

Supplementary Results

Data-driven characterization of individuals with delayed autism diagnosis

Dan Aizenberg, Ido Shalev, Florina Uzefovsky, and Alal Eran

This file includes

Figures 1 – 23

Tables 1 – 4

Supplementary results

I. Discovery cohort results

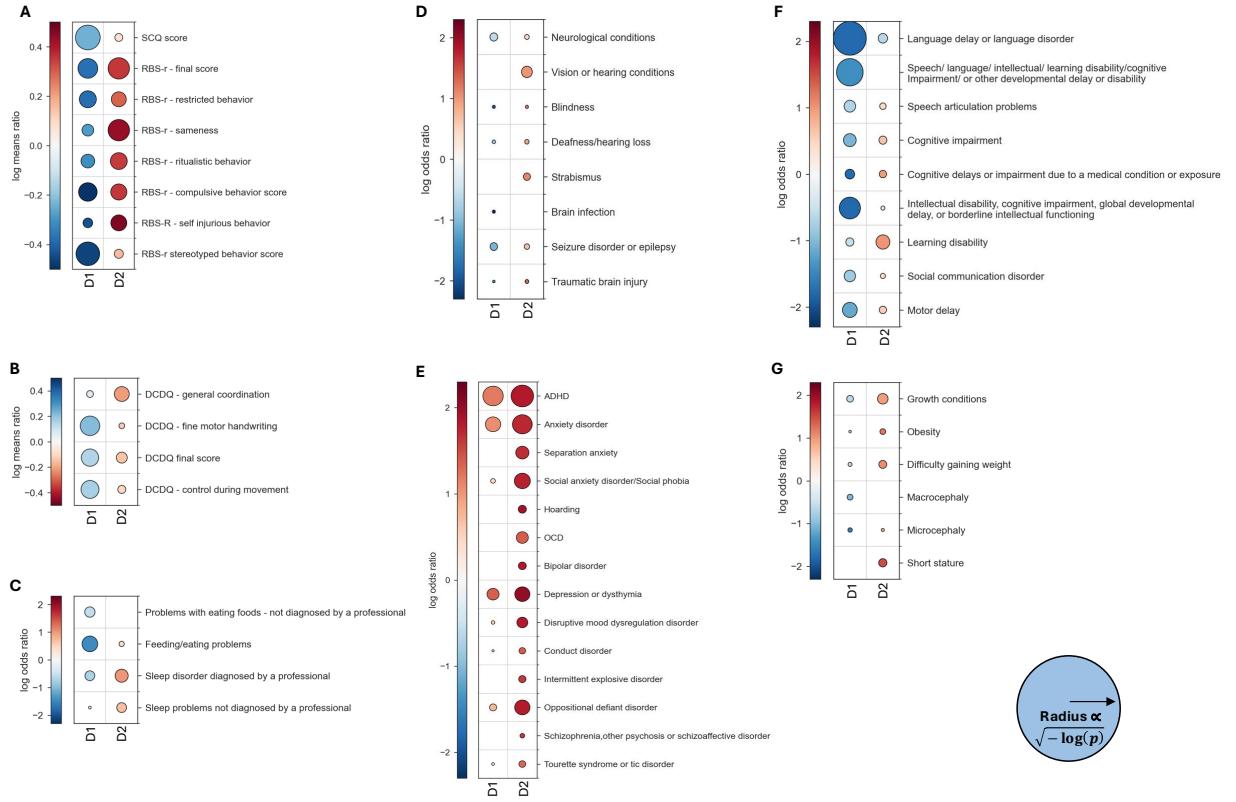


Figure 1. As compared to individuals with a timely ASD diagnosis, D1 individuals have fewer autistic traits and less co-occurring conditions, whereas D2 individuals have more autistic traits and more co-occurring conditions. Behavioral differences are consistent across all autism domains and multiple behavioral instruments, including (A) core autism domains, most pronounced in stereotyped and compulsive behaviors, and (B) motor development. Differences in co-occurring conditions include (C) sleep and eating problems, (D) sensory and neural problems, (E) neuropsychiatric conditions, (F) cognitive and motor delays, and (G) growth alterations. The color of each circle represents the effect size measured by the log ratio of a feature's mean in each group as compared to the feature's mean among timely diagnosed individuals. The size of the circles is proportional to $\sqrt{-\log(P \text{ value})}$. Thus, the larger the circle, the smaller the P value. The lowest possible P value in this analysis is 6.87×10^{-278} . Circles are shown only for significant results.

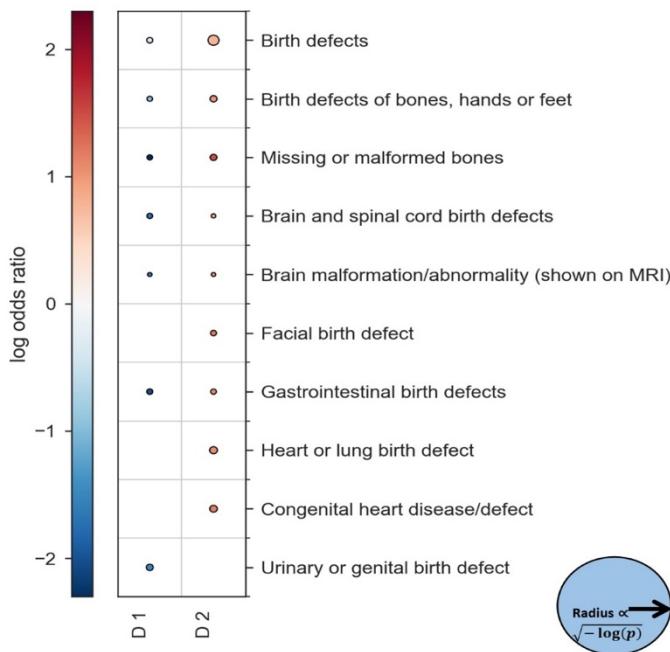


Figure 2. Differences in congenital disorders between individuals in each delayed diagnosis group and those that received a timely diagnosis in the discovery cohort. The color of each circle represents the effect size measured by the log ratio of a feature's mean in each group as compared to the feature's mean among timely diagnosed individuals. The size of the circles is proportional to $\sqrt{-\log(P \text{ value})}$. Thus, the larger the circle, the smaller the P value. The lowest possible P value in this analysis is 6.87×10^{-278} . Circles are shown only for significant results.

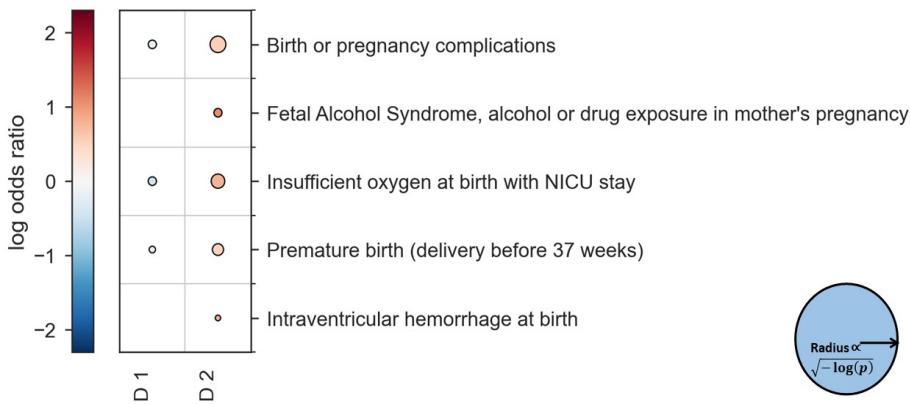


Figure 3. Differences in prenatal and perinatal complications between individuals in each delayed diagnosis group and those with a timely diagnosis in the discovery cohort. The color of each circle represents the effect size measured by the log ratio of a feature's mean in each group as compared to the feature's mean among timely diagnosed individuals. The size of the circles is proportional to $\sqrt{-\log(P \text{ value})}$. Thus, the larger the circle, the smaller the P value. The lowest possible P value in this analysis is 6.87×10^{-278} . Circles are shown only for significant results.

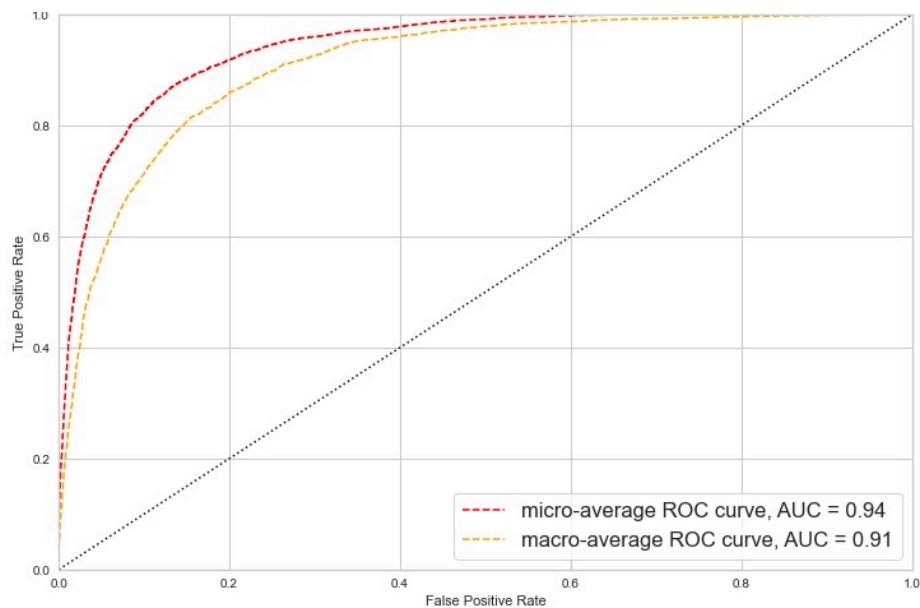


Figure 4. Receiver operating characteristics (ROC) curve for distinguishing the two delayed diagnosis groups using a random forest classifier.

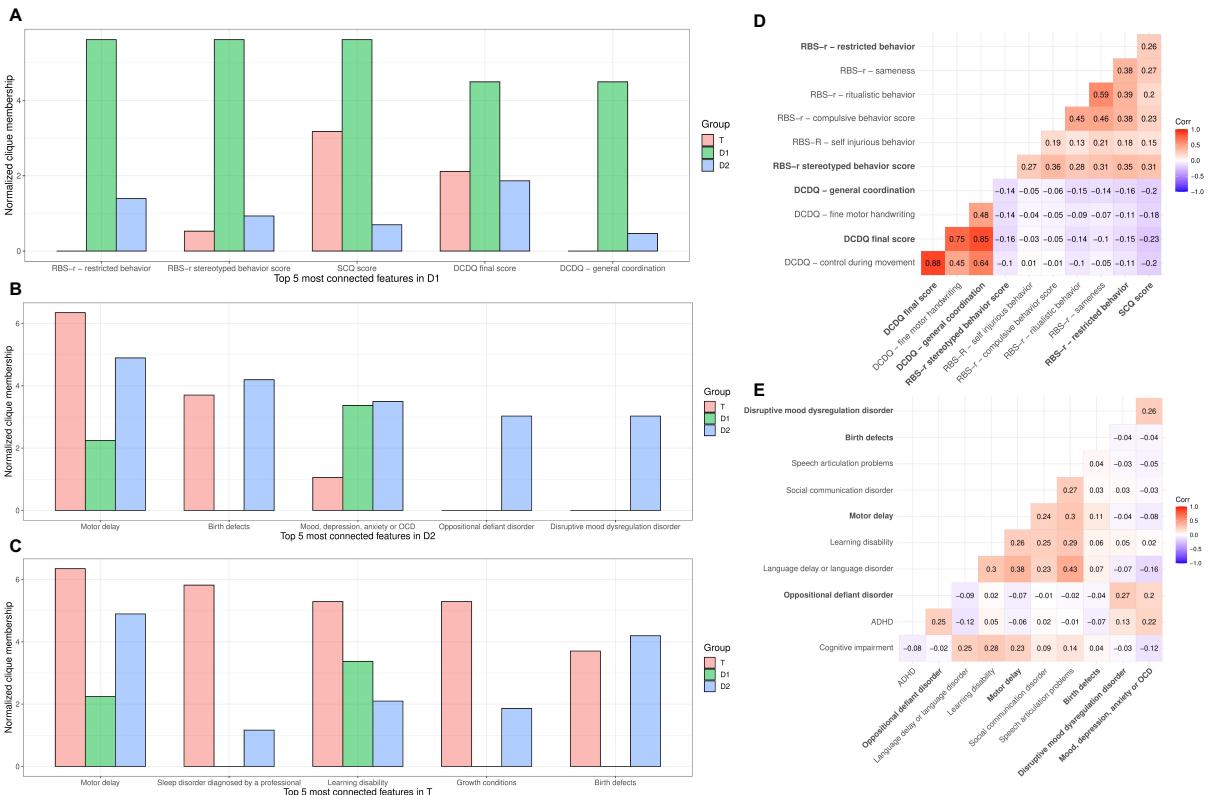


Figure 5. Feature interconnectivity and network structure of timely and delayed diagnosis groups. Normalized clique membership is depicted for the top five most connected features in (A) D1, (B) D2, and (C) the timely diagnosis group, T. Correlation heatmaps of the variables most strongly correlated with the top connected features are presented for (D) D1, and (E) D2. The top connected features are denoted in bold.

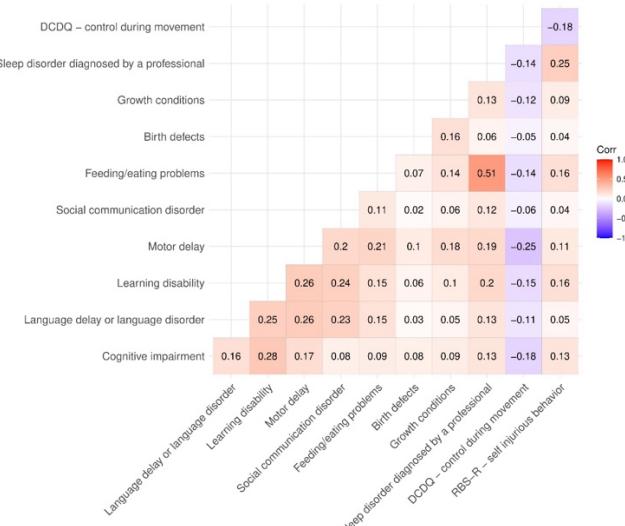


Figure 6. Heatmap of variables most strongly correlated with the top connected features of the timely diagnosis group.

Prevalance of psychiatric disorders by diagnosis age

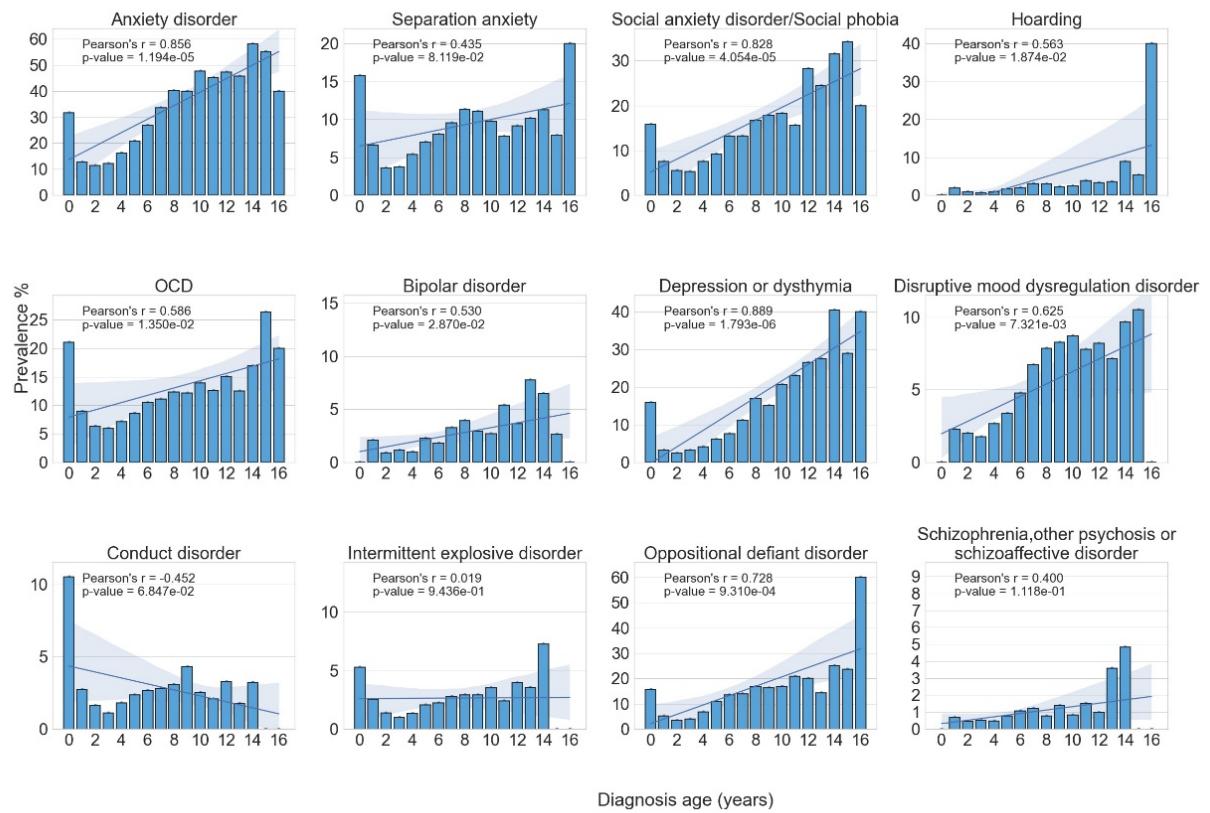


Figure 7. Prevalence of psychiatric disorders according to the age of autism diagnosis. For each disorder, Pearson's correlation was calculated between the age of autism diagnosis and the prevalence of the disorder in individuals diagnosed at that age. The least squares regression line is shown along its 95% confidence interval.

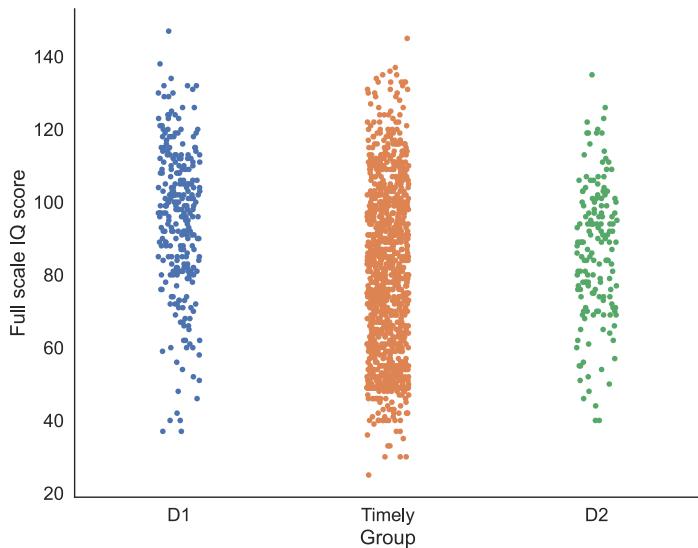


Figure 8: Full scale IQ scores in 296 D1 individuals, 1380 timely diagnosed individuals, and 172 D2 individuals. There was no IQ information on the remaining study participants. IQ was likely measured in non-randomly selected study participants, suggesting that the available IQ data may be biased. Nonetheless, the data at hand reflects a broad range of IQ in each group.

Table 1. Pairwise correlations between informative characteristics of the D1 group. This large table could be accessed at https://github.com/danaiz/Data-driven-characterization-of-individuals-with-a-delayed-autism-diagnosis/blob/main/net_corr_p_1_fdr.xlsx

Table 2. Pairwise correlations between informative characteristics of the D2 group. This large table could be accessed at https://github.com/danaiz/Data-driven-characterization-of-individuals-with-a-delayed-autism-diagnosis/blob/main/net_corr_p_2_fdr.xlsx

Table 3. Pairwise correlations between informative characteristics of the timely diagnosis group. This large table could be accessed at https://github.com/danaiz/Data-driven-characterization-of-individuals-with-a-delayed-autism-diagnosis/blob/main/net_corr_p_ed_fdr.xlsx

Table 4. Feature participation in pairwise correlation 4-cliques. Pairwise correlations between any two features were calculated separately within the T, D1, and D2 subgroups, and modeled as a dense graph of feature nodes connected by correlation edges. Each graph was analyzed via 4-cliques. This table summarizes the percent of 4-cliques each feature belongs to, as a measure of its information flow.

	T cliques	D1 cliques	D2 cliques
Motor delay	6.349206	2.247191	4.895105
Birth defects	3.703704	0	4.195804
Mood, depression, anxiety or OCD	1.058201	3.370787	3.496503
Oppositional defiant disorder	0	0	3.030303
Disruptive mood dysregulation disorder	0	0	3.030303
Depression or dysthymia	0	1.123596	2.797203
Cognitive impairment	2.645503	1.123596	2.564103
Gastrointestinal birth defects	2.116402	0	2.564103
Brain and spinal cord birth defects	1.587302	0	2.564103
DCDQ - fine motor handwriting	0	2.247191	2.564103
RBS-r - final score	2.645503	3.370787	2.331002
Social anxiety disorder or Social phobia	0.529101	3.370787	2.331002
speech	0.529101	3.370787	2.331002
Learning disability	5.291005	3.370787	2.097902
Birth defects of bones, hands or feet	1.587302	0	2.097902
Language delay or language disorder	0.529101	3.370787	2.097902
Spine deformity	0.529101	0	2.097902
RBS-r - compulsive behavior score	0	1.123596	2.097902
Growth conditions	5.291005	0	1.864802
DCDQ final score	2.116402	4.494382	1.864802
OCD	1.058201	2.247191	1.864802
Anxiety disorder	0.529101	3.370787	1.864802
Urinary or genital birth defect	0.529101	0	1.864802
RBS-r - ritualistic behavior	0	3.370787	1.864802
RBS-r - sameness	0.529101	2.247191	1.631702
Heart or lung birth defect	0.529101	0	1.631702
Separation anxiety	0	1.123596	1.631702
Attention or behavior disorders	3.174603	0	1.398601
Neurological conditions	2.116402	0	1.398601
Missing or malformed bones	1.058201	0	1.398601
Congenital heart disease or defect	0.529101	0	1.398601
RBS-r - restricted behavior	0	5.617978	1.398601
Intermittent explosive disorder	0	0	1.398601
Bipolar disorder	0	0	1.398601
Sleep disorder diagnosed by a professional	5.820106	0	1.165501
Premature birth (delivery before 37 weeks)	2.645503	3.370787	1.165501
Speech articulation problems	1.587302	2.247191	1.165501
Brain malformation or abnormality (shown on MRI)	0.529101	0	1.165501
Lung malformation	0.529101	0	1.165501

Seizure disorder or epilepsy	0	0	1.165501
Intellectual , cognitive , global	3.703704	1.123596	0.932401
Vision or hearing conditions	3.174603	0	0.932401
Difficulty gaining weight	2.116402	0	0.932401
RBS-r stereotyped behavior score	0.529101	5.617978	0.932401
Conduct disorder	0	0	0.932401
SCQ score	3.174603	5.617978	0.699301
Year of birth	2.645503	2.247191	0.699301
Birth or pregnancy complications	2.116402	2.247191	0.699301
Cognitive delays or impairment due to a medical condition or exposure	1.058201	0	0.699301
Facial birth defect	1.058201	0	0.699301
RBS-R - self injurious behavior	0.529101	1.123596	0.699301
Short stature	0.529101	0	0.699301
Hoarding	0	0	0.699301
Twin participating in SPARK	2.645503	3.370787	0.4662
Twin	2.645503	3.370787	0.4662
No twin	2.645503	3.370787	0.4662
DCDQ - control during movement	2.116402	2.247191	0.4662
Insufficient oxygen at birth with NICU stay	1.587302	1.123596	0.4662
ADHD	0.529101	0	0.4662
Intraventricular hemorrhage at birth	0.529101	0	0.4662
DCDQ - general coordination	0	4.494382	0.4662
Strabismus	0	0	0.4662
Pyloric stenosis	0	0	0.4662
Extra fingers and or or extra toes	0	0	0.4662
Missing kidney	0	0	0.4662
Monozygotic twin	1.058201	2.247191	0.2331
Social communication disorder	1.058201	1.123596	0.2331
Suspected cause of ASD - Birth or delivery complication	1.058201	1.123596	0.2331
Dizygotic twin	1.058201	1.123596	0.2331
Suspected cause of ASD - Problems during pregnancy	0.529101	1.123596	0.2331
Macrocephaly	0	0	0.2331
Intestinal malrotation	0	0	0.2331
Cleft palate	0	0	0.2331
Microcephaly	0	0	0.2331
Feeding or eating problems	3.174603	0	0
Nuclear family ASD	1.058201	1.123596	0

II. Validation cohort results

Two distinct subgroups of individuals with delayed ASD diagnosis were detected in the validation cohort. The first group, *VD1*, includes 1,801 individuals with lower support needs as compared to 7,453 individuals with a timely ASD diagnosis (which we here term *VT*). The second subgroup, *VD2*, includes 939 individuals with higher support needs, as consistently reflected by all considered behavioral instruments, and exemplified by RBS-R quantified RRBs (*VD1*: MR = 0.661, 95% CI = 0.658-0.663; *VD2*: MR = 1.436, 95% CI = 1.429-1.445, **Figure 9**).

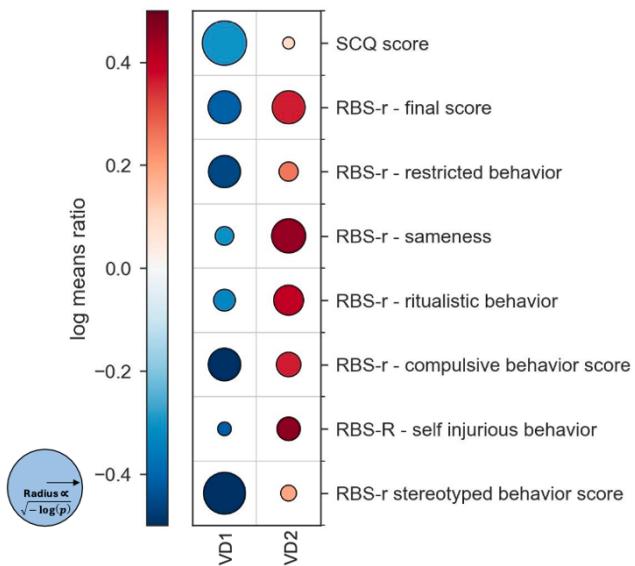


Figure 9. Differences in core autism domains between individuals in each delayed diagnosis group and those with a timely diagnosis, all from the validation cohort. The color of each circle represents the effect size measured by the log ratio of a feature's mean in each group as compared to the feature's mean among timely diagnosed individuals. The size of the circles is proportional to $\sqrt{-\log(P \text{ value})}$. Thus, the larger the circle, the smaller the P value. The lowest possible P value in this analysis is 6.87×10^{-278} . Circles are shown only for significant results.

Overall, recapitulating findings from the discovery cohort, *VD1* individuals were found to have fewer autistic traits as compared to *VT* individuals, whereas *VD2* individuals have more. These include difficulties in social communication and RRBs (**Figure 9**), motor development (**Figure 10**), and feeding and sleeping problems (**Figure 11**). Moreover, *VD1* individuals have lower odds of co-occurring neurodevelopmental conditions than *VT* individuals, whereas *VD2* individuals have higher odds for such conditions, including cognitive and motor delays (**Figure 12**), growth abnormalities (**Figure 13**), and neural problems (**Figure 14**).

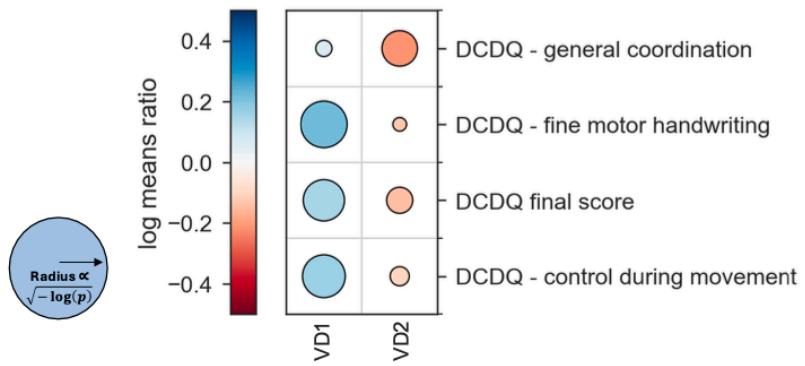


Figure 10. Differences in motor development between individuals in each delayed diagnosis group and those with a timely diagnosis, all from the validation cohort. Color coding and circle size are the same as in Figure 9.

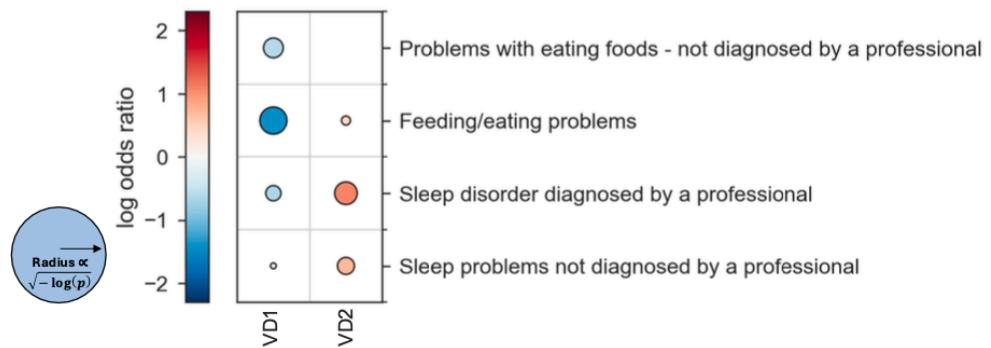


Figure 11. Differences in feeding and sleeping problems between individuals in each delayed diagnosis group and those with a timely diagnosis, all from the validation cohort. Color coding and circle size are the same as in Figure 9.

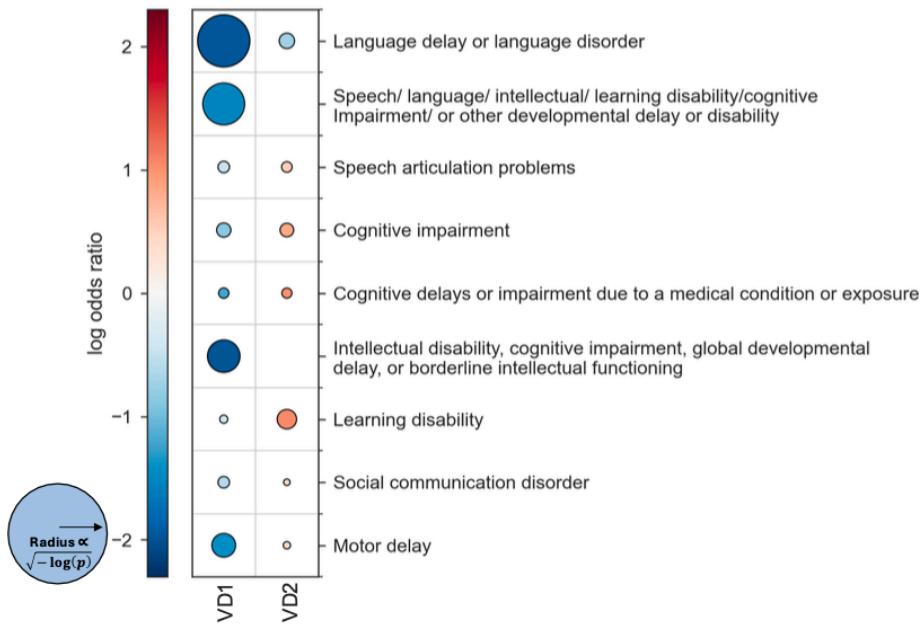


Figure 12. Differences in cognitive and motor delays between individuals in each delayed diagnosis group and those with a timely diagnosis, all from the validation cohort. Color coding and circle size are the same as in Figure 9.

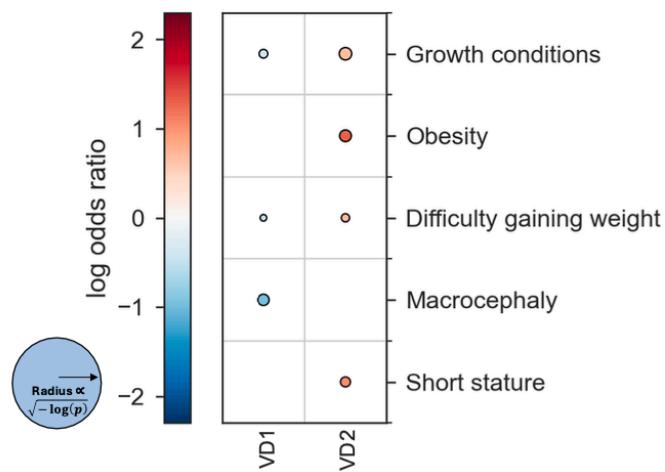


Figure 13. Differences in growth abnormalities between individuals in each delayed diagnosis group and those with a timely diagnosis, all from the validation cohort. Color coding and circle size are the same as in Figure 9.

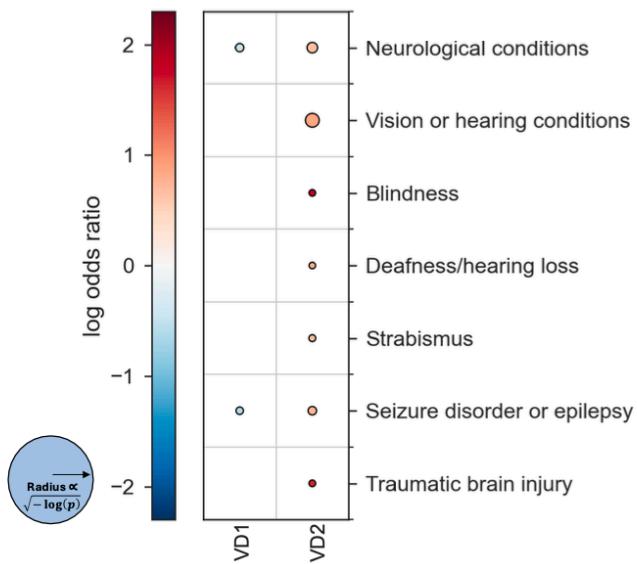


Figure 14. Differences in sensory and neural problems between individuals in each delayed diagnosis group and those that received a timely diagnosis, all from the validation cohort. Color coding and circle size are the same as in Figure 9.

Further confirming our findings in the discovery cohort, both delayed diagnosis groups of the validation cohort have higher rates of ADHD, anxiety, and depression or dysthymia as compared to timely diagnosed individuals. However, VD2 individuals have higher rates of neuropsychiatric disorders such as bipolar disorder and schizophrenia, while VD1 individuals do not (Figure 15). Finally, prenatal and perinatal complications are less common among VD1 individuals as compared to VT individuals, whereas VD2 individuals have higher odds for such conditions, including fetal alcohol syndrome, premature birth, and insufficient oxygen at birth (Figure 16).

Consistent with findings in the discovery cohort, when considering all comorbidities, individuals in VD1 have fewer comorbidities compared to those in VT, while individuals in VD2 have more co-occurring conditions (VD1: mean = 3.42, $t = 6.62$; VD2: mean = 7.41, $t = 31.84$, VT mean = 3.93, $P < 1.79 \times 10^{-210}$).

Taken together, the validation cohort confirms the existence of two distinct groups of individuals with delayed ASD diagnosis, differing by their level of autistic traits and the degree of co-occurring conditions. It is important to note that the independent validation cohort is drawn from the same population as the discovery cohort, and as such is subject to the same selection bias. Therefore, further efforts are required to generalize these findings to broader populations.

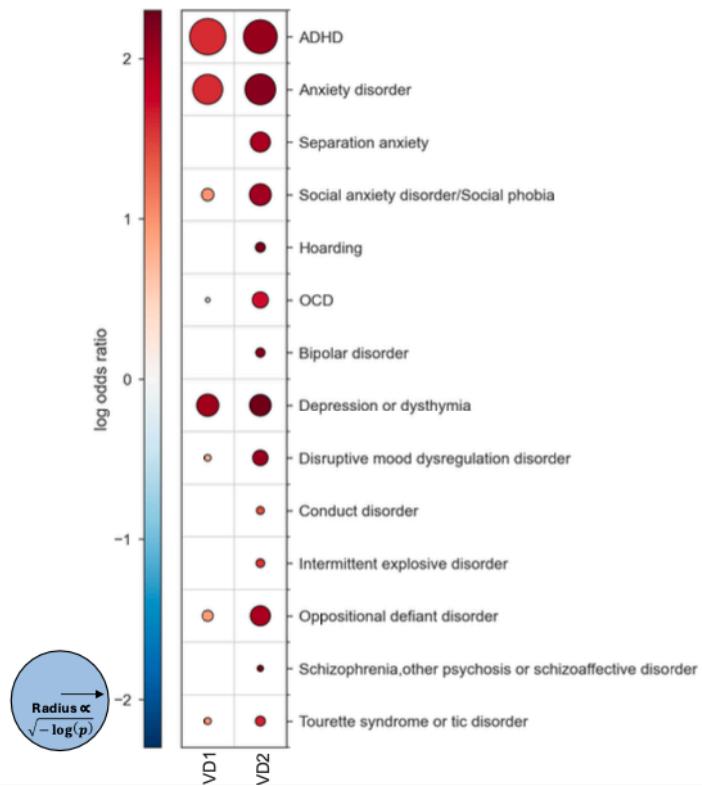


Figure 15 Differences in neuropsychiatric conditions between individuals in each delayed diagnosis group and those that received a timely diagnosis, all from the validation cohort. Color coding and circle size are the same as in Figure 9.

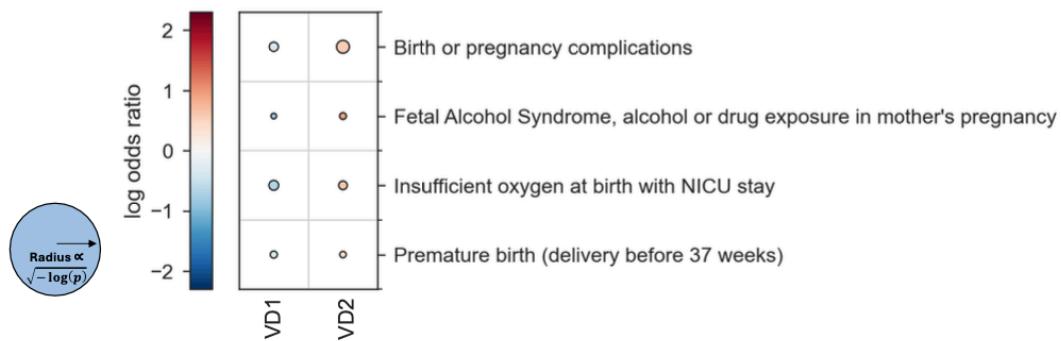


Figure 16 Differences in reported prenatal and perinatal complications between individuals in each delayed diagnosis group and those that received a timely diagnosis, all from the validation cohort. Color coding and circle size are the same as in Figure 9.

III. Replication analysis

An independent clustering approach, agglomerative clustering, identified two distinct groups of individuals with delayed diagnosis, AD1 and AD2, differing in their level of autistic traits and comorbidity burdens. Again, the strongest differences were in RRBs (**Figure 17**), with replicated differences detected in motor development (**Figure 18**), feeding and sleeping problems (**Figure 19**), growth abnormalities (**Figure 20**), neural problems (**Figure 21**), prenatal and perinatal conditions (**Figure 22**), and neuropsychiatric conditions (**Figure 23**).

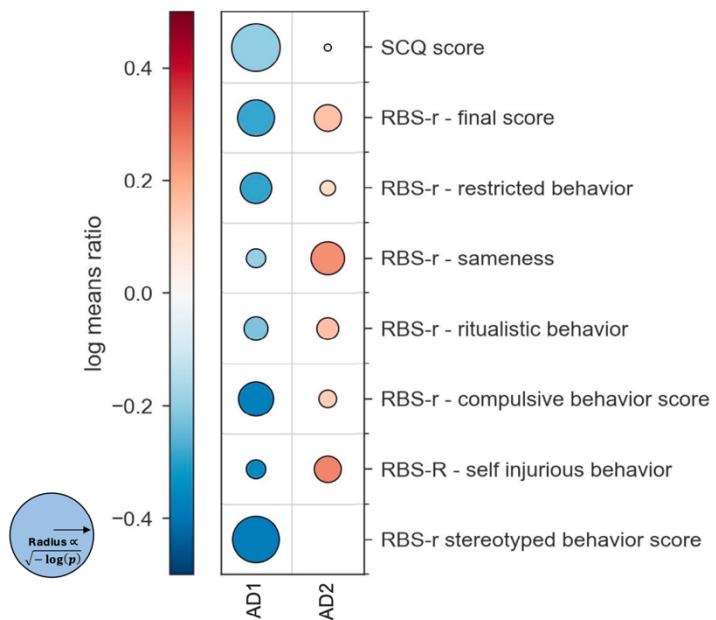


Figure 17. Differences in core autism domains between individuals in each delayed diagnosis group identified using agglomerative clustering, and individuals with a timely diagnosis. The color of each circle represents the effect size measured by the log ratio of a feature's mean in each group as compared to the feature's mean among timely diagnosed individuals. The size of the circles is proportional to $\sqrt{-\log(P \text{ value})}$. Thus, the larger the circle, the smaller the P value. The lowest possible P value in this analysis is 6.87×10^{-278} . Circles are shown only for significant results.

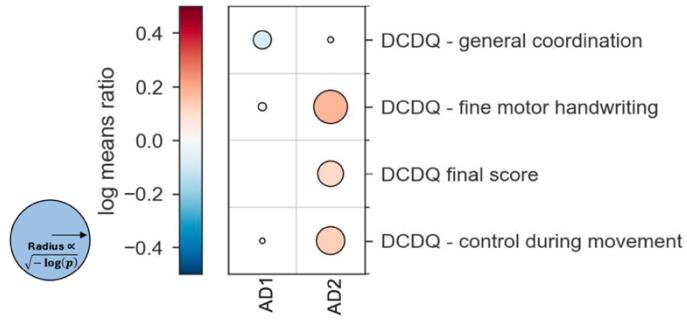


Figure 18. Differences in motor development between individuals in each delayed diagnosis group identified using agglomerative clustering, and individuals with a timely diagnosis. Color coding and circle size are the same as in Figure 17.

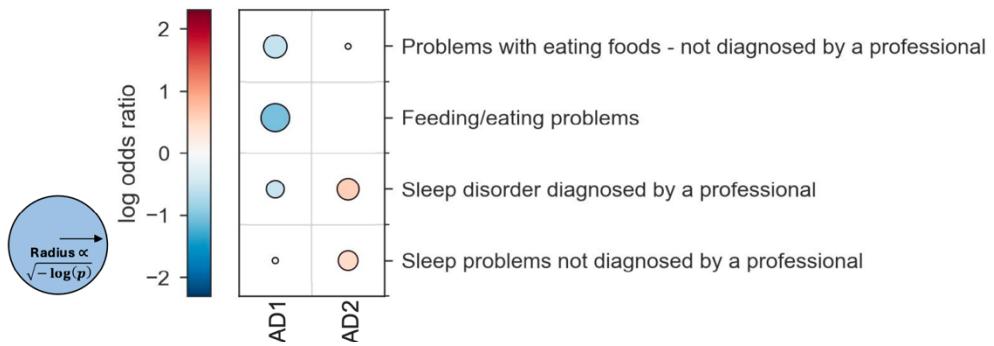


Figure 19. Differences in feeding and sleeping problems between individuals in each delayed diagnosis group identified using agglomerative clustering, and individuals with a timely diagnosis. Color coding and circle size are the same as in Figure 17.

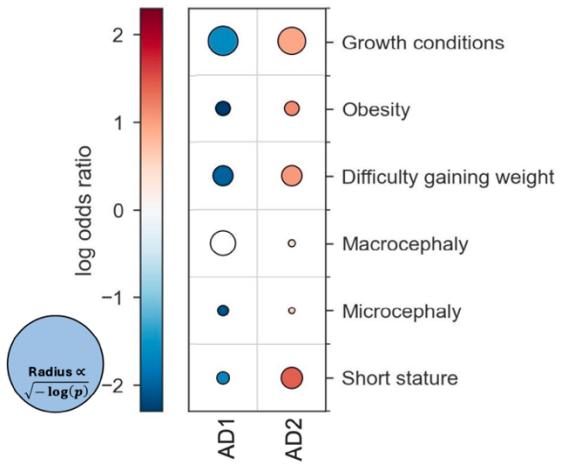


Figure 20. Differences in growth abnormalities between individuals in each delayed diagnosis group identified using agglomerative clustering, and individuals with a timely diagnosis. Color coding and circle size are the same as in Figure 17.

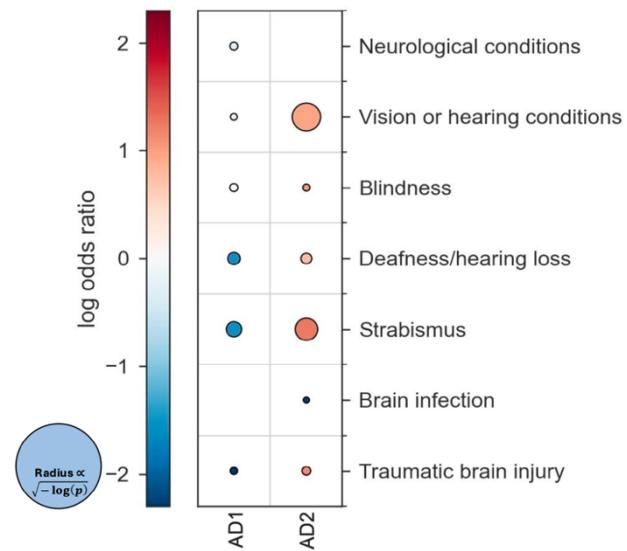


Figure 21. Differences in sensory and neural problems between individuals in each delayed diagnosis group identified using agglomerative clustering, and individuals with a timely diagnosis. Color coding and circle size are the same as in Figure 17.

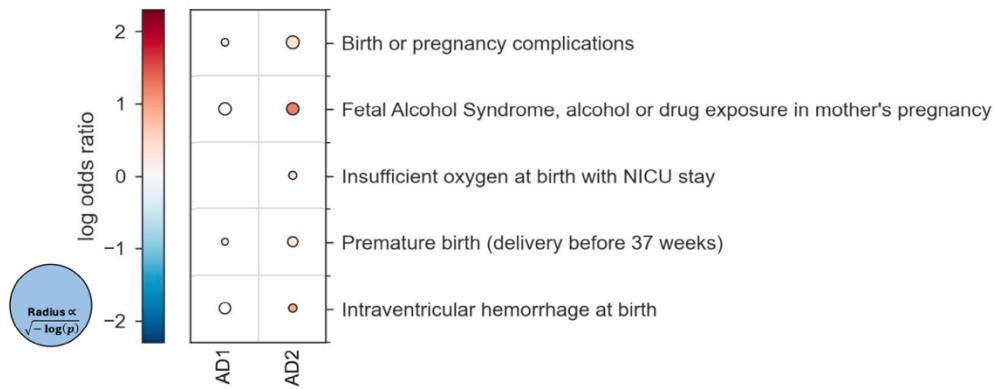


Figure 22. Differences in reported prenatal and perinatal complications between individuals in each delayed diagnosis group identified using agglomerative clustering, and individuals with a timely diagnosis. Color coding and circle size are the same as in Figure 17.

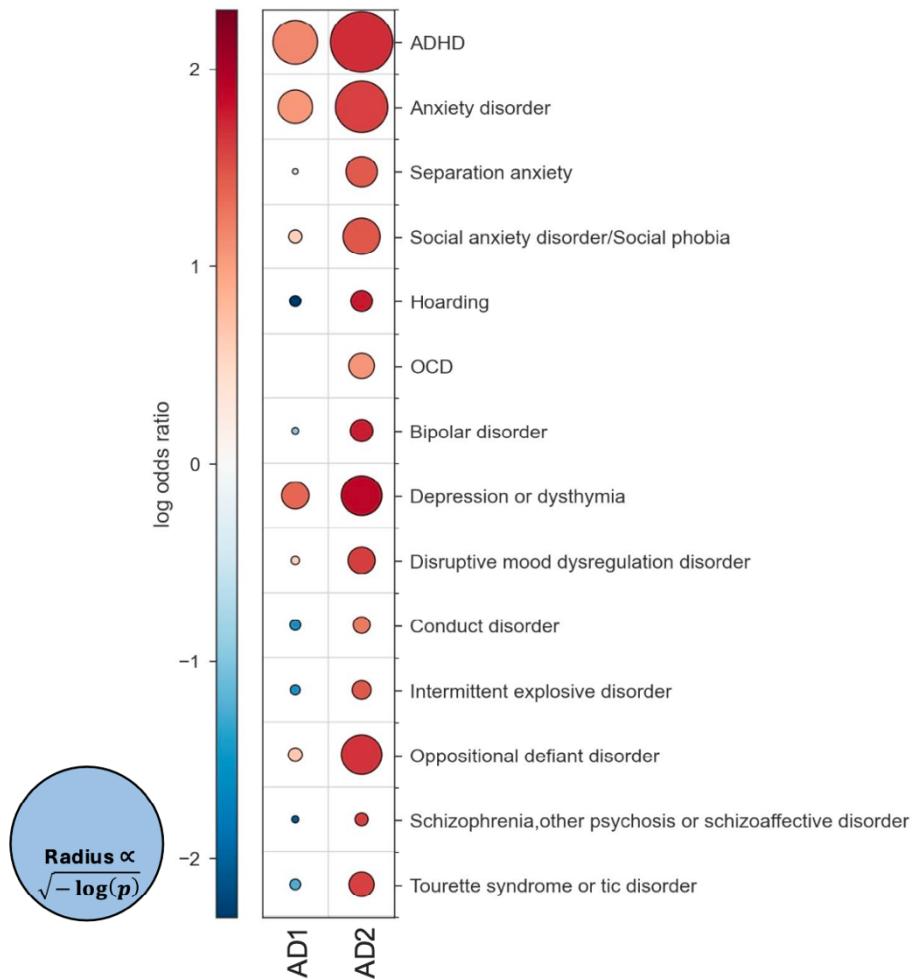


Figure 23. Differences in neuropsychiatric conditions between individuals in each delayed diagnosis group identified using agglomerative clustering, and individuals with a timely diagnosis. Color coding and circle size are the same as in Figure 17.