. Alesha Hill, BS,
Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California – Davis

Monique M. Hill, MA,
Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California – Davis

Meghan Miller, PhD,
Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California – Davis

Dr. A.J. Schwichtenberg, PhD,
Department of Human Development and Family Studies, Purdue University

Dr. Mary Beth Steinfeld, MD,
Department of Pediatrics, University of California – Davis

Chandni Parikh, PhD, and
Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California – Davis

Dr. Ana-Maria Iosif, PhD
Department of Public Health Sciences, University of California – Davis

Abstract

Objective—While previous studies suggested that regressive forms of onset were not common in autism spectrum disorder (ASD), more recent investigations suggest that the rates are quite high

Corresponding Author: Sally Ozonoff PhD, 2825 50th Street, Sacramento CA 95817; 916-703-0259 (phone), sozonoff@ucdavis.edu.
The authors have no conflicts of interest to declare.

and may be under-reported using certain methods. The current study undertook a systematic investigation of how rates of regression differed by measurement method.

Method—Infants with (*n*=147) and without a family history of ASD (*n*=83) were seen prospectively for up to 7 visits in the first three years of life. Reports of symptom onset were collected using four measures that systematically varied the informant (examiner v. parent), the decision type (categorical [regression absent or present] v. dimensional [frequency of social behaviors]), and the timing of the assessment (retrospective v. prospective). Latent class growth models were used to classify individual trajectories to see whether regressive onset patterns were infrequent or widespread within the ASD group.

Results—A majority of the sample was classified as having a regressive onset using either examiner (88%) or parent (69%) prospective dimensional ratings. Rates of regression were much lower using retrospective or categorical measures (from 29 to 47%). Agreement among different measurement methods was low.

Conclusions—Declining trajectories of development, consistent with a regressive onset pattern, are common in children with ASD and may be more the rule than the exception. The accuracy of widely used methods of measuring onset is questionable and the present findings argue against their widespread use.

Keywords

Regression; Early Signs; Infants

Introduction

The onset of behavioral signs of autism spectrum disorder (ASD) is usually conceptualized as occurring in one of two ways: an early onset pattern, in which children demonstrate social-communication delays early in life, and a regressive pattern, in which children develop typically for some period and then experience a substantial decline in or loss of previously developed skills. While it was long believed that the majority of children with ASD demonstrated an early onset pattern, more recent studies suggest that regressive onset occurs more frequently than previously recognized (Brignell et al., 2017; Hansen et al., 2008; Kern, Geier & Geier, 2015; Shumway et al., 2011; Thurm et al., 2014; for a review, see meta-analysis by Barger, Campbell & McDonough, 2013). Studies occasionally also identify a third onset pattern, that of developmental stagnation or plateau (Shumway et al., 2011), that is characterized by intact early skills that fail to progress or transform into more advanced developmental achievements. This onset pattern is distinct from regression, in that the child does not lose acquired skills, but instead fails to make expected gains.

The most common procedure for collecting information about the timing of early symptoms is parent report. A number of factors can influence report validity, however, including awareness of the child's eventual diagnosis and knowledge of developmental milestones. It has long been documented that retrospective reports are subject to problems of memory and interpretation (Finney, 1981; Henry et al., 1994; Pickles et al., 1996), including in studies of autism (Andrews et al., 2002). Multiple studies have documented the ways in which recall problems and other biases can influence parent report. For example, several studies have

demonstrated changes in recall over time, with parents less likely to report regression and more likely to report early delays as their children grow older (Hus et al., 2011; Lord et al., 2004; Ozonoff et al., in press). In a study that compared classification of onset based on objective coding of family movies to onset type as recalled by parents (Ozonoff et al., 2011), less than half of children whose home video displayed clear evidence of a major decline in social-communication behavior were reported to have had a regression by parents. Similarly, only 40% of participants with clear evidence of early delays in social-communication and little evidence of skill decline on video were reported by parents to show an early onset pattern.

Questions about when and how behavioral signs of autism emerge may be better answered through prospective investigations, which do not rely on recall. A previous investigation compared onset type based on behavioral coding of videotaped prospective standardized assessments to onset type as recalled by parents (Ozonoff et al., 2010). By parental recall, only 17% of the children were classified as having regressive onset, but using prospective observational data, 86% of the same sample showed decreasing rates of eye contact, social smiles, and vocalizations over time. Decelerations in development, in which key social-communicative behaviors decrease in frequency over time, are consistent with a regressive onset. Multiple independent research teams, using a variety of different prospective methods, have replicated the finding of largely intact early development, followed by developmental declines and later onset of symptoms (for a review, see Jones et al., 2014). For example, Landa et al. (2013) found no differences on a variety of prospectively collected measures at 6 months between infants later diagnosed with ASD and non-ASD participants. Approximately half of the children with ASD demonstrated differences from the non-ASD cases at 14 months but the other half did not diverge from typical infants until 24 months. A similar pattern of no differences from non-ASD infants at 7 months, followed by divergence from typical development in the second year of life, was reported by Gammer et al. (2015). Jones and Klin (2013), using eye tracking, demonstrated that mean levels of looking at eyes did not differ between infants later diagnosed with ASD and infants with typical development at 6 months or earlier, but a negative slope, indicating a declining trajectory of gaze to eyes was evident between 2 and 24 months.

Collectively, the results of these prospective studies indicate that many children with ASD demonstrate declining trajectories of development and suggest that the rate of regression is underreported when relying upon retrospective recall. The previous literature is challenging to interpret, however, because earlier studies confounded assessment methods and informants with assessment timing: all retrospective measures used parent report, while all prospective measures used expert clinical opinion. In addition, retrospective measures relied on dichotomous categorical distinctions, asking parents if their child lost skills or not, whereas prospective measures were rated by examiners on a dimensional scale, quantifying the frequency of behaviors along a continuum. In the current study, we disentangled these variables by systematically manipulating the informant (examiner v. parent), the decision type (categorical [regression absent or present] v. dimensional [frequency of key behaviors]), and the timing of the assessment (retrospective v. prospective). Additionally, we used analytic methods (i.e., latent class growth models) that permitted classification of individual trajectories to see whether the pattern of decelerating development reported in previous

studies (Gammer et al., 2015; Jones & Klin, 2013; Landa et al., 2013; Ozonoff et al., 2010) is widespread across individuals or if the group results are driven by a few dramatic outliers.

Method

Participants

Two groups of infant siblings participated, one with and one without a family history of ASD. The sole inclusion criterion for the High-Risk (HR) group was status as a younger sibling of a child with ASD. Diagnosis of the affected older sibling was confirmed by meeting ASD criteria on both the Autism Diagnostic Observation Schedule (ADOS-2; Lord et al., 2012) and the Social Communication Questionnaire (SCQ; Berument et al., 1999). Exclusion criteria for the HR group included birth before 32 weeks of gestation and a known genetic disorder (e.g., Fragile X syndrome) in the older affected sibling. The primary inclusion criterion for the Low-Risk (LR) group was status as a younger sibling of a child (or children) with typical development, confirmed by an intake screening questionnaire and scores below the ASD range on the SCQ. Exclusion criteria for the LR group were birth before 36 weeks of gestation, developmental, learning, or medical conditions in any older sibling, and ASD in first-, second-, or third-degree relatives. The study was approved by the UC Davis IRB and parents signed a consent form prior to participation.

Participants were recruited at a mean age of 3.2 months ($SD = 4.7$), with 84% enrolled by 6 months and 100% by 9 months. Data were collected at up to seven ages (6, 9, 12, 15, 18, 24, and 36 months), with 36% of the participants seen at all seven ages, 37% at six ages, 17% at five ages, 8% at four ages, and 1% at three ages. At the 36-month visit, participants were classified into one of two outcome groups. The ASD group ($n = 32$; 30 HR, 2 LR; $n = 11$ females) met *DSM-5* criteria for autism spectrum disorder and obtained a score over the ASD cutoff on the ADOS. All other participants were classified as Non-ASD and further stratified by risk status (HR-Non-ASD, $n = 117$, 54 females; LR-Non-ASD, $n = 81$, 32 females).

Measures

Four instruments were used to measure onset patterns, varying the informant (examiner v. parent), the type of discrimination made (categorical v. dimensional), and the timing of the assessment (retrospective v. prospective). See Table 1.

Session Summary Report (Ozonoff et al., 2010) – examiner-rated prospective dimensional measure—At the end of each visit, examiners rated the frequency of eye contact, shared affect, and overall social engagement (number of initiations and responses) on a 5-point scale (1 = none, 2 = rare, 3 = occasional, 4 = frequent, 5 = very frequent). These scores were summed to create a composite that could range from 3 to 15.

Early Development Questionnaire (EDQ), Part 1 (Ozonoff, Williams & Landa, 2005) – parent-rated prospective dimensional measure—Parents completed this instrument prior to each visit. Part 1 of the EDQ consists of 45 questions about the child's current functioning in social, communication, and repetitive behavior domains. Each item is

rated on a 0 – 3 frequency scale (0=behavior never occurs, 3=behavior often occurs). Three items, comparable in content to the examiner ratings, were summed: item 1 ("my child looks at me during social interactions"), item 4 ("my child smiles back at me when I smile at him/her"), and item 13 ("when I call my child's name, he/she looks at me right away"). In addition to being parallel to the behaviors rated by examiners, these items were selected because they represent early-appearing behaviors that are relevant and developmentally appropriate across all ages of the study (6–36 months). In contrast to other EDQ items that measure later-developing skills (e.g., joint attention, language), the items selected for the composite measure behaviors present in the first year of life (Inada, Kamio & Koyama, 2010). Since this study was focused on regression, rather than developmental plateau or stagnation, it was critical that the behaviors measured have the potential to demonstrate decreases over time as ASD signs emerge. This dimensional variable, quantifying parent report of the frequency of key early social behaviors, had a potential range of 0 – 9.

Early Development Questionnaire, Part 2 – parent-rated prospective categorical measure—The EDQ also yields a categorical variable, in which the parent is asked, "Since your last visit, has your child shown significant decreases in XX," followed by a list of 10 social-communication behaviors, including speech, gestures, social interest, shared affect, eye contact, etc. The variable is scored dichotomously, with a "yes" response to a significant decrease in any of the ten listed behaviors coded as 1 and "no" responses to all ten behaviors coded as 0.

Autism Diagnostic Interview – Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) – parent-rated retrospective categorical measure—The ADI-R is a standardized parent interview developed to assess the presence and severity of autism symptoms, as well as collect information about the onset of symptoms. Specific ADI-R questions ask about skill decline and loss in language and social development. Question 11 ("Were you ever concerned that XX might have lost language skills during the first years of her/his life?") and question 25 ("Has there ever been a period when XX seemed to get markedly worse or dropped further behind in her/his social engagement, responsiveness, social relatedness, interest or involvement?") were used to define regression ("no loss" on both questions = no regression, "probable loss" or "definite loss" on either question = regression). All ADI-R raters were trained to research reliability by the instrument's developers or authorized trainers and maintained reliability at 80% or higher agreement throughout the study.

Statistical Analyses

The actual range of the prospective dimensional ratings in the sample was 5 to 15 for examiner ratings and 3 to 9 for parent ratings (with higher scores indicating better social engagement). The data were skewed, with many children scoring at the maximum of each scale. Following our previous modelling approach (Ozonoff et al., 2010), we rescaled the data by subtracting each score from the corresponding maximum (changing the range to 0–10 for examiner ratings and 0–6 for parent ratings) and analyzed them using generalized mixed-effect models for count data (McCulloch, Searle & Neuhaus, 2008). Mixed-effects Poisson regression models, using a log link, were used to estimate patterns of change in

examiner and parent prospective dimensional ratings and test whether group was related to the initial level or rate of change in these variables. This approach makes use of all available data and produces valid inference under the assumption that the missing data mechanism is ignorable. Separate models were fit for examiner and parent ratings. Both models included fixed effects for group (ASD, HR-Non-ASD, and LR-Non-ASD), linear and quadratic effects of age (centered at 6 months), and interactions between group and age. The interaction term between group and the linear effect of age allowed for differences between groups in linear trajectory and the interaction between group and the quadratic effect of age allowed for group differences in curvilinear trajectory. To account for the correlated nature of the data, the models included a random effect for child-specific intercepts.

To identify distinct onset patterns within the ASD sample, group-based trajectory analysis for count data was performed for the examiner and parent prospective dimensional ratings (Jones, Nagin & Roeder, 2001). We ran models with an increasing number of trajectories, with each having different intercepts and linear slopes. We selected the optimal number of trajectories using statistical goodness-of-fit criteria, which penalize more complex models (Bayesian information criterion [BIC] and Akaike information criterion [AIC]), as well as the bootstrap likelihood ratio test (BLRT; Nylund, Asparouhov, & Muthen, 2007). We also examined reduced models, in which we removed terms that did not add significantly to the model to see if it improved model fit. The local maximum problem was addressed by using a large number of starting points (up to 5000) to replicate each model. The selected models were used to calculate, for each child, estimates of the posterior probabilities of belonging to each trajectory; each child was then assigned to the trajectory for which they had the highest posterior probability.

All tests were two-sided, with $\alpha = 0.05$. Residual analyses and graphical diagnostics determined that the model assumptions were adequately met. Poisson mixed-model analyses were implemented using PROC GLIMMIX in SAS Version 9.4 (SAS Institute Inc., Cary, NC). Group-based trajectory analysis was performed in Mplus version 8 (Muthén & Muthén, 2017).

Results

Longitudinal Differences between the ASD and Non-ASD Groups

Table 2 and Figure 1 summarize the results of the Poisson mixed-effects models for the prospective dimensional ratings of the ASD and Non-ASD groups. For easier interpretation, we back-transformed the predicted scores from the Poisson models to the original scales in Figure 1.

For the examiner prospective dimensional ratings, all three groups had comparable values at baseline (6 months of age). The ASD group demonstrated a decrease in scores with age, while the HR-Non-ASD group showed stable scores over time and the LR-Non-ASD group demonstrated increasing scores longitudinally. By 12 months, the two Non-ASD groups had significantly higher scores than the ASD group and these differences widened over time.

For the parent prospective dimensional ratings, there were also no group differences at 6 months. The ASD group again showed a decline in social-communication with age, while the two Non-ASD groups demonstrated small gains in social-communication over time. The ASD group's scores were significantly lower than both Non-ASD groups by 12 months and the differences increased with age.

Differences within the ASD Group in Onset Patterns

The initial analyses replicated previous studies (Gammer et al., 2015; Jones & Klin, 2013; Landa et al., 2013; Ozonoff et al., 2010) and demonstrated that declining trajectories were seen only in the ASD group but did not clarify how widespread such patterns were within that group. To address this question, we next employed latent class growth models to examine potential within-group variation in onset patterns for the ASD participants ($n = 32$).

Examiner Prospective Dimensional Ratings (Session Summary Report)—Table 3 presents the model fit statistics. The best fit for examiner ratings was a two-group solution. This model had the lowest AIC and BIC indices. The p-value of the BLRT for a three-group solution was .67, also indicating that the two-group solution was better. Table 4 presents the details of the final model. Using their highest posterior group probability, the 32 participants were classified into two trajectories (see Figure 2A, which also presents the Low and High Risk Non-ASD groups as contrasts): an Early Onset/No Regression group ($n = 4$; 13%), in which examiners prospectively reported low levels of social behavior at all ages, and a Regression group ($n = 28$; 88%), in which examiners prospectively rated initially high levels of social engagement that declined over time.

Parent Prospective Dimensional Ratings (EDQ-Part 1)—In contrast to the examiner ratings, the best fit for parent ratings was a three-group solution (see Tables 3 and 4). This model had the lowest AIC and BIC and the BLRT p-value was < .001, indicating a better fit than a two-group solution. Using their highest posterior group probability, the 32 participants were classified into three trajectories: Group 1, an Early Onset trajectory, Group 2, a Declining trajectory, and Group 3, an Improving trajectory (see Figure 2B). Parents prospectively reported low levels of social-communication skills at all ages for Group 1 ($n = 4$; 13% of the sample). The majority of the sample ($n = 22$; 69%) was classified in Group 2; these children were prospectively reported by parents to show high rates of social engagement early in life, which significantly declined over time. Parents of children in Group 3 ($n = 6$; 19%) prospectively reported low levels of skills at early ages that then significantly increased over time. For the correspondence analyses below, the Group 1/Early Onset and Group 3/Improving trajectories were combined into a No Regression group ($n = 10$; 31%) and the Group 2/Decline trajectory was classified as the Regression group ($n = 22$; 69%).

Correspondence Analyses

Finally, we examined correspondence across informants (examiner v. parent), rating types (dimensional v. categorical), and timing (prospective v. retrospective) in a systematic fashion so that each analysis manipulated only one variable (e.g., examined the effect of informant on ratings that were both prospective and dimensional). See Table 3. Because traditional

kappa statistics are highly sensitive to imbalances in the marginal totals of 2×2 comparisons, we also analyzed data in terms of positive and negative agreement (Cicchetti & Feinstein, 1990). Sensitivity and specificity indices were not calculated since no particular rating could be regarded as a gold standard.

Correspondence across informants (examiner v. parent) on prospective dimensional ratings—We compared the classifications, based on trajectory analyses, of examiner prospective dimensional ratings (Session Summary Report) to parent prospective dimensional ratings (EDQ-Part 1). Overall agreement was 75% (48/64 ratings agreed on classifications of Regression or No Regression across informants), but *kappa* was poor ($\kappa = 0.30$) due to the low base rate of No Regression in both examiner and parent ratings. As seen in Table 3, positive agreement was high: When either informant's prospective ratings, on the basis of the trajectory analyses, were classified as Regression, 84% of the time there was agreement with the other rater. In contrast, negative agreement was much lower (43%), largely due to the very low frequency of No Regression classifications by both types of informants.

Correspondence across rating types (dimensional v. categorical) on parent prospective measures—Next we compared dimensional prospective parent ratings (EDQ-Part 1) to categorical prospective parent ratings (EDQ-Part 2). There was agreement for only 38/64 ratings (59%; $\kappa = 0.21$) in onset classification (Regression, No Regression) across rating types. Agreement was again higher for the presence of regression than its absence, but both positive and negative agreement statistics were low to moderate. As can be seen in Table 3, only a little over half of parents whose dimensional ratings, over time, were consistent with a regressive course also endorsed a regression when asked concurrently in a categorical manner.

Correspondence across timing (prospective v. retrospective) on parent categorical ratings—In the third set of contrasts, we compared prospective categorical parent ratings (EDQ-Part 2) to retrospective categorical parent ratings (ADI-R; *n*=31, one missing). For 42/62 ratings (68%) there was agreement in onset classification (Regression, No Regression) across prospective and retrospective measures ($\kappa = 0.35$). Of 15 parents who rated their child prospectively as having declining skills since the last visit, only 7 (47%) reported any regression when asked retrospectively when their child was 36 months old.

Discussion

The results of this study replicate earlier findings of high rates of regression in ASD (Brignell et al., 2017; Hansen et al., 2008; Ozonoff et al., 2010; Thurm et al., 2014) and suggest that a regressive pattern of onset may be much more common than previously thought, perhaps the rule rather than the exception. A majority of the prospective dimensional ratings of both parents and examiners showed a pattern of initially good social development that progressively declined over time among the infants who developed ASD, with parents rating 69% of the sample in this way and examiners rating 88% similarly. This suggests that the group-based findings of declining trajectories in previous studies (Gammer

et al., 2015; Jones & Klin, 2013; Landa et al., 2013; Ozonoff et al., 2010) were likely not due to a few dramatic outliers, but are a more general finding across individuals.

While previous studies have reported inconsistencies in onset classification between retrospective and prospective data (Ozonoff et al., 2010, 2011), the source of the discrepancies was not clear. Inconsistencies could be due to differences between reporters, with parents being more knowledgeable about their child's everyday functioning and examiners having greater expertise in differentiating typical from atypical development and diagnosing autism. Or discrepancies could be caused by methodological challenges, with expert-administered interviews that require categorical judgments having stricter standards for regression than parent ratings of behavior frequency. Or inconsistencies could arise from the timing of the assessments, with ratings of current behavior likely to be more accurate, because they do not rely upon memory, than retrospective ratings. This is the first study to disentangle these variables by collecting reports across informants, rating methods, and time. Best agreement was found when ratings were simultaneously prospective and dimensional. Agreement was markedly lower when ratings were retrospective or categorical or both. This suggests that the discrepancies previously identified between parent report of regression and other indices (Ozonoff et al., 2010, 2011, in press) are due to multiple factors, both the *way* the question is asked (categorical v. dimensional) and *when* it is asked (concurrently or in retrospect). Similarly, studies of other disorders have converged on the conclusion that reports of psychopathology are much more likely when measured prospectively than retrospectively (Moffitt et al., 2010).

It is striking that, while 69% of parents rated their child in a manner consistent with regression on a prospective dimensional measure, only 46% of them rated that same child as having lost skills using a prospective categorical measure, and only 29% reported a regression using a retrospective categorical measure. Parents were able to *implicitly* identify the changes in their child's development over time when making ratings of the frequencies of current behaviors, but often did not *explicitly* label these changes as skill loss or regression when asked in a more categorical way, whether concurrently or at a later date. This may be because categorical ratings, which inherently require comparison to previous functioning and also require an all-or-none judgment (presence or absence of a phenomenon), are more difficult discriminations to make than frequency ratings of current behavior. Although the ADI-R does not define skill loss as complete absence of a behavior, it is possible that parents may interpret ADI-R questions more strictly, which would result in under-reporting of regression on this measure.

The results of the current study have important implications beyond clarifying onset patterns and may be useful for improving screening and early detection of ASD as well. Previous studies that tracked the emergence of early symptoms (Gammer et al., 2015; Jones & Klin, 2013; Landa et al., 2013; Ozonoff et al., 2010) used prospective measures, such as behavioral coding, eye tracking, and serial developmental assessments, but these methods are labor intensive and not practical for routine use. The current results demonstrated that parent report (which *is* feasible for clinical use) can differentiate children who will ultimately be diagnosed with ASD as early as 12 months. The parent report measure used in this study had several key features that likely contributed to its utility in this regard: 1) it was

intended for prospective use, 2) it asked only about current functioning, 3) it used ratings of behavior frequency, and 4) it focused on early-appearing social behaviors, present in the first year of life (Inada et al., 2010), that have the potential to demonstrate decreases over time as ASD signs emerge. Such a brief dimensional rating scale, administered longitudinally at regular well-child health care visits, could provide a clinically feasible and cost-effective screening tool capable of detecting ASD in the first or second years of life. While several excellent ASD screeners exist (Robins et al., 2014; Wetherby et al., 2008), none place the same focus on identifying declines over time and include many behaviors (e.g., language, gestures, joint attention, play) that do not develop until the second year of life or later. If practitioners were able to identify infants at risk for ASD, *during* the decline of skills, rather than after the decline was over, it might be possible to disrupt these trajectories prior to the full onset of symptoms. Children could be provided immediate access to infant interventions (Fein et al., 2016; Rogers et al., 2014), capitalizing on still-preserved skills and harnessing the brain plasticity of early infancy to improve outcomes, lessen disability, and perhaps, prevent the full disorder from developing.

Acknowledgments

References

Andrews N, Miller E, Taylor B, Lingam R, Simmons A, Stowe J, et al. Recall bias, MMR, and autism. Archives of Diseases of Childhood. 2002; 87:493–494.

Barger BD, Campbell JM, McDonough JD. Prevalence and onset of regression within autism spectrum disorders: A meta-analytic review. Journal of Autism and Developmental Disorders. 2013; 43:817–828. [PubMed: 22855372]

Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: Diagnostic validity. British Journal of Psychiatry. 1999; 175:444–451. [PubMed: 10789276]

Brignell A, Williams K, Prior M, Donath S, Reilly S, et al. Parent-reported patterns of loss and gain in communication in 1- to 2-year-old children are not unique to autism spectrum disorder. Autism. 2017; 21:344–356. [PubMed: 27178996]

Cicchetti DV, Feinstein AR. High agreement but low kappa: Resolving the paradoxes. Journal of Clinical Epidemiology. 1990; 43:551–558. [PubMed: 2189948]

Fein, D., Helt, M., Brennan, L., Barton, M. The Activity Kit For Babies And Toddlers At Risk: How To Use Everyday Routines To Build Social And Communication Skills. New York, NY: Guilford Press; 2016.

Finney HC. Improving the reliability of retrospective survey measures: Results of a longitudinal field survey. Evaluation Review. 1981; 5:207–229.

Gammer I, Bedford R, Elsabbagh M, Garwood H, Pasco G, Tucker L, et al. Behavioral markers for autism in infancy: Scores on the Autism Observational Scale for Infants in a prospective study of at-risk siblings. Infant Behavior and Development. 2015; 38:107–115. [PubMed: 25656952]

Hansen RL, Ozonoff S, Krakowiak P, Angkustsiri K, Jones C, Deprey LJ, et al. Regression in autism: Prevalence and associated factors in the CHARGE study. Ambulatory Pediatrics. 2008; 8:25–31. [PubMed: 18191778]

Henry B, Moffitt TE, Caspi A, Langley J, Silva PA. On the "remembrance of things past" A longitudinal evaluation of the retrospective method. Psychological Assessment. 1994; 6:92–101.

Hus V, Taylor A, Lord C. Telescoping of caregiver report on the autism diagnostic interview-revised. Journal of Child Psychology and Psychiatry. 2011; 52:753–760. [PubMed: 21410473]

Inada N, Kamio Y, Koyama T. Developmental chronology of preverbal social behaviors in infancy using the M-CHAT: Baseline for early detection of atypical social development. Research in Autism Spectrum Disorders. 2010; 4:605–611.

Jones BL, Nagin DS, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. Sociological Methods and Research. 2001; 29:374–393.

Jones EJH, Gliga T, Bedford R, Charman T, Johnson MH. Developmental pathways to autism: A review of prospective studies of infants at risk. Neuroscience and Biobehavioral Reviews. 2014; 39:1–33. [PubMed: 24361967]

Jones W, Klin A. Attention to eyes is present but in decline in 2 to 6 month old infants later diagnosed with autism. Nature. 2013; 504:427–431. [PubMed: 24196715]

Kern JK, Geier DA, Geier MR. Evaluation of regression in autism spectrum disorder based on parental reports. North American Journal of Medical Sciences. 2015; 6:41–47.

Landa RJ, Stuart EA, Gross AL, Faherty A. Developmental trajectories in children with and without autism spectrum disorders: The first 3 years. Child Development. 2013; 84:429–442. [PubMed: 23110514]

Lord, C., Rutter, M., DiLavore, PC., Risi, S., Gotham, K., Bishop, SL. Autism Diagnostic Observation Schedule Manual. 2. Torrance, CA: Western Psychological Services; 2012.

Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of Autism and Developmental Disorders. 1994; 24:659–685. [PubMed: 7814313]

Lord C, Shulman C, DiLavore P. Regression and word loss in autistic spectrum disorders. Journal of Child Psychology and Psychiatry. 2004; 45:936–955. [PubMed: 15225337]

McCulloch, CE., Searle, SR., Neuhaus, JM. Generalized, Linear, and Mixed Models. 2. New York, NY: John Wiley & Sons; 2008.

Moffitt TE, Caspi A, Taylor A, Kokaua J, Milne BJ, Polanczyk G, et al. How common are mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. Psychological Medicine. 2010; 40:899–909. [PubMed: 19719899]

Muthén, LK., Muthén, BO. Mplus User's Guide. 8. Los Angeles, CA: Muthén & Muthén; 1998–2017.

Nylund KL, Asparouhov A, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling. 2007; 14:535–569.

Ozonoff S, Iosif A, Baguio F, Cook IC, Hill MM, Hutman T, et al. A prospective study of the emergence of early behavioral signs of autism. Journal of the American Acadamy of Child and Adolescent Psychiatry. 2010; 49:258–268.

Ozonoff S, Iosif A, Young GS, Hepburn S, Thompson M, Colombi C, et al. Onset patterns in autism: Correspondence between home video and parent report. Journal of the American Acadamy of Child and Adolescent Psychiatry. 2011; 50:796–806.

Ozonoff S, Li D, Deprey L, Hanzel EP, Iosif A. Reliability of parent recall of ASD symptom onset and timing. Autism. (in press).

Ozonoff S, Williams BJ, Landa R. Parental report of the early development of children with regressive autism: The "delays-plus-regression" phenotype. Autism. 2005; 9:495–520. [PubMed: 16287702]

Pickles A, Pickering K, Taylor C. Reconciling recalled dates of developmental milestones, events and transitions: A mixed generalized linear model with random mean and variance functions. Journal of the Royal Statistical Society, Series A. 1996; 159(Part 2):225–234.

Robins DL, Casagrande K, Barton M, Chen CA, Dumont-Mathieu T, Fein D. Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). Pediatrics. 2014; 133:37–45. [PubMed: 24366990]

Rogers SJ, Vismara L, Wagner AL, McCormick C, Young G, Ozonoff S. Autism treatment in the first year of life: A pilot study of Infant Start, a parent-implemented intervention for symptomatic infants. Journal of Autism and Developmental Disorders. 2014; 44:2981–2995. [PubMed: 25212413]

SAS Institute Inc.. SAS/STAT Version 9.4. Cary, NC: 2002–2012.

Shumway S, Thurm A, Swedo SE, Deprey L, Barnett LA, Amaral DG, et al. Symptom onset patterns and functional outcomes in young children with autism spectrum disorders. Journal of Autism and Developmental Disorders. 2011; 41:1727–1732. [PubMed: 21360021]

Thurm A, Manwaring SS, Luckenbaugh DA, Lord C, Swedo SE. Patterns of skill attainment and loss in young children with autism. Development and Psychopathology. 2014; 26:203–214. [PubMed: 24274034]

Wetherby AM, Brosnan-Maddox S, Peace V, Newton L. Validation of the infant-toddler checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. Autism. 2008; 12:487–511. [PubMed: 18805944]

Lay Summary

This study examines different ways of measuring the onset of symptoms in autism spectrum disorder (ASD). The present findings suggest that declining developmental skills, consistent with a regressive onset pattern, are common in children with ASD and may be more the rule than the exception. The results question the accuracy of widely used methods of measuring symptom onset and argue against their widespread use.
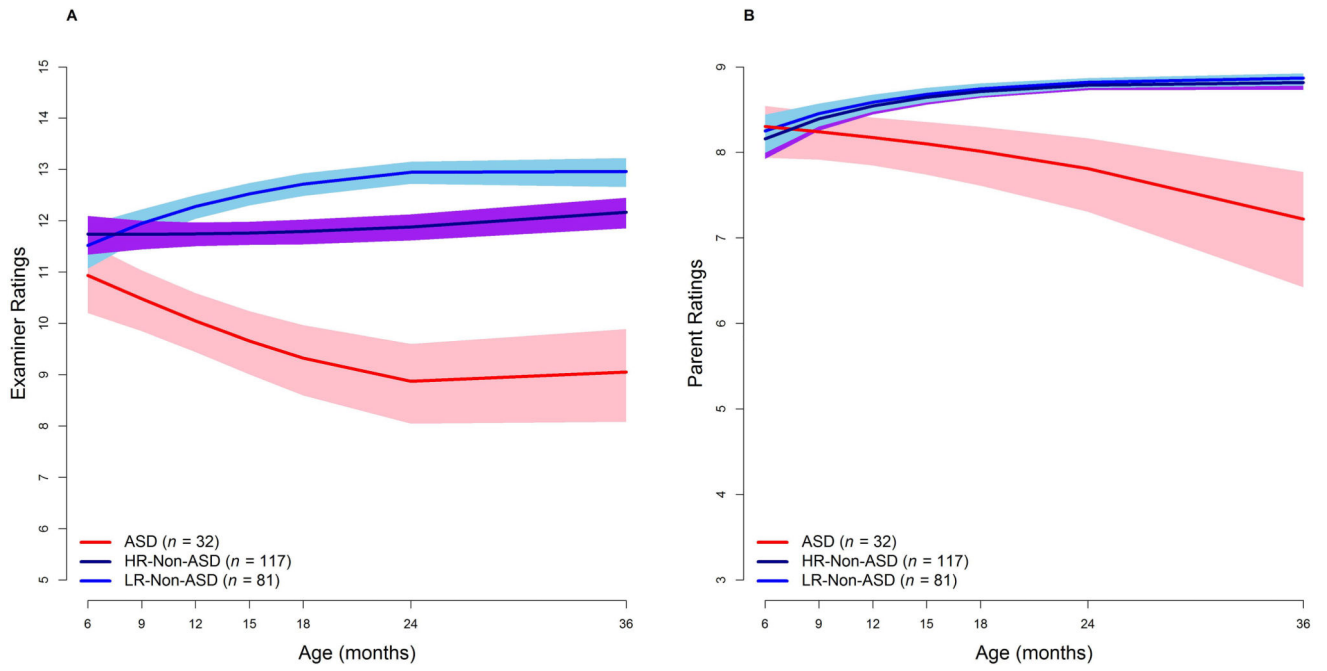
Figure 1.
Trajectories and 95% confidence intervals for the ASD and Non-ASD groups in examiner
and parent prospective dimensional ratings
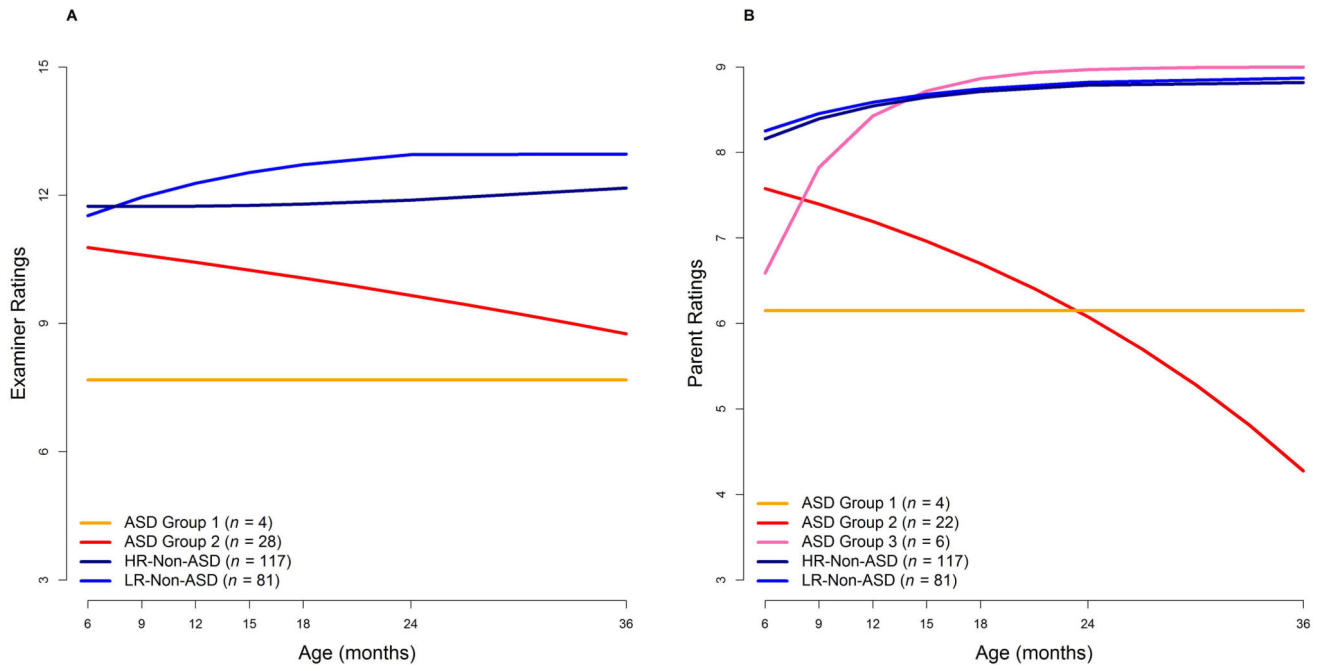
Figure 2.
Trajectories within the ASD group for examiner and parent prospective dimensional ratings

Table 1

Methods of measuring onset

Assessment Instrument	Informant	Timing	Type
Session Summary Report	Examiner	Prospective	Dimensional
EDQ - Part 1	Parent	Prospective	Dimensional
EDQ - Part 2	Parent	Prospective	Categorical
Autism Diagnostic Interview-Revised	Parent	Retrospective	Categorical

Note: EDQ = Early Development Questionnaire

Prospective = rated at each visit

Retrospective = rated at the end of the study (36 months)

Dimensional = frequency of social behaviors

Categorical = presence or absence of skill decline or loss

Table 2

Parameter estimates (SE)[a] for the mixed-effects Poisson regression models predicting prospective dimensional ratings of social-communication

	Examiner Ratings	Parent Ratings
Estimated trajectory for LR-Non-ASD group		
Baseline (6 months)	1.247 (0.063)***	−0.289 (0.150)[†]
Linear change with age (per 6 months)	−0.279 (0.056)***	−0.659 (0.146)***
Quadratic change with age (per 6 months)	0.035 (0.010)***	0.062 (0.030)*
Estimated difference between ASD and LR-Non-ASD groups		
Baseline (6 months)	−0.156 (0.105)	−0.070 (0.262)
Linear change with age (per 6 months)	0.507 (0.086)***	0.822 (0.203)***
Quadratic change with age (per 6 months)	−0.065 (0.015)***	−0.057 (0.039)
Estimated difference between HR-Non-ASD and LR-Non-ASD groups		
Baseline (6 months)	−0.065 (0.086)	0.117 (0.195)
Linear change with age (per 6 months)	0.285 (0.075)***	−0.032 (0.184)
Quadratic change with age (per 6 months)	−0.041 (0.013)**	0.015 (0.038)

Note:

*
$p < .05$,

**
$p < .01$,

***
$p < .001$,

[†]
$p < .10$.

SE = standard error

[a]
From mixed-effect Poisson regression models with a log link fitted to the rescaled data with person as a random effect. Ratings on the original scales (ranging from 5–15 for examiners and 3–9 for parents) were rescaled (to 0–10 and 0–6, respectively) by subtracting the scores from the maximum of the corresponding scales, so that zero scores correspond to no deficits in social-communication.

Table 3

Model fit statistics and the number of children assigned to each group for latent class growth models with two to four classes for examiner and parent prospective dimensional ratings

Number of classes	Number of parameters	BIC [1]	AIC [1]	BLRT [2]	Number of children assigned to each class			
					1	2	3	4
Examiner prospective dimensional ratings								
Two	12	856.2	838.6	<.0001	6	26	–	–
Two [3]	**11**	**853.2**	**837.0**	**<.0001**	**4**	**28**	–	–
Three	15	863.2	841.2	.29	3	25	4	–
Parent prospective dimensional ratings								
Two	12	568.1	550.5	<.0001	6	26	–	–
Three	15	561.5	539.5	<.0001	8	18	6	–
Three [3]	**14**	**556.1**	**535.6**	**<.0001**	**4**	**24**	**6**	–
Four	18	564.2	537.9	.18	2	13	4	13

BIC = Bayesian Information Criterion; AIC = Akaike Information Criterion; BLRT = Parametric bootstrapped likelihood ratio test *p*-values. Bold numbers correspond to final models.

[1] Lower numbers indicate more optimal model fit.

[2] Small *p*-values of the BLRT test support the model with *k* vs. the model with *k*–1 classes

[3] Slope in group 1 was fixed at 0 in these models

Table 4

Final latent class growth models for parent and examiner prospective dimensional ratings

	Parameter	Estimate	*SE*	*p*-value
Examiner Ratings, Group (%)				
Group 1 (20.3%)	Intercept	1.991	0.298	<.001
	Linear	–	–	–
Group 2 (80.7%)	Intercept	1.442	0.205	<.001
	Linear	0.013	0.006	.019
Parent Ratings, Group (%)				
Group 1 (16.1%)	Intercept	1.048	0.064	<.001
	Linear	–	–	–
Group 2 (66.3%)	Intercept	−0.353	0.218	.105
	Linear	0.040	0.008	<.001
Group 3 (17.6%)	Intercept	0.880	0.547	.108
	Linear	−0.239	0.089	.007

SE = standard error. Ratings on the original scales (ranging from 5–15 for examiners and 3–9 for parents) were rescaled (to 0–10 and 0–6, respectively) by subtracting the scores from the maximum of the corresponding scale, so that zero scores correspond to no deficits in social-communication. Latent class growth models for count data were fitted to the rescaled data.

Table 5

Agreement among methods of measuring onset

Correspondence across Informants (on prospective dimensional measures)

		Parent Rating	
		Regression	No Regression
Examiner Rating	Regression	21	7
	No Regression	1	3
	κ (95% *CI*): 0.30 (−0.03 to 0.64)		
	Percent agreement: 75%	Positive agreement: 84%	Negative agreement: 43%

Correspondence across Rating Types (on parent prospective measures)

		Dimensional Rating	
		Regression	No Regression
Categorical Rating	Regression	12	3
	No Regression	10	7
	κ (95% *CI*): 0.21 (−0.10 to 0.51)		
	Percent agreement: 59%	Positive agreement: 65%	Negative agreement: 52%

Correspondence across Timing (on parent categorical measures)

		Retrospective Rating [1]	
		Regression	No Regression
Prospective Rating	Regression	7	8
	No Regression	2	14
	κ (95% *CI*): 0.35 (0.04 to 0.65)		
	Percent agreement: 68%	Positive agreement: 58%	Negative agreement: 74%

[1] Retrospective rating missing for 1 participant

Note:

Positive agreement is the proportional agreement in positive ("Regression") ratings, calculated as:

$$\frac{2 * \text{Number of "Regression" ratings by both raters}}{\text{Number of "Regression" ratings by rater 1} + \text{Number of "Regression" ratings by rater 2}}$$

Negative agreement is the proportional agreement in negative ("No Regression") ratings, calculated as:

$$\frac{2 * \text{Number of "No Regression" ratings by both raters}}{\text{Number of "No Regression" ratings by rater 1} + \text{Number of "No Regression" ratings by rater 2}}$$