

Clinical Signs Associated with Earlier Diagnosis of Children with Autism Spectrum Disorder

Nachum Sicherman, PhD, Columbia University.

Jimmy Charite, BA., Columbia University.

Gil Eyal, PhD, Columbia University.

Magdalena Janecka, PhD, Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai.

George Loewenstein, PhD, Carnegie Mellon University.

Kiely Law, MD, MPH, Kennedy Krieger Institute and Johns Hopkins School of Medicine.

Paul H. Lipkin, MD, Kennedy Krieger Institute and Johns Hopkins School of Medicine.

Alison R. Marvin, PhD, Kennedy Krieger Institute and Johns Hopkins Bloomberg School of Public Health.

Joseph D. Buxbaum, PhD, Seaver Autism Center for Research and Treatment, Department of Psychiatry, Friedman Brain Institute, Mindich Institute for Child Health and Development, Icahn School of Medicine at Mount Sinai.

Abstract

The objective of this study is to gain new insights into the relationship between clinical signs and age at diagnosis utilizing a new, large, online survey of parents of children diagnosed with ASD. Using multiple statistical approaches, we find that clinical signs and symptoms that most strongly predict early diagnosis are not necessarily specific to autism, but rather those that initiate the process that eventually leads to an ASD diagnosis. Given the high correlations between symptoms, only a few symptoms are found to be important in predicting early diagnosis. Even though our data are drawn from parents' retrospective accounts, we provide evidence that parental recall bias and/or hindsight bias did not play a significant role in shaping our results.

Introduction

Given evidence that the timing of interventions can have a large impact on trajectories of children diagnosed with autism spectrum disorder (ASD) (Filipek *et al.*, 1999; Dawson, 2008; Warren *et al.*, 2011; Anderson, Liang and Lord, 2014; Bradshaw *et al.*, 2014; McPheeters *et al.*, 2016; Siu *et al.*, 2016), timing of diagnosis has assumed increasing significance. Timely diagnosis not only enables early clinical and educational intervention, but also relieves parental distress (Gray, 2002; Pisula, 2003), which in turn ameliorates secondary impacts of ASD (Howlin, 1999). Obtaining an accurate estimate of the number of children affected across different age categories is also essential to guide policies and plan service needs (Wiggins, Baio and Rice, 2006).

There is extensive evidence, from both population and clinical samples, that ASD can be reliably diagnosed before the age of 2 (Howlin and Asgharian, 1999; Sivberg, 2002; Watson *et al.*, 2007; Zuckerman, Lindly and Sinche, 2015; Christensen *et al.*, 2016; Pierce, Courchesne and Bacon, 2016), yet the Center for Disease Control and Prevention (CDC) estimates that the median age at diagnosis is 3 years 10 months for more severe autistic disorders, 4 years 1 month for pervasive developmental disorder – not otherwise specified, and 6 years 2 months for Asperger disorder (Christensen *et al.*, 2016). Thus, most children who are ultimately diagnosed with ASD are not diagnosed until after the age of 4, despite the fact that parents often express concerns a year or two before this age (Zuckerman, Lindly and Sinche, 2015). Family members and caretakers are typically the first to raise concerns, usually before the second year of age (Frith and Soares, 1993; Howlin and Asgharian, 1999; Sivberg, 2002).

Numerous studies show that both the prevalence and age of diagnosis are correlated with family socioeconomic status (SES). Children of more educated and wealthier parents are more likely to be diagnosed with ASD, and to be diagnosed earlier. With regard to race and ethnicity, the evidence is mixed. While minority children are less likely to be diagnosed with ASD (Mandell *et al.*, 2009), differences in age of diagnosis across racial and ethnic groups are not significant after controlling for parents' wealth and education (Mandell *et al.*, 2002, 2009; Wiggins, Baio and Rice, 2006; Shattuck *et al.*, 2009). Other factors shown to be correlated with age of diagnosis are family structure (e.g., the presence of grandparents in the household) (Sicherman *et al.*, 2018), birth order, the type of diagnosis (e.g., Asperger being diagnosed later than other conditions), and level of urbanization; children in urban areas are diagnosed significantly earlier (Mandell and Palmer, 2005).

In the US, the median age at diagnosis is highly variable across states (Shattuck *et al.*, 2009; Christensen *et al.*, 2016). Rates of diagnosis in different states are affected by the availability of resources (Mandell and Palmer, 2005; Palmer *et al.*, 2005), distribution of clinics, density of doctors, existence of early screening and intervention programs, and strength of parents' organizations. Rates of diagnosis also correlate with the type of welfare regime in the state, and the state's history of de-institutionalization for intellectual disabilities (Eyal *et al.*, 2010). Across

states, however, the age of diagnosis is decreasing over time (Mandell and Palmer, 2005; Christensen *et al.*, 2016), suggesting continuous improvements in ASD detection.

Given a lack of clinical biomarkers for ASD, currently, diagnosis typically relies on both behavioral observation by a qualified clinician and parental report of developmental history and current presentation. Clinical signs are clearly important, given that family members, caretakers, and health-care professionals, who are typically the first to raise concerns, necessarily rely on observations of signs to initiate the processes that eventually lead to diagnosis and treatment. However, only a small number of studies have examined the correlation between the age of diagnosis and the presence and severity of various specific signs, either separately or jointly (Mandell, Novak and Zubritsky, 2005; Fountain, King and Bearman, 2011). Several studies have shown that *overall* severity leads to earlier diagnosis (Wiggins, Baio and Rice, 2006), and that specific or broad categories of signs are associated with either earlier or later diagnosis (Daniels and Mandell, 2014). However, we are not aware of any study that systematically looks at a large set of symptoms and signs and their level of severity, their correlation with one-another, and their connection to age of diagnosis.

The objective of the research reported in this article is to disentangle the relationship between symptoms, clinical signs, and age of diagnosis, utilizing a new, large, online survey of parents of children diagnosed with ASD. In addition to asking about the age at which their child was diagnosed with ASD, the survey presented parents with a list of 25 clinical signs and symptoms, and asked them, using a slider, to indicate which signs their child exhibited around the time of diagnosis as well as the level of severity. We find that the clinical signs that most strongly predict early diagnosis are not necessarily specific to autism (e.g., speech delay), but more likely index what parents expect of normal development in the home in the first 2 years of life. Signs not specifically associated with ASD can have this effect if they lead parents to seek input from professionals who are knowledgeable about ASD. We discuss practical implications of our findings for efforts to reduce the age of ASD diagnosis.

Since our analysis is based on a retrospective survey, there are two types of potential biases. The first is recall bias: The accuracy and quality of responses could diminish with the time elapsed since the child was diagnosed until the survey date. We conducted several statistical analyses to test whether our data are subject to recall bias and these did not yield evidence of such bias (see Appendix B for detailed analysis). Second, parents' recall of the type of clinical signs and symptoms exhibited by their child prior to diagnosis could be distorted by their awareness, at the time of completing the survey, of their child's diagnosis. However, our results run contrary to what one would expect if such a bias were present. If parents' recall was influenced by current diagnosis in a fashion consistent with hindsight bias – the tendency to believe that one knew in the past what one knows in the present – then they would be more likely to emphasize symptoms that are specific to autism, i.e., that differentiate autism from other conditions, yet we find the opposite. We caution, however, that the same does not necessarily hold for parents' recall of the *severity* of their children's signs. Consequently, we have less confidence in our results regarding

severity. Finally, it should be noted that prospective epidemiological research on this question is impractical. Only a small fraction of all children are diagnosed with ASD; hence an epidemiological sample of parents of children who have not yet been diagnosed with ASD will contain too few cases of children who will end up with ASD diagnosis later. For this reason, we believe that despite the well-known problems with retrospective studies, our approach – especially given the direction of our results and the additional tests we conducted to rule out retrospective bias (see Appendix B) – is the best way to shed light on this crucial question.

Methods

Sample

An online survey, approved by the Columbia University Institutional Review Board, was completed by 1,815 parents of children who were previously diagnosed with ASD. Potential participants, who were contacted by email, were selected from the Interactive Autism Network (IAN) Research Database and Registry at the Kennedy Krieger Institute, Baltimore, MD.

The IAN Research database and research registry was designed to facilitate ASD research efforts by informing participants about studies for which they qualify. All the participants in the database are parents of children whose ASD diagnosis has been clinically validated (Lee *et al.*, 2010; Marvin *et al.*, 2014) as well as verified by a review of parent- and professional-provided medical records (Daniels *et al.*, 2012). To date, IAN Research has provided recruitment and/or data services for more than 500 research studies. IAN Research is governed by a Johns Hopkins Medicine IRB (NA_00002750; PI: Dr. Paul H. Lipkin).

It is difficult to estimate response rate. We estimate that the recruitment email went to approximately 11,300 e-mail addresses (excluding emails of participants reported as deceased, bounced emails, and emails excluded per participant request). However, we cannot tell how many emails were actually received and read. Thus, the response rate is at least 16%.

Demographic and socio-economic information on parents collected in the study included ethnicity, education, household income, and urban/rural residence. Data from the survey was supplemented with information previously collected by IAN on each of the children in our sample.

In cases where more than one child in the family was diagnosed, parents were asked to answer the survey with regard to their first diagnosed child. After deleting cases for which crucial variables were missing (e.g., age of diagnosis), the sample size was reduced to 1,743. Descriptive statistics of the sample are reported in Table 1.

[PLACE TABLE 1 HERE]

Survey

This survey is part of a larger project that addresses various factors that could affect the age of diagnosis. In this study we focus on a limited number of questions from the survey. Parents were provided with a list of 25 signs and asked to report those exhibited by their child around the time of his/her diagnosis. Most signs were taken from the diagnostic criteria for autism from the Diagnostic and Statistical Manual for Mental Disorders (DSM). Some modifications to the list were made following a pilot study in which parents also listed signs that are not unique to individuals with autism (Sicherman *et al.*, 2018). This prior study found that there were some signs that, while not unique to ASD, were frequently cited by parents, and may have led them to seek professional help and to obtain a diagnosis.

For each clinical sign or symptom, the respondent could indicate, using a slider, the level of severity ranging from “none” to “severe.” Other intermediate levels were labeled (at equal intervals) as “minor,” “moderate,” and “serious.” Respondents could position the slider anywhere along the scale. Responses were translated to a numerical value on a 0-10 scale. A zero response indicates that a child did not display that particular symptom. The list of the symptoms that parents were asked to assess is presented in Table 2.

[PLACE TABLE 2 HERE]

Covariates

In addition to the detailed description of clinical signs, we control in some statistical analyses for socio-economic variables such as ethnicity, parents’ education, family income, and location (urban vs. rural). Other covariates that we include in some analyses are year of birth, to account for the overall increase in diagnosis over time, and a dummy variable for a diagnosis of Asperger, to account for the fact that such children tend to be diagnosed later. Given the retrospective nature of the study, we also use the time elapsed between the time of diagnosis and the date of survey to check for potential biases (see the Study Limitations section and Appendix B).

Statistical analyses

To explore characteristics of the data we examined the median and the distribution of all items on the clinical signs list, as well as the correlation structure between individual signs.

Subsequently, we used two statistical methods to study the link between symptoms and age of diagnosis: factor analysis and regression trees. Each of these methods takes a different approach to addressing the same problem: identifying the patterns of signs that lead, most reliably, to early diagnosis of ASD.

Factor analysis

Factor analysis is designed to distill multiple variables into smaller sets of factors representing key sources of unique – i.e., non-redundant – variance. For example, a large number of questions designed to measure personality could be distilled, using factor analysis, into measures of a much smaller number of specific traits, such as introversion and impulsivity. This approach allowed us to avoid the problems associated with estimating the effects of multiple signs that are highly correlated by replacing them with the scores of a limited number of uncorrelated factors.

Regression Trees

Regression trees is a class of machine learning models that make it possible to estimate non-linear relationships in an easily interpretable fashion. Clinical signs whose severity allows for best prediction of early vs late diagnosis were identified in an iterative and hierarchical way. The regression tree analysis first identifies, for each sign, the level of severity that best splits the sample between children who are diagnosed earlier and those diagnosed later. Then, among all signs, the signs that allows for the best separation of the sample into children with early versus late diagnosis is selected as the first node in the tree. Two branches are then created, one for children with early diagnosis and one for children with late diagnosis. The above procedure is repeated for each branch, creating two new nodes. This process continues at each node of the tree, until no symptom can further split the sample into two groups such that the differences in age of diagnosis between the groups (early and late diagnosis) are statistically different. In the Results section, we provide more details on the tree construction.

Results

Sample demographic and symptom characteristics

Table 1 reports the socio-economic characteristics of the sample and compares them to that of the US population. Similar to other studies, the children in our sample are from wealthier, more educated, and more urban families than the US average. Representation of minority groups, especially Hispanic and Black, is also lower than their share of the US population. The effects of various socioeconomic variables on the age of diagnosis in our sample are reported in Appendix Table 1.

The percent of children that, at time of diagnosis, displayed each of the signs listed in the survey, as well as (for those who reported them) their median values of severity, are presented in Table 2. Figure 1 presents the value distributions of each signs. As can be seen, the modal responses for most signs are either zero (indicating that a child did not exhibit the symptom) or the highest severity level (10). However, there are sufficient numbers of intermediate severity values reported to render severity level a potentially useful input into statistical analyses.

[PLACE FIGURE 1 HERE]

Correlations between severity and age of diagnosis (Univariate Analysis).

Before presenting results using factor analysis and regression trees, we report some basic correlations between each sign and the age of diagnosis. Figure 2 displays graphically, for each sign, the average effect of a one unit increase in reported severity on the age of diagnosis (in months). Each horizontal line describes the 95% confidence interval for the effect, and the dot in the center displays the point estimate (see Appendix A for more details).

[PLACE FIGURE 2 HERE]

For most signs, higher severity was predictive of an earlier age of diagnosis (the lines are on the left side of the panel, indicating a negative relationship between severity and age of diagnosis). Delayed speech, delay in response to own name, and lack of gesture had the strongest effects. Most regressions of skills were associated with an earlier age of diagnosis, except for loss of motor and daily living skills, whose effects were non-significant. However, given that regressive symptoms were less frequently reported by parents, their usefulness for early diagnosis might be limited.

Interestingly, signs associated with aggression as well as “need for sameness” and sensory hyperreactivity are *positively* correlated with the age of diagnosis; children exhibiting these symptoms are diagnosed later, on average. In the Discussion Section we provide possible explanations to this finding.

A limitation of the univariate analysis is that the existence of some signs may be correlated with the existence of others, so that individual signs provide redundant clues about a child’s condition. We use two methodologies to estimate the joint effects of various symptoms on the age of diagnosis: factor analysis and regression trees.

A seemingly natural approach to dealing with this problem would be to estimate the effect of one sign controlling for the effect of each of the others in a multiple regression. This approach proved infeasible due to multicollinearity. The high correlations between many of the signs (see Table 3) increases the variance of the coefficient estimates and makes the coefficient estimates unstable and difficult to interpret. We, therefore, use factor analysis and regression trees to avoid such problems. The estimation results using a multiple regression are reported in Appendix A.

Factor analysis

Factor analysis is most suitable when the correlation between variables of interest is relatively high. A visual examination of the correlations across signs (Table 3) shows a relatively large number of correlations with values of 0.3 and higher (35.7% of the pairwise correlations). Using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, the standard test of whether a data set is appropriate for factor analysis, yielded a value of 0.91, suggesting a high degree of suitability for factor analysis (Kaiser and Rice, 1974).

[PLACE TABLE 3 HERE]

Using varimax rotation, we identified 5 distinct factors. Table 4 reports the mapping of the 25 signs to these five factors and the labels we chose for these factors. Three of those factors mapped well onto a triad of ASD diagnostic impairments (social interaction, communication, and restrictive/repetitive behaviors). Two additional factors represented items relevant to developmental regression and aggressive behaviors.

[PLACE TABLE 4 HERE]

The next step in our analysis was to use the weights from the factor analysis to generate factor scores for each child and each factor. The factor scores are then included as independent variables in a regression model in which, as before, the dependent variable is age of diagnosis. Therefore, instead of including all 25 symptoms in one regression, we use the 5 factor scores. This avoids the problem of multicollinearity, since factors are, by construction, uncorrelated with one-another, and, by reducing the number of independent variables, increases statistical power.

The results of this regression are reported in Table 5, Column 1. The factor representing communication difficulties was the strongest predictor of age of diagnosis, with higher levels of severity associated with significantly lower age of diagnosis. Unfortunately, one of the drawbacks of factor analysis is that there is no interpretation to the values of the estimated coefficients.

[PLACE TABLE 5 HERE]

Developmental regression and restricted-and-repetitive-behaviors (RRBs) were also predictive of earlier age at diagnosis, although much less predictive than communication difficulties. Presence of aggressive behaviors, on the other hand, was associated with a delayed diagnosis. Adding to the regression socio-economic indicators, the child's year of birth, and an indicator for an Asperger diagnosis (Table 5, Column 2), did not much affect the results.

To test the hypothesis that individual signs play an important role in predicting age of diagnosis beyond what is captured by the overall severity of the child condition, we constructed a measure of overall severity by counting the number of signs that the parents reported with a positive level of severity. The results, reported in Table 5, column 3, show that, on average, a one unit increase in the number of signs reduced the age of diagnosis by almost one month. However, when the regression also includes the five factors representing the effects of the individual signs (Table 5, columns 4 and 5), the effect of overall severity becomes insignificant.

Regression Trees

We conducted a regression tree analysis using a package called RPART (Therneau and Atkinson, 2017). As described earlier, the first step in constructing the tree is to find, for each sign, the level of severity (on the 0-10 scale) that best split the sample into two groups, those who are

diagnosed earlier and those who are diagnosed later. The estimation procedure is similar to running numerous Ordinary Least Squares (OLS) regressions for each sign, where the dependent variable is the age of diagnosis and the independent variable is a dummy variable indicating high vs. low severity of the sign. In each regression we use a different level of severity to construct the cutoff for the high/low dummy variable. The best cutoff point is produced by the RPART package using “Gini Index” and is, intuitively, similar to choosing a cutoff point that produces the highest explained variation (R^2) for each sign.

The results of this first step are presented in Table 6. For example, for “delayed speech” the level of severity of 5.75 best divides the sample by age of diagnosis. For children with severity levels of 5.75 or below, the mean age of diagnosis is 63.7 months (median=58), while for children with levels of severity above 5.75 the mean age of diagnosis is 34.9 (median=30). The extremely low p-values indicate that, with almost certainty, the population mean age of diagnosis for children above the split is different (and lower) than those with severity level below the split.

[PLACE TABLE 6 HERE]

At the bottom of the table are signs for which no level of severity could separate the sample into two groups where the age of diagnosis of one group was significantly different from that of the other.

In creating the tree (Figure 3) we work from top down, first picking the sign for the top of the tree that best divides the sample between children who are diagnosed earlier and those who are diagnosed later, based on the criteria discussed above. This is “delayed speech,” which, with a cutoff level of 5.8 (all cutoffs numbers are rounded in the figure) splits our sample into two groups: 62% ($n=740$) with a severity level of 5.8 or above have an average age of diagnosis of 35 months, and 38% of the sample ($n=463$), with a severity level below 5.8, have an average age of diagnosis of 64 months.

[PLACE FIGURE 3 HERE]

Next, for each of these sub-groups, we again split the sample, using the same criteria. We repeat this process until we cannot split the sample into two sub-groups such that the difference in the age of diagnosis is statistically significant.

From Figure 3, we can see that when we limit the sample to the children with high severity of delayed speech (node 2, where severity level is ≥ 5.8), the only remaining sign that further splits this sub-sample is “lack of gestures” where the cutoff level of 5.7 splits our sample into two groups (nodes 4 and 5): 42% ($n=500$) with a severity level of 5.7 or above have an average age of diagnosis of 32 months, and 20% of the sample ($n=240$), with a severity level below 5.7, have an average age of diagnosis of 42 months. As the graph shows, these two groups, both relatively large, cannot be further divided.

Going back up the tree, of children with relatively low severity of delayed speech (node 3), the symptom that best divides this sample is “delayed response to name.” Notice that here the cutoff level is quite low (0.75), suggesting that among children with no (or low level of) delayed speech, any level of delayed response to name is important in predicting the age of diagnosis. The cutoff level of 0.75 splits the subsample into two groups (nodes 6 and 7): 22% (n=264) with a severity level of 0.75 or above have an average age of diagnosis of 55 months (node 6), and 17% of the sample (n=199), with a severity level below 0.75, have an average age of diagnosis of 75 months (node 7). Again, the rationale for such a split seems intuitive, at least after the fact. Speech could be delayed for a variety of reasons and the delay can take many forms, for example, an ability to receive and understand speech without the capacity to speak. If the child does not respond to his name, however, the lack of response is likely to be more diagnostic, to be indicative of a lack of attachment or lack of social awareness, “autistic aloneness.” Combined with difficulties in initiating relationships we get the “Asperger” type and a later diagnosis (node 8).

Looking at the sample represented by node 6, we see that the sign that best splits this sample is “difficulties in initiating and/or maintaining relationships.” The cutoff level of 6.0 splits our sample into two groups (nodes 8 and 9). Fourteen percent (n=170) with a severity level of 6 or above have an average age of diagnosis of 61 months, and 8% of the sample (n=94), with a severity level below 6, have an average age of diagnosis of 45 months. It should be noted that here, children that exhibit the sign at higher severity are actually diagnosed at an older age.

Moving up to the sample represented by node 7 (low levels of delayed speech and low levels of delayed response to own name), the symptom that best split the sample is “need for sameness,” and the cutoff level of 2.3 splits our sample into two groups (nodes 11 and 15). Here again, children who exhibit the sign at higher severity are diagnosed at an older age. The 14% (n=174) with a severity level of 2.3 or above have an average age of diagnosis of 79 months, and the 2% of the sample (n=25), with a severity level below 2.3, have an average age of diagnosis of 48 months. This group cannot be further split, most likely due to small sample size.

Going back to node 11, the sign that best split this sample is “played with toys or objects in an unusual way.” The cutoff level of 5.6 splits our sample into two groups (nodes 10 and 13). Children with the higher level of severity have a lower age of diagnosis. The 6% of our sample (n=67) with a severity level of 5.6 or greater have an average age of diagnosis of 68 months, and 9% of the sample (n=107), with a severity level below 5.6, have an average age of diagnosis of 86 months. Notice that this sub-group has the highest age of diagnosis so far. Looking at this specific subgroup we see that once again “delayed speech” is the sign that best splits it. Note that we have here a small sub-group of children with relatively low levels of delayed speech. However, when limiting the sample to this sub-group, those with relatively more delayed speech are diagnosed earlier. A cutoff level of 3.8 divides the group into our last two subgroups, represented by nodes 12 and 14. Children with delayed speech levels of 3.8 or above are

diagnosed at an average age of 56 months, and children with severity level below 3.8 are diagnosed at an average age of 91 months. The groups' sizes are 15 (1%) and 92 (8%) respectively.

In sum, the regression tree analysis identifies speech delay, lack of gestures and delayed response to name – all components of factor 3 in the factor analysis - as the key signs leading to early diagnosis (the lighter the node's color in Figure 3, the earlier is the diagnosis).

Discussion

Our study analyzed, using several different analytical approaches, variation in age of diagnosis as a function of a child's signs and symptoms at time of diagnosis. Leveraging a unique sample of over 1,600 children with an ASD diagnosis, for whom detailed demographic and clinical information was available, allowed us to explore a variety of clinical factors associated with age of diagnosis, while controlling for key environmental factors identified by previous studies (SES, urban/rural residence and cohort).

The different methodologies we employed yielded overlapping, but informatively distinct findings. All analyses strongly point towards deficits in early communication as an important factor in early diagnosis, findings that corroborate and build on previous research. We showed that deficits in early communication are predictive of age at diagnosis over and above overall symptom severity. Importantly, these symptoms were observed by parents much earlier than the current median age of diagnosis. This suggests the potential value of designing targeted tools that could enable family members to identify key early symptoms of ASD before they come to the attention of healthcare specialists.

Deficits in communication as the earliest indicators of ASD

Our results indicated that early deficits in communication – particularly responding to one's name, lack of gestures, and delayed language, – are all strongly associated with an earlier age at diagnosis. Using a machine learning approach, we show that children whose parents reported the presence of high severity of those signs were diagnosed with ASD 33 months earlier compared to children whose parents either did not observe those signs or observed only their milder presentations.

The finding that early deficits in communication are predictive of early diagnosis is consistent with earlier studies which found that problems with verbal and nonverbal communication, rather than abnormal social interactions or presence of restricted/repetitive behaviors, are key for early detection of ASD (Mandell, Novak and Zubritsky, 2005). Using, to our knowledge, one of the biggest sample to date, we replicated the finding from a number of studies that parental concerns about child's communication are the earliest signs of an underlying ASD (Giacomo and

Fombonne, 1998; Chawarska *et al.*, 2009; Pierce *et al.*, 2011). As we noted earlier, the most likely explanation for this finding is that these symptoms index what parents expect of normal development in the home in the first 2 years of life. If we think of child development not only as a process of biological maturation, but also as a social “career” (Goffman 1961, Eyal *et al.* 2010, pp. 21-22) in the course of which children transition from simpler to more complex social situations, it stands to reason that the factors predictive of early diagnosis would involve the most basic skills of social communication. These signs are likely to trigger a referral to early intervention programs, where such programs exist (recall the finding that early diagnosis is strongly correlated with urban residence). Other signs more strongly indicative of ASD will then become more evident against the background of the type of interactions expected in these programs, in which therapists work one-on-one with children and often involve children in group activities. Given the experience of early intervention workers with many types of disabilities, they are much more likely to attend to these more specific signs and to suggest a referral to diagnosis.

Daniels and Mandell (2014) speculated that the later age of diagnosis in Asperger’s Syndrome vs. autism is due to lack of speech delay in the former, making it less evident until children move into the more socially demanding environment of school. Chawarska and colleagues provide evidence for such heterogeneity in the early course of ASD, and propose that the lack of such communication problems in early development may mask social difficulties until later in development, contributing to a delay in the diagnosis (Chawarska *et al.*, 2014; Kim *et al.*, 2016). Similarly, several studies suggested that restrictive/repetitive behaviors appear to worsen with age (Lord, 1995; Cox *et al.*, 1999; Filipek *et al.*, 1999; Guthrie *et al.*, 2013). Lord (1995) speculates that, since these behaviors are also present in normally developing young children, the worsening may reflect the fact that “parents become better able to recognize the absence of normal development as children grow older” (page 1377). This improved ability owes, no doubt, to a greater range of experiences and comparisons as children begin to exit the home environment, and as normally developing children respond more appropriately to the demands of new environments. Collectively, this evidence suggests that some of the efforts towards reducing the age at diagnosis should better reflect how the disorder manifests differently over time due to both maturational changes, and the transitions in social environments that children typically undergo. The full set of signs required for a formal diagnosis are not likely to emerge until later in development; yet by this time an important window of opportunity for early intervention may have been missed. This argument is supported by studies showing that clinician judgement, rather than use of standardized diagnostic criteria, was associated with an earlier and more stable diagnosis (Lord *et al.*, 2006). It is also in agreement with a tri-level model of autism screening and diagnosis (5), the first level of which consist of generic developmental surveillance as soon and as often as possible and the second which involves referral to early intervention when necessary. Only at the third level, which would typically follow early intervention at a somewhat later age, would the full and formal diagnostic schedule become relevant.

Factors associated with a delayed diagnosis

Our results suggest that temper tantrums and aggression as well as “need for sameness” and sensory hyperreactivity are *positively* correlated with the age of diagnosis; children exhibiting these symptoms are diagnosed later, on average.

Note, moreover, that these clinical signs are relatively uncorrelated with all other signs in Table 4 (apart from “need for sameness”). This suggests that this cluster is associated with a particular ASD subtype, the presentation of which is milder, and hence does not become obvious until the child is older and has transitioned into relatively complex and socially demanding settings. Indeed, the diagnostic delay associated with these symptoms may itself be a cause of those symptoms, which emerge as a consequence of non-recognized ASD-related difficulties.

An alternative explanation is that such clinical signs may mask other ASD-associated signs. For example, an aggressive child would be first treated for aggression and not much else and then, over time, additional diagnoses would be considered. This in turn indicates that children showing excess temper tantrums and/or aggression should be evaluated for autism, rather than awaiting more obviously ASD-linked signs.

Study limitations

One limitation of our study is that the sample is limited to children who were eventually diagnosed with ASD, and hence does not include information on the prevalence and severity of clinical signs among children who were not diagnosed with ASD. As a consequence, we could only establish their connection to age at diagnosis, but not their overall usefulness in predicting which children will go on to develop ASD. Collecting such information would, however, be difficult. The relevant population to sample for signs could include all children, children identified as being “at risk,” or children with a demographic profile similar to that of children who are diagnosed with ASD. Use of each of these comparison populations would likely yield different results.

Given the retrospective nature of our study, the accuracy and quality of responses could diminish with the time elapsed since the child was diagnosed until the survey date. We conducted several statistical analyses to test for such bias (see Appendix B). First, we tested, for each sign, whether the probability that it would be recalled by the parent was influenced by the passage of time; it was not. Second, we conducted a similar test, but for the severity of each sign. Again, reported severity levels were unrelated to time passed since diagnosis. Finally, we included time elapsed since diagnosis as a covariate in our regression analyses. Doing so did not affect our findings.

Another limitation is that questions about signs refer to only one point in time: around the time of the diagnosis. Clinical signs that appear later and might have raised concerns had the child not already been diagnosed, as well as signs that occurred earlier but did not reach the threshold for

parents or other care-givers to delve further or seek help, were not reported, and hence not used in the analysis. In addition, given that the signs and their level of severity are reported by the parents, they are, by definition, subjective and could be biased.

We also cannot infer causation between the sign profile and age of diagnosis, and the direction of many of the effects reported in our study remains to be established. For example, playing alone could *cause* children to be diagnosed later, perhaps because it reduces the salience of other cues, such as delayed speech. However, it is more likely that playing alone is associated with late diagnosis because it becomes apparent only later, when children transition to environments where they are expected to play with other kids.

Even when we suspect that a causal chain does exist, identification of the exact pathway of causation cannot be established. It is likely, for example, that delayed speech contributes to early diagnosis not because parents identify it as a cause for concern about ASD, but because it leads to referral to speech experts, who are more attune to the possibility of ASD as well as familiar with other patterns of signs that predict it.

Finally, similar to most studies of children with ASD, the indices of socioeconomic status were higher in our sample than in the general US population. If this discrepancy results, in part, from a greater willingness to participate in surveys on the part of high SES individuals, then this could affect the generalization of our findings to the broad population. However, the general findings of our study should still hold unless the path to diagnosis is systematically different for lower SES families.

Conclusion

Utilizing a large survey of parents of children with ASD, this is the first study that systematically looks at a large set of clinical signs and symptoms, their presence and level of severity, the correlations between them, and how they are related to the age of diagnosis. We show that individual signs play an important role in predicting age of diagnosis beyond what is captured by the overall severity of the child condition.

Since many signs are highly correlated, we show that most of the variation in age of diagnosis can be captured by a small number of signs. Using regression tree, we can both rank-order the signs and show how they interact with each other to provide an earlier diagnosis.

While some of our findings are consistent with practitioners' experience, this study provides statistical support for existing intuitions, as well as new insights. A key new insight is that clinical signs not specifically associated with ASD can lead to earlier diagnosis if they bring the child to the attention of professionals who have greater experience with, and hence greater skill in diagnosing, ASD. Another key insight is that careful attention should be paid to children showing excessive tantrums or aggression, as these behaviors may interfere with an early ASD

diagnoses. Similarly, in the subset of children without early deficits in communication, diagnosis is delayed, and this might be improved if more attention will be given to clinical signs that are not necessarily considered as ASD symptoms.

References

- Anderson, D. K., Liang, J. W. and Lord, C. (2014) 'Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 55(5), pp. 485–494. doi: 10.1111/jcpp.12178.
- Bradshaw, J. *et al.* (2014) 'Feasibility and Effectiveness of Very Early Intervention for Infants At-Risk for Autism Spectrum Disorder: A Systematic Review.', *Journal of autism and developmental disorders*. doi: 10.1007/s10803-014-2235-2.
- Chawarska, K. *et al.* (2009) 'A prospective study of toddlers with ASD: Short-term diagnostic and cognitive outcomes', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 50(10), pp. 1235–1245. doi: 10.1111/j.1469-7610.2009.02101.x.
- Chawarska, K. *et al.* (2014) '18-Month Predictors of Later Outcomes in Younger Siblings of Children With Autism Spectrum Disorder: a Baby Siblings Research Consortium Study.', *Journal of the American Academy of Child and Adolescent Psychiatry*. Elsevier Inc, 53(12), p. 1317–1327.e1. doi: 10.1016/j.jaac.2014.09.015.
- Christensen, D. *et al.* (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 Surveill Summ*, 65(No. SS-3), p. 1023. doi: <http://dx.doi.org/10.15585/mmwr.ss6503a1>.
- Coughlin, S. S. (1990) 'Recall bias in epidemiologic studies.', *Journal of clinical epidemiology*, 43(1), pp. 87–91. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2319285> (Accessed: 11 September 2018).
- Cox, A. *et al.* (1999) 'Autism Spectrum Disorders at 20 and 42 Months of Age : Stability of Clinical and ADI-R Diagnosis', *Journal of Child Psychology and Psychiatry*, 40(5), pp. 719–732.
- Daniels, A. M. *et al.* (2012) 'Verification of Parent-Report of Child Autism Spectrum Disorder Diagnosis to a Web-Based Autism Registry', *Journal of Autism and Developmental Disorders*, 42(2), pp. 257–265. doi: 10.1007/s10803-011-1236-7.
- Daniels, A. M. and Mandell, D. S. (2014) 'Explaining differences in age at autism spectrum disorder diagnosis: A critical review', *Autism*, 18(5), pp. 583–597. doi: 10.1177/1362361313480277.
- Dawson, G. (2008) 'Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder.', *Development and psychopathology*, 20(3), pp. 775–803. doi: 10.1017/S0954579408000370.
- Eyal, G. *et al.* (2010) *The Autism Matrix*. Cambridge, MA: Policy Press.
- Filipek, P. A. *et al.* (1999) 'The Screening and Diagnosis of Autistic Spectrum Disorders', *Journal of Autism & Developmental Disorders*, 29(6), p. 439. doi: 10.1023/A:1021943802493.
- Fountain, C., King, M. D. and Bearman, P. S. (2011) 'Age of diagnosis for autism: individual and community factors across 10 birth cohorts.', *Journal of epidemiology and community health*,

65(6), pp. 503–510. doi: 10.1136/jech.2009.104588.

Frith, U. and Soares, I. (1993) 'Research into earliest detectable signs of autism: what parents say', *Communication*, 27, pp. 17–18.

Giacomo, A. De and Fombonne, E. (1998) 'Parental recognition of developmental abnormalities in autism.', *European child & adolescent psychiatry*, 136, pp. 131–136.

Goffman, E. (1961) *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*. 1st edn. Garden City, NY: Anchor Books.

Gray, D. (2002) 'Ten years on: A longitudinal study of families of children with autism', *Journal of Intellectual and Developmental Disability*, 27(3), pp. 215–222. doi: 10.1080/136682502100000863.

Guthrie, W. *et al.* (2013) 'Early diagnosis of autism spectrum disorder: Stability and change in clinical diagnosis and symptom presentation', *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 54(5), pp. 582–590. doi: 10.1111/jcpp.12008.

Howlin, P. (1999) *Children with autism and Asperger syndrome: A guide for practitioners and carers*. Wiley.

Howlin, P. and Asgharian, a (1999) 'The diagnosis of autism and Asperger syndrome: findings from a survey of 770 families.', *Developmental medicine and child neurology*, 41(12), pp. 834–839. doi: 10.1111/j.1469-8749.1999.tb00550.x.

Kaiser, H. F. and Rice, J. (1974) 'Little Jiffy, Mark IV', *Educational and Psychological Measurement*, (1), pp. 111–117.

Kim, S. H. *et al.* (2016) 'Examining the phenotypic heterogeneity of early autism spectrum disorder: subtypes and short-term outcomes', *Journal of Child Psychology and Psychiatry*, 57(1), pp. 93–102. doi: 10.1111/jcpp.12448.

Lee, H. *et al.* (2010) 'Accuracy of phenotyping of autistic children based on internet implemented parent report', *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 153B(6), p. n/a-n/a. doi: 10.1002/ajmg.b.31103.

Lord, C. (1995) 'Follow-up of two-year-olds referred for possible autism.', *Journal of Child Psychology & Psychiatry*, 36(8), pp. 1365–82. doi: 10.1111/j.1469-7610.1995.tb01669.x.

Lord, C. *et al.* (2006) 'Autism From 2 to 9 Years of Age', *Archives of General Psychiatry*, 63(6), p. 694. doi: 10.1001/archpsyc.63.6.694.

Mandell, D. *et al.* (2002) 'Race Differences in the Age at Diagnosis Among Medicaid-Eligible Children With Autism', *Journal of the American Academy of Child & Adolescent Psychiatry*. The American Academy of Child and Adolescent Psychiatry, 41(12), pp. 1447–1453. doi: 10.1097/00004583-200212000-00016.

Mandell, D. S. *et al.* (2009) 'Racial/ethnic disparities in the identification of children with autism spectrum disorders', *American Journal of Public Health*, 99(3), pp. 493–498. doi: 10.2105/AJPH.2007.131243.

Mandell, D. S., Novak, M. and Zubritsky, C. (2005) 'Factors Associated With Age of Diagnosis

Among Children With Autism Spectrum Disorders', *Pediatrics*, 116(6), pp. 1480–1486. doi: 10.1542/peds.2005-0185.

Mandell, D. S. and Palmer, R. (2005) 'Differences Among States in the Identification of Autistic Spectrum Disorders', *Archives of pediatrics & adolescent medicine*, 159, pp. 4–7.

Marvin, A.R., Law, P.A., Law, J.K., Arthur, E.M., Mortenson, E.L., Abbacchi, A.M., Watson, T.T., Westreich, A.A., T. Gray, T., Zhang, Y., Marvin, D. J., Levin, S.N., & Constantino, J. . (2014) *Non-Verbal Children with ASD (NV-ASD): Validating a Registry and Characterizing a Population*. Atlanta, GA: International Society for Autism Research. Available at: <https://imfar.confex.com/imfar/2014/webprogram/Paper16978.html>.

McPheeters, M. L. *et al.* (2016) 'Screening for Autism Spectrum Disorder in Young Children: A Systematic Evidence Review for the U.S. Preventive Services Task Force', *AHRQ Publication*, 13–05185–E(129), p. 202.

Palmer, R. F. *et al.* (2005) 'School district resources and identification of children with autistic disorder', *American Journal of Public Health*, 95(1), pp. 125–130. doi: 10.2105/AJPH.2003.023077.

Pierce, K. *et al.* (2011) 'Detecting, studying, and treating autism early: the one-year well-baby check-up approach', *The Journal of Pediatrics*, 159(3), pp. 458–465. doi: 10.1016/j.jpeds.2011.02.036.Detecting.

Pierce, K., Courchesne, E. and Bacon, E. (2016) 'To Screen or Not to Screen Universally for Autism is not the Question: Why the Task Force Got It Wrong', *Journal of Pediatrics*. Elsevier, 176, pp. 182–194. doi: 10.1016/j.jpeds.2016.06.004.

Pisula, E. (2003) 'Parents of children with autism: review of current research.', *Archives of Psychiatry and Psychotherapy*, 5.4, pp. 51–63.

Shattuck, P. T. *et al.* (2009) 'Timing of Identification Among Children With an Autism Spectrum Disorder: Findings From a Population-Based Surveillance Study', *Journal of the American Academy of Child & Adolescent Psychiatry*. American Academy of Child and Adolescent Psychiatry, 48(5), pp. 474–483. doi: 10.1097/CHI.0b013e31819b3848.

Sicherman, N. *et al.* (2018) 'Grandma knows best: Family structure and age of diagnosis of autism spectrum disorder', *Autism*. SAGE PublicationsSage UK: London, England, 22(3), pp. 368–376. doi: 10.1177/1362361316679632.

Siu, A. L. *et al.* (2016) 'Screening for Autism Spectrum Disorder in Young Children', *JAMA : the journal of the American Medical Association*, 315(7), p. 691. doi: 10.1001/jama.2016.0018.

Sivberg, B. (2002) 'Family System and Coping Behaviors', *Autism*, 6(4), pp. 397–409. doi: 10.1177/1362361302006004006.

Therneau, T. M. and Atkinson, E. J. (2017) *An Introduction to Recursive Partitioning Using the RPART Routines*.

Warren, Z. *et al.* (2011) 'A Systematic Review of Early Intensive Intervention for Autism Spectrum Disorders', *Pediatrics*, 127(5), pp. e1303–e1311. doi: 10.1542/peds.2011-0426.

Watson, L. R. *et al.* (2007) 'The First Year Inventory: Retrospective parent responses to a questionnaire designed to identify one-year-olds at risk for autism', *Journal of Autism and Developmental Disorders*, 37(1), pp. 49–61. doi: 10.1007/s10803-006-0334-4.

Wiggins, L. D., Baio, J. and Rice, C. (2006) 'Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample.', *Journal of developmental and behavioral pediatrics*, 27(2), pp. 79–87. doi: 10.1097/00004703-200604002-00005.

Wilcox, A. J. and Horney, L. F. (1984) 'Accuracy of spontaneous abortion recall.', *American journal of epidemiology*, 120(5), pp. 727–33. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6541871> (Accessed: 11 September 2018).

Zuckerman, K. E., Lindly, O. J. and Sinche, B. K. (2015) 'Parental concerns, provider response, and timeliness of autism spectrum disorder diagnosis', *Journal of Pediatrics*. Elsevier Inc, 166(6), p. 1431–1439.e1. doi: 10.1016/j.jpeds.2015.03.007.

Table 1. Sample Socio-Economic Characteristics
Compared with the US Population

A. Race and ethnicity

	Freq.	Percent	US Ages 0-25	US Total
White only	1273	77.7%	55.6%	64.3%
Hispanic	180	10.98%	20.8%	15.6%
Other/Mixed	299	18.2%	4.4%	2.9%
Asian only	27	1.7%	4.3%	4.5%
Black only	36	2.2%	14.9%	12.7%

B. Parents' education

			U.S. (25+) Adults	
	Father	Mother	Men	Women
Less than HS	4.5%	0.9%	12.7%	12.0%
HS Grad	16.5%	6.5%	30.8%	30.0%
Some college	27.3%	27.8%	25.1%	27.5%
College grad	25.2%	33.5%	19.8%	19.9%
More than college	26.2%	31.2%	11.6%	10.7%

C. Household income

	Freq.	Percent	US Total
Below \$25,000	162	10.5%	25.1%
\$25,000-\$49,999	393	25.4%	24.9%
\$50,000-\$99,999	585	37.1%	29.1%
\$100,000-\$199,999	312	20.2%	16.8%
\$200,000 or higher	76	4.9%	4.2%

D. Level of Urbanization

	Freq.	Percent	US (2010)
Urban	1392	87.6%	80.70%
Rural, Large	115	7.2%	19.30%
Rural, Small	47	3.0%	
Rural, Isolated	36	2.3%	

Table 2: Reported Clinical Signs & Symptoms and Level of Severity

		Level of Severity on a 0-10 scale			
Clinical Signs & Symptom	Obs.	Mean	Median	Std.	% Non Zero
Social and Communication					
Delayed speech	1609	6.18	7.40	3.58	0.83
Delayed response to name	1609	4.66	4.90	3.51	0.77
Poor eye contact	1609	6.07	6.30	2.99	0.91
Lack of gestures (e.g., pointing, nodding or shaking head)	1609	5.23	5.30	3.54	0.82
Difficulty understanding gestures	1609	5.08	5.20	3.38	0.83
Preference to play alone or play with objects rather than with others	1609	7.10	7.80	2.90	0.93
More focus on objects than people	1609	6.79	7.50	2.96	0.92
Failure to initiate or respond to social interactions	1609	6.94	7.70	2.80	0.94
Difficulties in initiating and/or maintaining relationships and friendships	1609	7.48	8.30	2.81	0.94
Restricted and Repetitive behaviors					
Played with toys or objects in an unusual way (e.g. repetitive play, lining up toys)	1609	6.19	6.60	3.02	0.92
Need for sameness (e.g. difficulties with changes in routine)	1330	6.30	6.90	3.01	0.76
Unusual motor mannerisms (e.g. hand flapping, spinning)	1330	5.29	5.40	3.41	0.71
Unusual interest in specific objects or toys (e.g. high in intensity or focus)	1330	6.45	7.10	3.00	0.75
Sensory Symptoms					
Unusual interest in sensory aspects of the environment (e.g. excessive smelling of objects or people, fascination with lights or movement)	1609	4.77	4.90	3.16	0.86
Sensory hyperreactivity (e.g. excessive or adverse response to specific sounds, lights, touch, smell or tastes)	1609	6.16	6.50	3.01	0.93
Sensory hyporeactivity (e.g. insensitivity or indifference to sensory pain or temperature, slow response to sensory stimuli in the environment)	1609	4.81	5.00	3.39	0.83
Aggression					
Temper tantrums	1609	4.85	5.00	3.20	0.85
Aggression toward self	1609	2.48	1.30	3.01	0.60
Aggression toward others	1609	2.65	1.80	3.01	0.64
Regression (At some point in time, did x's behavior regress (deteriorate) in any of the following ways?)					
Loss of skills	1609	2.31	0.10	3.12	0.50
Loss of language (words only)	1609	2.68	0.10	3.60	0.49
Loss of language (phrases)	1609	2.16	0.00	3.46	0.41
Less social engagement	1609	2.96	1.20	3.49	0.56
Loss of motor skills	1609	1.18	0.00	2.36	0.35
Loss of daily living skills	1609	1.56	0.00	2.80	0.38

The smaller number of observations for three of the clinical signs was the result of a technical problem that resulted in a loss of data.

Figure 1: Distribution of Clinical Signs & Symptoms by Levels of Severity

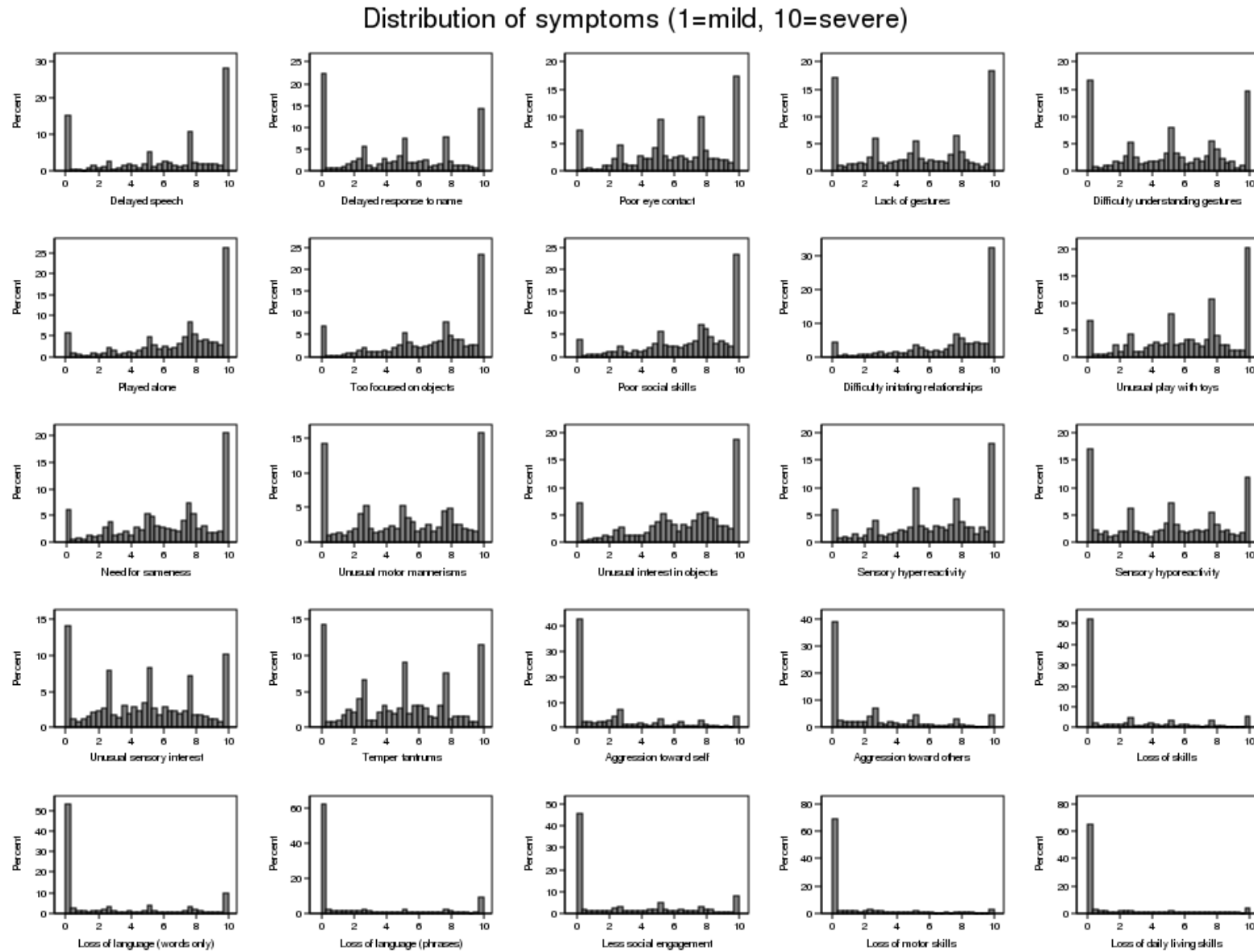


Figure 2: Correlation Between Clinical Signs & Symptoms Severity and Age of Diagnosis

Note: “Negative” months indicate earlier diagnosis

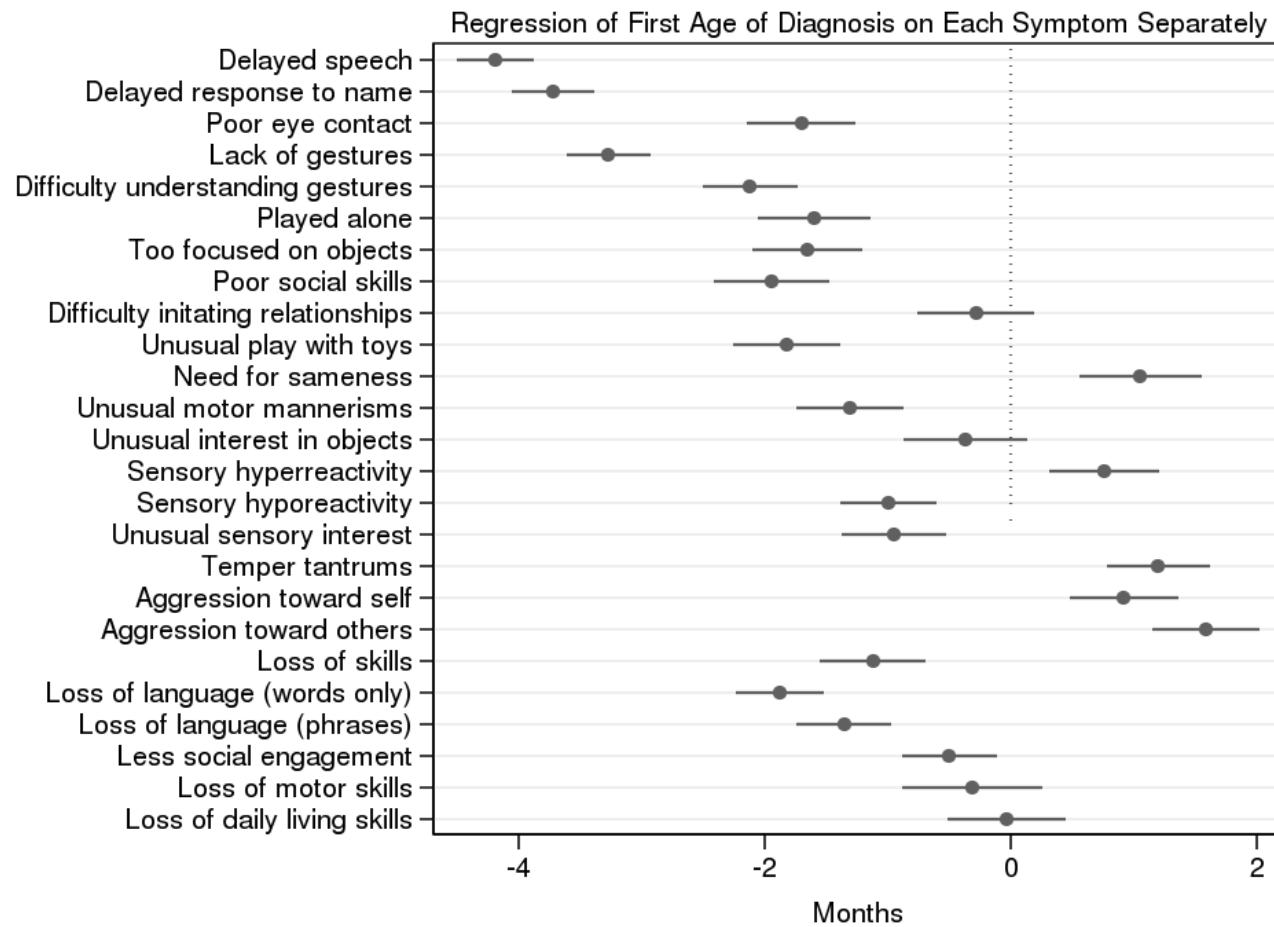


Table 3: Pairwise Correlations Between Clinical Signs & Symptoms

(those highlighted in red are significant at the .05 level).

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15	s16	s17	s18	s19	s20	s21	s22	s23	s24	s25
Delayed speech	1.00																								
Delay resp to name	0.64	1.00																							
Poor eye contact	0.38	0.52	1.00																						
Lack of gestures	0.58	0.66	0.50	1.00																					
Difficulty understanding gestures	0.48	0.61	0.52	0.72	1.00																				
Played alone	0.32	0.38	0.49	0.41	0.50	1.00																			
Too focused on objects	0.32	0.42	0.50	0.42	0.50	0.74	1.00																		
Poor social skills	0.35	0.42	0.49	0.43	0.49	0.61	0.65	1.00																	
Difficulty initiating relationships	0.22	0.29	0.37	0.30	0.42	0.54	0.54	0.63	1.00																
Unusual play with toys	0.35	0.32	0.34	0.31	0.34	0.48	0.55	0.43	0.42	1.00															
Need for sameness	0.07	0.07	0.22	0.06	0.15	0.27	0.33	0.28	0.34	0.43	1.00														
Unusual motor mannerisms	0.29	0.30	0.29	0.29	0.34	0.34	0.36	0.32	0.32	0.42	0.33	1.00													
Unusual interest in objects	0.16	0.20	0.29	0.17	0.25	0.40	0.50	0.33	0.35	0.63	0.52	0.49	1.00												
Sensory hyperreactivity	0.05	0.09	0.22	0.09	0.16	0.22	0.28	0.22	0.27	0.32	0.46	0.33	0.40	1.00											
Sensory hyporeactivity	0.22	0.29	0.25	0.28	0.34	0.29	0.36	0.31	0.30	0.30	0.28	0.32	0.33	0.34	1.00										
Unusual sensory interest	0.19	0.26	0.23	0.24	0.29	0.30	0.35	0.32	0.33	0.42	0.36	0.46	0.45	0.47	0.46	1.00									
Temper tantrums	0.01	0.03	0.14	0.01	0.12	0.15	0.16	0.15	0.27	0.20	0.46	0.15	0.28	0.37	0.23	0.26	1.00								
Aggression toward self	0.05	0.09	0.13	0.09	0.18	0.19	0.20	0.15	0.22	0.20	0.34	0.22	0.23	0.29	0.27	0.24	0.57	1.00							
Aggression toward others	-0.06	-0.01	0.07	0.00	0.10	0.07	0.10	0.09	0.22	0.13	0.33	0.12	0.20	0.27	0.20	0.19	0.63	0.54	1.00						
Loss of skills	0.27	0.29	0.24	0.28	0.29	0.23	0.25	0.22	0.20	0.24	0.15	0.27	0.18	0.18	0.24	0.22	0.16	0.20	0.13	1.00					
Loss of language (words only)	0.39	0.34	0.24	0.31	0.27	0.23	0.22	0.23	0.16	0.21	0.09	0.21	0.13	0.10	0.18	0.16	0.08	0.12	0.04	0.71	1.00				
Loss of language (phrases)	0.32	0.30	0.22	0.25	0.24	0.24	0.22	0.21	0.17	0.20	0.12	0.19	0.14	0.11	0.17	0.17	0.11	0.14	0.09	0.68	0.82	1.00			
Less social engagement	0.19	0.23	0.24	0.23	0.25	0.24	0.25	0.23	0.21	0.18	0.11	0.18	0.14	0.14	0.21	0.16	0.14	0.19	0.13	0.71	0.65	0.65	1.00		
Loss of motor skills	0.14	0.16	0.19	0.16	0.19	0.16	0.19	0.15	0.17	0.16	0.15	0.21	0.17	0.17	0.21	0.17	0.14	0.22	0.17	0.61	0.46	0.49	0.53	1.00	
Loss of daily living skills	0.14	0.17	0.17	0.15	0.18	0.16	0.20	0.14	0.19	0.18	0.20	0.23	0.20	0.21	0.27	0.22	0.22	0.27	0.23	0.58	0.46	0.52	0.53	0.61	1.00

Table 4: Factor Loadings, 5 Factors, Orthogonal Varimax Rotation, loading>.3

Variable		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
		Regressive Autism	Social Awkwardness	Communication difficulties	Sensory reactivity and need for sameness	Aggressive Behavior
1	Delayed speech			0.6610		
2	Delayed response to name			0.7492		
3	Poor eye contact		0.4324	0.4654		
4	Lack of gestures			0.7728		
5	Difficulty understanding gestures		0.3613	0.6916		
6	Played alone		0.7254			
7	Too focused on objects		0.7400			
8	Poor social skills		0.6906	0.3184		
9	Difficulty initiating relationships		0.6150			
10	Unusual play with toys		0.4290		0.3545	
11	Need for sameness				0.4949	0.3918
12	Unusual motor mannerisms				0.5210	
13	Unusual interest in objects				0.6638	
14	Sensory hyperreactivity				0.4932	0.3374
15	Sensory hyporeactivity				0.3659	
16	Unusual sensory interest				0.5741	
17	Temper tantrums					0.7523
18	Aggression toward self					0.6514
19	Aggression toward others					0.7224
20	Loss of skills	0.8197				
21	Loss of language (words only)	0.8146				
22	Loss of language (phrases)	0.8222				
23	Less social engagement	0.7676				
24	Loss of motor skills	0.6562				
25	Loss of daily living skills	0.6496				

See Table 2 for the full labels of symptoms.

**Table 5: Regression Models for the effects of factor scores on age of diagnosis
using 5 factor Varimax Rotation Model**

	(1)	(2)	(3)	(4)	(5)
Factor 1: Regressive Autism	-2.43*** (0.66)	-2.01*** (0.65)		-1.77** (0.79)	-1.64** (0.78)
Factor 2: Social Awkwardness	0.53 (0.73)	-0.90 (0.72)		0.88 (0.76)	-0.69 (0.76)
Factor 3: Communication Difficulties	-15.7*** (0.70)	-13.5*** (0.74)		-15.3*** (0.77)	-13.2*** (0.79)
Factor 4: Sensory Reactivity & Need for Sameness	-1.97*** (0.76)	-2.55*** (0.74)		-1.61** (0.80)	-2.34*** (0.77)
Factor 5: Aggressive Behavior	7.46*** (0.73)	6.25*** (0.73)		7.81*** (0.76)	6.46*** (0.76)
Both Parents Graduated College		-4.04** (1.60)			-4.07** (1.60)
Only Mother Graduated College		0.52 (1.78)			0.48 (1.78)
Only Father Graduated College		-3.65 (2.45)			-3.61 (2.45)
Black		1.49 (2.97)			1.42 (2.97)
Hispanic/Latino		3.20 (1.97)			3.24 (1.97)
Other		3.02 (3.46)			3.06 (3.46)
Family Income > \$100,000		-1.45 (1.54)			-1.48 (1.54)
Lived in Urban Area		-2.76 (1.79)			-2.76 (1.79)
Asperger Diagnosis		10.9*** (1.84)			10.8*** (1.85)
Year of Birth		-0.75*** (0.10)			-0.75*** (0.10)
Number of Symptoms Reported			-0.97*** (0.14)	-0.31 (0.21)	-0.18 (0.21)
Constant	46.7*** (0.62)	1,549*** (202)	64.3*** (2.63)	52.7*** (4.12)	1,553*** (202)
Observations	1,359	1,330	1,669	1,359	1,330
R-squared	0.327	0.387	0.030	0.328	0.387

Each column reports the estimated coefficients of one regression

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

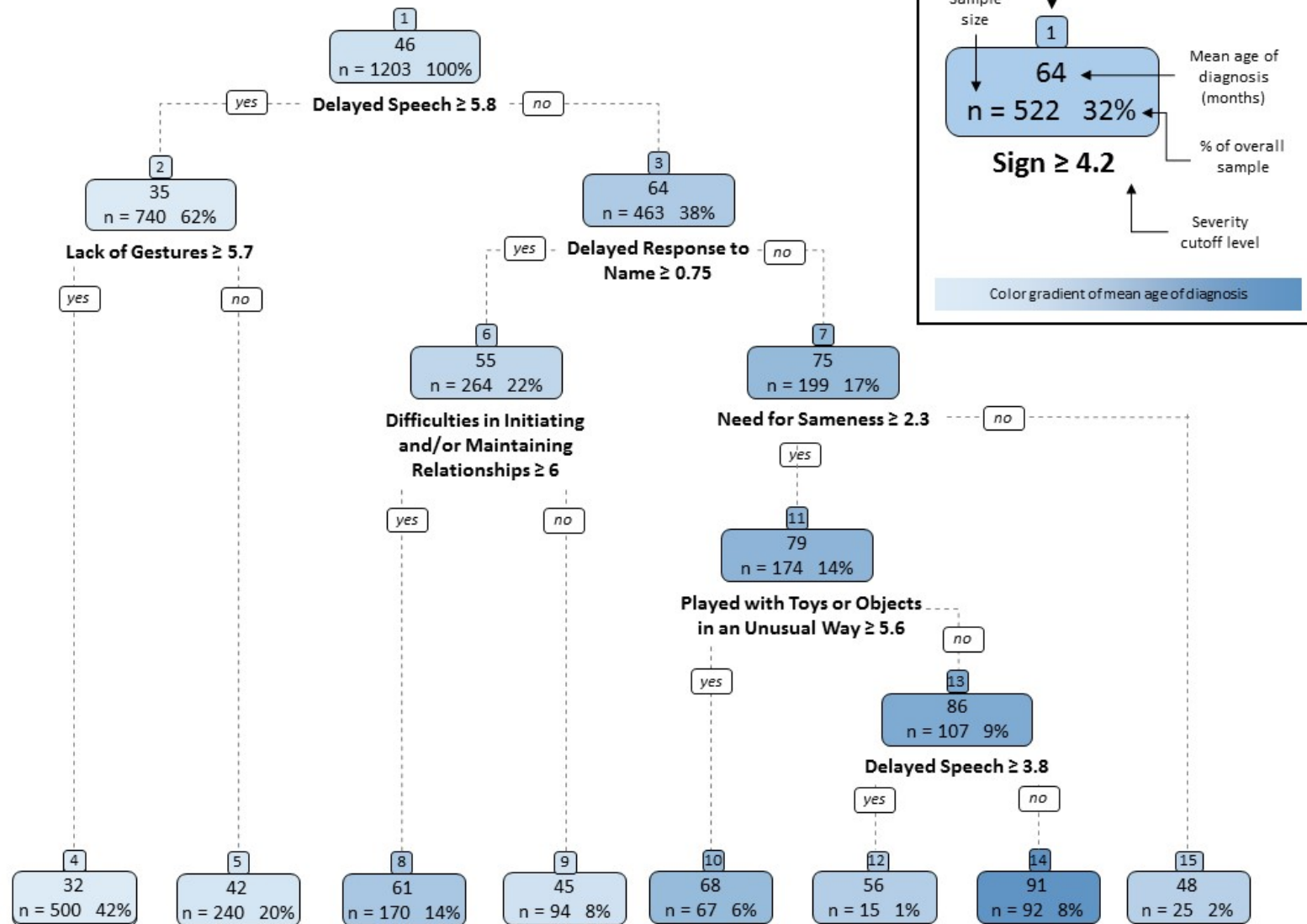
Table 6: Optimal Splits and Mean/Median Age at First Diagnosis Above/Below Split

Clinical Signs & Symptoms	Split	Mean Above	Mean Below	p-value*	Median Above	Median Below
Delayed speech	5.75	34.9	63.7	0.00000	30	58
Delayed response to own name	3.85	35.6	60.4	0.00000	32	52
Poor eye contact	6.15	40.5	52.1	0.00000	33	42
Lack of gestures	6.05	33.6	56.7	0.00000	30	48
Difficulty understanding gestures	3.75	41.1	54.9	0.00000	34	47
Preference to play alone	7.55	42.3	50.7	0.00000	36	40
Too focused on objects	1.05	44.8	61.6	0.00000	36	54
Poor social skills	5.25	43.2	53.9	0.00000	36	42
Unusual play with toys or objects	5.55	41.3	52.7	0.00000	35	42
Need for sameness	2.35	47.4	35.1	0.00000	38	31
Unusual motor mannerisms	3.15	42.6	52.7	0.00000	35	40
Unusual sensory interest	1.65	47.3	35.2	0.00000	38	28
Sensory hyperreactivity	8.55	39.4	47.3	0.00019	33	38
Sensory hyporeactivity	1.75	44.5	51.2	0.00042	36	39
Temper tantrums	2.35	48.2	39.6	0.00000	38	32
Aggression toward self	2.65	51.4	43.3	0.00000	40	36
Aggression toward others	2.15	50.8	41.9	0.00000	41	34
Loss of skills	5.15	39.8	47.5	0.00011	35	38
Loss of language (words only)	5.8	34.8	49.3	0.00000	30	39
Loss of language (phrases)	7.25	35.7	48.0	0.00000	32	38
Difficulty initiating or maintaining relationships	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."
Unusual interest in objects or toys	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."
Less social engagement	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."
Loss of motor skills	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."
Loss of daily living skills	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."	"n.a."

* The p-value is for the difference between the means

Figure 3: Regression Tree of Age of Diagnosis and Clinical Signs & Symptoms

Age at First Diagnosis and Clinical Signs



Appendix Table 1: The Effects of Various Variables on Age of Diagnosis

	(1)	(2)	(3)	(4)	(5)
<u>Parents Education (excluded: Neither parent graduated college)</u>					
Both Parents Graduated College	-5.19*** (1.48)				-3.97** (1.59)
Only Mother Graduated College	0.42 (1.78)				0.99 (1.79)
Only Father Graduated College	-2.00 (2.50)				-2.49 (2.54)
<u>Race/Ethnicity (excluded: White)</u>					
Black		0.64 (3.00)			-0.056 (3.02)
Hispanic/Latino		1.97 (2.00)			1.67 (2.01)
Other		5.11 (3.42)			4.76 (3.42)
Family Income > \$100,000			-4.36*** (1.46)		-2.25 (1.55)
Lived in Urban Area				-5.86*** (1.85)	-5.33*** (1.86)
Asperger Diagnosis	25.5*** (1.72)	25.5*** (1.72)	25.7*** (1.72)	25.5*** (1.72)	25.9*** (1.72)
Year of Birth	-0.87*** (0.10)	-0.87*** (0.10)	-0.82*** (0.10)	-0.86*** (0.10)	-0.85*** (0.10)
Constant	1,793*** (201)	1,786*** (203)	1,694*** (204)	1,766*** (202)	1,745*** (205)
Observations	1,669	1,669	1,669	1,631	1,631
R-squared	0.181	0.174	0.177	0.182	0.192

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Each column reports the estimation results of one regression. The dependent variable in all 5 regressions is the age of diagnosis (in months). The independent variables vary across regressions and are listed in the first column. For example, parents' education dummies are included in regressions (1) and (5) only.

Appendix A: Regression Analyses

In Appendix table A1 we report the estimation results of various regression analyses where the dependent variable is the age of diagnosis (in months).

Column 1 reports the estimation results of 25 separate regressions. Each cell reports the estimated coefficient and standard errors used to draw the point estimates and confidence intervals in Figure 2 in the paper. Each coefficient measures the average change in the age of diagnosis (in months) for a one-unit change in the severity of the clinical signs and symptom. For example, a one unit increase in the level of severity of delayed speech is associated with, on average, about four months decline in the age of diagnosis.

[PLACE APPENDIX TABLE A1 HERE]

In Appendix Table A1, column 2, we report the estimation result of a multiple regression model. Unlike the first column, here we report the estimated coefficients of a single regression, where the partial effects of all clinical signs and symptoms on the age of diagnosis are estimated simultaneously.

Comparing the results in column 2 to those in column 1, several observations can be made. For many symptoms, while the sign of the coefficient did not change, the effect became weaker and sometimes less significant as it was in the case when it was the only variable in the regression. This is natural, given that multiple regression looks for a symptom's unique predictive power. In two cases there was a dramatic change in the coefficient. In the univariate regressions, "poor eye contact" and "difficulty understanding gestures" were each negatively (and significantly different from zero) correlated with the age of diagnosis. However, when including all other clinical signs and symptoms in the regressions (i.e., holding other signs symptoms constant), higher severity levels of these two symptoms significantly *increase* the age of diagnosis.

[PLACE APPENDIX FIGURE A1 HERE]

In order to account for other factors that are likely to affect the age of diagnosis, the analyses reported in Appendix Table A1, columns 1 and 2, were repeated by adding the following covariates: parents' education, race/ethnicity, family income (if over \$100,000), residence location (if urban), an Asperger diagnosis, and year of birth.

In Appendix Table 1 we report the effects of these variables on the age of diagnosis, both separately and jointly. Our findings, which are similar to those reported in other studies, show that parents' education and urban residence, are the two primary facilitators of early diagnosis. They neatly capture the interaction between parent-early intervention and infrastructure that is necessary for early diagnosis. The improvement in diagnosis over time is reflected in the "year of birth" coefficient, showing that, in our sample, for every year that passed, the average age of diagnosis went down by 0.85 months.

Comparing the results in columns 3 and 4 to those in 1 and 2, respectively, show that the inclusion of these additional covariates, while affecting the point estimates, did not change the general findings.

Appendix Table A1: Regressions of First Age of Diagnosis on Symptom Severity

	(1)	(2)	(3)	(4)
Clinical Signs & Symptoms	Separate	Together	Separate	Together
Delayed speech	-4.173*** (0.159)	-2.550*** (0.231)	-3.709*** (0.174)	-2.314*** (0.241)
Delay resp to name	-3.707*** (0.169)	-1.886*** (0.259)	-2.993*** (0.173)	-1.535*** (0.260)
Poor eye contact	-1.637*** (0.222)	0.769*** (0.261)	-1.469*** (0.205)	0.613** (0.257)
Lack of gestures	-3.252*** (0.173)	-1.632*** (0.277)	-2.598*** (0.170)	-1.423*** (0.275)
Difficulty understanding gestures	-2.086*** (0.194)	1.409*** (0.281)	-1.722*** (0.185)	1.185*** (0.280)
Played alone	-1.563*** (0.232)	-0.0932 (0.323)	-1.611*** (0.215)	-0.233 (0.316)
Too focused on objects	-1.609*** (0.227)	-0.404 (0.338)	-1.595*** (0.212)	-0.382 (0.332)
Poor social skills	-1.880*** (0.239)	-0.462 (0.318)	-1.614*** (0.222)	-0.326 (0.315)
Difficulty initiating relationships	-0.267 (0.241)	1.229*** (0.285)	-0.551** (0.225)	0.952*** (0.282)
Unusual play with toys	-1.801*** (0.219)	-0.789*** (0.279)	-1.867*** (0.205)	-0.971*** (0.276)
Need for sameness	1.095*** (0.251)	1.011*** (0.260)	0.518** (0.239)	0.868*** (0.258)
Unusual motor mannerisms	-1.324*** (0.219)	-0.561*** (0.216)	-1.211*** (0.207)	-0.512** (0.215)
Unusual interest in objects	-0.390 (0.251)	0.113 (0.287)	-0.758*** (0.234)	0.104 (0.283)
Sensory hyperreactivity	0.820*** (0.223)	0.618** (0.241)	0.378* (0.209)	0.559** (0.238)
Sensory hyporeactivity	-1.065*** (0.198)	-0.280 (0.207)	-1.047*** (0.184)	-0.410** (0.203)
Unusual sensory interest	-0.871*** (0.213)	-0.0802 (0.242)	-0.850*** (0.197)	-0.0289 (0.239)
Temper tantrums	1.240*** (0.210)	-0.0449 (0.264)	0.825*** (0.196)	0.0405 (0.262)
Aggression toward self	0.875*** (0.224)	0.583** (0.253)	0.548** (0.215)	0.459* (0.255)
Aggression toward others	1.570*** (0.221)	0.166 (0.261)	1.126*** (0.206)	0.152 (0.256)
Loss of skills	-1.081*** (0.216)	0.438 (0.325)	-0.628*** (0.204)	0.549* (0.320)
Loss of language (words only)	-1.875*** (0.183)	-0.775** (0.311)	-1.367*** (0.174)	-0.678** (0.306)
Loss of language (phrases)	-1.367*** (0.193)	-0.0991 (0.308)	-1.008*** (0.181)	-0.303 (0.305)
Less social engagement	-0.438** (0.195)	0.290 (0.260)	-0.215 (0.181)	0.234 (0.257)
Loss of motor skills	-0.246 (0.286)	-0.0634 (0.345)	0.00384 (0.264)	0.0311 (0.339)
Loss of daily living skills	0.0169 (0.240)	0.663** (0.291)	0.214 (0.222)	0.703** (0.287)
Observations		1360		1331
Adjusted R-squared		0.410		0.444

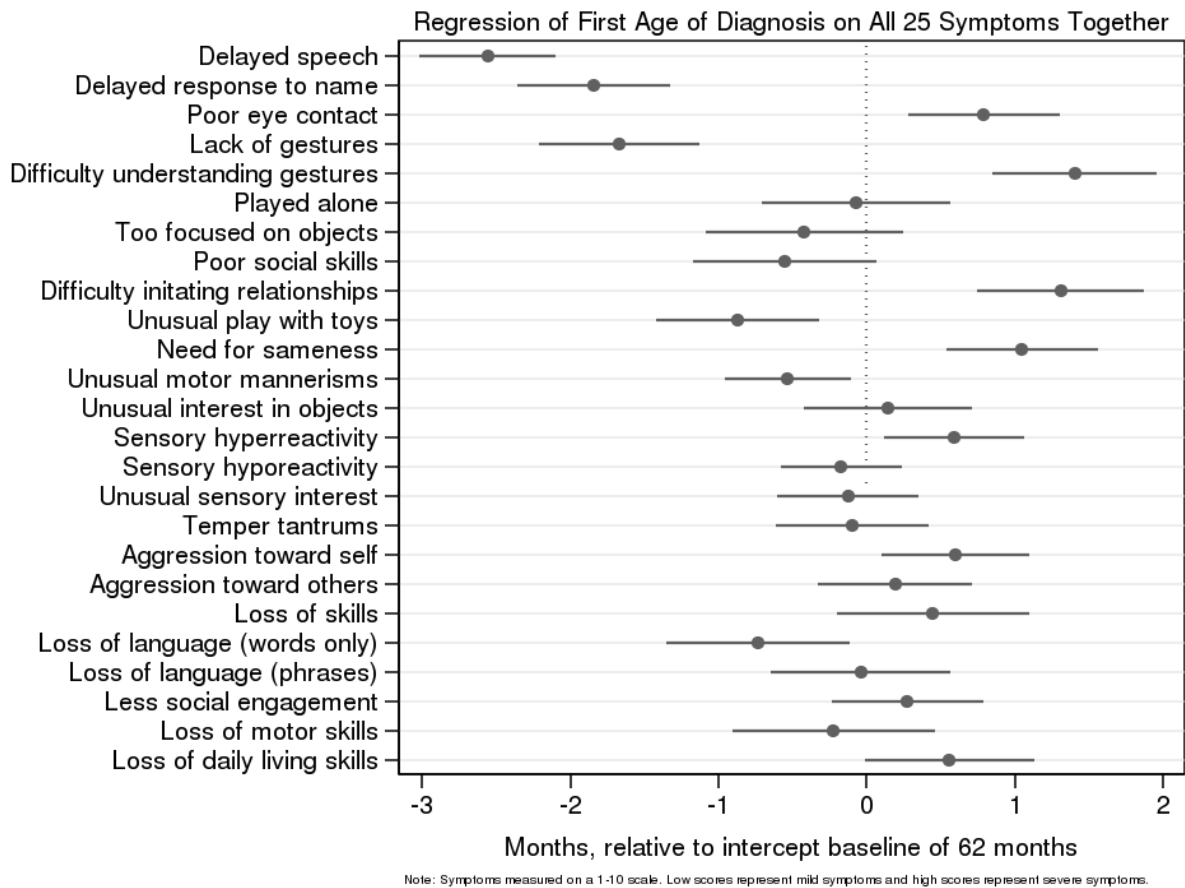
Models (3) & (4) include, as covariates, parents education, race, income, urban, Asperger, and year of birth
Standard errors in parentheses.

*p<0.10, **p<0.05, ***p<0.01

In Columns (1) & (3) each cell reports the results of a separate regression estimating the effect of one symptom on age of diagnosis

In Columns (2) & (4) each column reports the estimation result of one regression where all symptoms are included

Appendix Figure A1: The Effect of all Clinical Signs & Symptoms (Combined) on Age of Diagnosis



Appendix B: Testing for potential recall-bias

Given the retrospective nature of our survey data, it is possible that parents' ability to recall events that took place when their child was diagnosed will decrease over time. Such a diminishing ability could have two distinct effects: Parents recollection becomes less accurate over time and their recollection becomes biased in a systematic way. In this appendix we test whether time elapsed between the time the child was diagnosed and the survey date, affected parents' recollections concerning the clinical signs their child exhibited at the time of diagnosis.

Appendix Table B1 provides some descriptive statistics on the distribution values of the age of diagnosis, the child age at the time of survey, and the time elapsed between the two dates.

Appendix Figures 1-4 provide visual description of these variables.

In order to test for potential recall bias, we ask the following questions:

1. Does time elapsed since diagnosis affect the likelihood that a clinical sign or a symptom will be reported by parents?
2. Does time elapsed since diagnosis affect the intensity of a sign reported by parents?

In Appendix Table B2 we test these two questions directly. In the first column we report the estimation results of 25 separate regressions. In each regression we estimate the effect of time elapsed on the likelihood that a clinical sign will be reported. For none of the clinical signs, the time elapsed decreases the likelihood that the sign will be recalled by the parents.

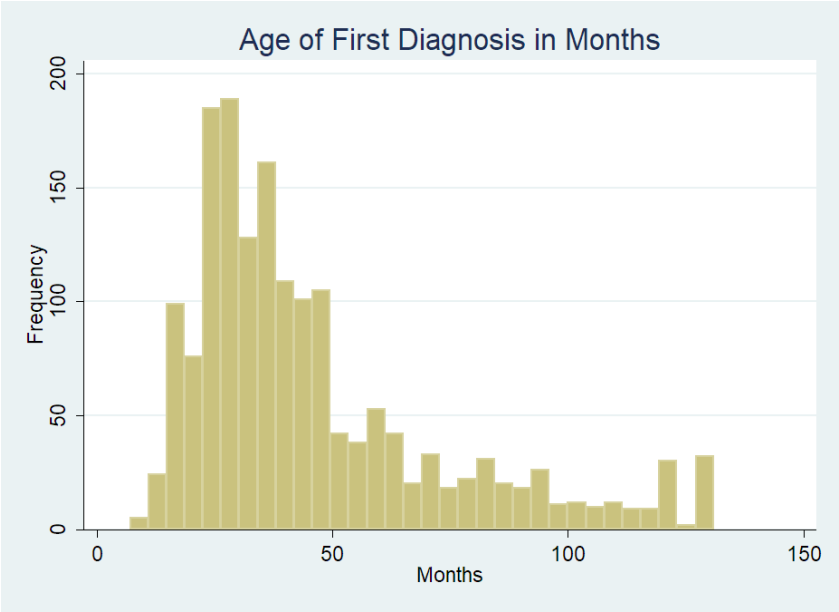
In the second column, we estimate, for each clinical sign or symptom, the effect of time elapsed on the level of severity recalled by the parents. For none of the signs the reported severity is decreasing with time. In fact, for many signs and symptoms, the level of reported severity is increasing with time elapsed, although none of the coefficients is significantly different from zero.

To summarize, we find no evidence that time elapsed between the time of diagnosis and the survey has any effect on the two central variables of our analysis, the signs and symptoms reported by the parents and the levels of severity as recalled by the parents.

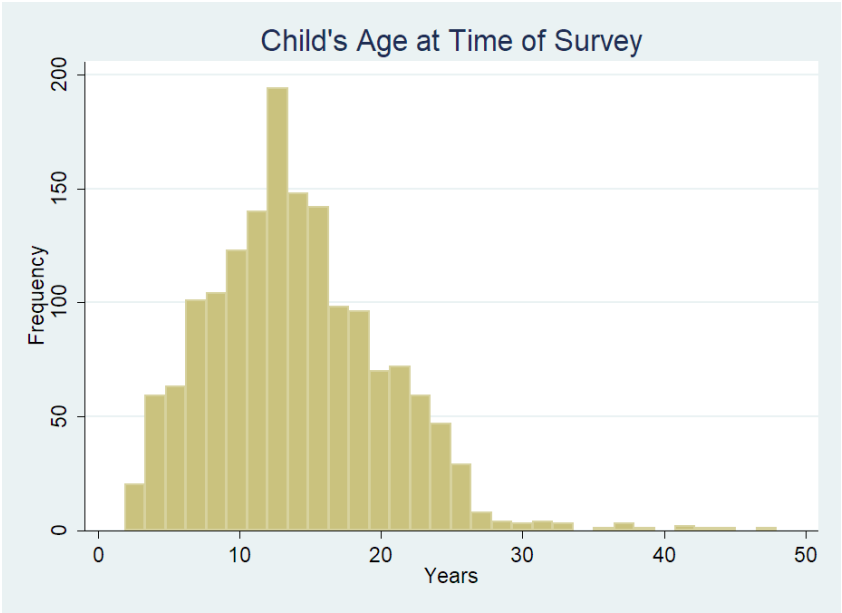
Our findings are consistent with the idea that as past events are more significant, recall accuracy is higher (Coughlin, 1990). For example, Wilcox and Horney found that the major determinant of spontaneous abortion recall was the length of pregnancy at the time of abortion (Wilcox and Horney, 1984). Only 54% of abortions which accrued within the first 6 weeks of gestation were recalled, compared to 93% of those occurring after 13 weeks.

Appendix Table B1						
Child age at diagnosis, at time of survey, and time elapsed						
	Count	Mean	Median	25% perc.	75% perc.	90% perc.
Age of Diagnosis (in months)	1672	46	36	27	57	89
Age at time of survey	1597	168	161	117	215	267
Time Elapsed	1597	122	115	70	163	221

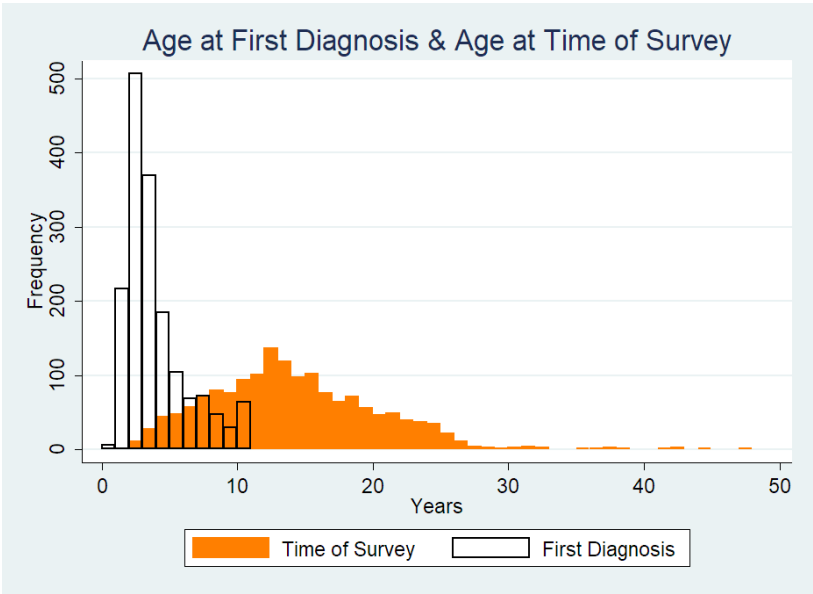
Appendix Figure B1



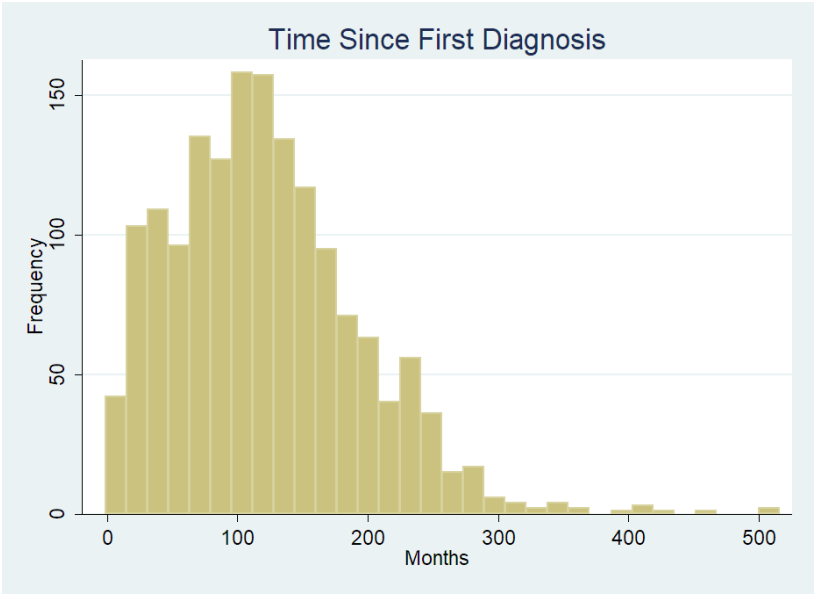
Apnnedix Figure B2



Appendix Figure B3



Appendix Figure B4



Appendix Table B2

The effects of Time Elapsed on the Likelihood that a sign is recalled
and the reported severity of the symptom

	A Symptom is reported	Reported intensity
Delayed speech	0.00124	-0.00497
	(0.00230)	(0.0183)
Delayed response to name	0.00107	0.0166
	(0.00266)	(0.0213)
Poor eye contact	-0.00210	0.0434*
	(0.00171)	(0.0187)
Lack of gestures	0.00212	0.0309
	(0.00248)	(0.0218)
Difficulty understanding gestures	0.00236	0.0120
	(0.00254)	(0.0220)
Played alone	0.000899	0.0243
	(0.00150)	(0.0176)
Too focused on objects	-0.000876	0.0348
	(0.00166)	(0.0178)
Poor social skills	0.000526	0.0128
	(0.00141)	(0.0181)
Difficulty initiating relationships	-0.000146	0.00891
	(0.00143)	(0.0179)
Unusual play with toys	-0.00229	0.00887
	(0.00160)	(0.0193)
Need for sameness	-0.000828	-0.00349
	(0.00142)	(0.0218)
Unusual motor mannerisms	-0.00211	0.00176
	(0.00212)	(0.0248)
Unusual interest in objects	0.000216	0.00495
	(0.00138)	(0.0212)
Unusual sensory interest	0.00155	-0.00251
	(0.00146)	(0.0199)
Sensory hyperreactivity	-0.00314	0.0124
	(0.00253)	(0.0229)
Sensory hyporeactivity	-0.00254	-0.00674
	(0.00234)	(0.0212)
Temper tantrums	-0.00220	0.0203
	(0.00228)	(0.0212)
Aggression toward self	0.00102	-0.00366
	(0.00352)	(0.0280)
Aggression toward others	-0.00131	0.0218
	(0.00340)	(0.0250)
Loss of skills	-0.000139	0.0288
	(0.00356)	(0.0299)
Loss of language (words only)	-0.00282	-0.00745
	(0.00352)	(0.0345)
Loss of language (phrases)	0.00393	-0.00346
	(0.00351)	(0.0400)
Less social engagement	0.00423	0.0138
	(0.00353)	(0.0286)
Loss of motor skills	0.00392	-0.0301
	(0.00346)	(0.0345)
Loss of daily living skills	0.00393	-0.0283
	(0.00350)	(0.0375)

Standard errors in parentheses

* p<0.05 ** p<0.01 *** p<0.001