



LARANA, INC.

# **SEVERITY OR SUBTYPE?**

**AN INVESTIGATION OF HETEROGENEITY IN AUTISM  
SPECTRUM DISORDER**

**A PROJECT BY JAVONNE PORTER**

**SUPERVISED BY LLUIS, SARAH, AND CECI**

# CONTENTS

**01**

WELCOME

**02**

CONTENTS

**03**

AUTISM SPECTRUM DISORDER

**04**

HETEROGENEITY IN ASD

**05**

MY APPROACH (ADIDE, ANALYSIS PLAN)

**07**

MY RESULTS (OUTPUTS, ANALYSIS FINDINGS)

**10**

DISCUSSION

# AUTISM SPECTRUM DISORDER (ASD)

## The DSM-V

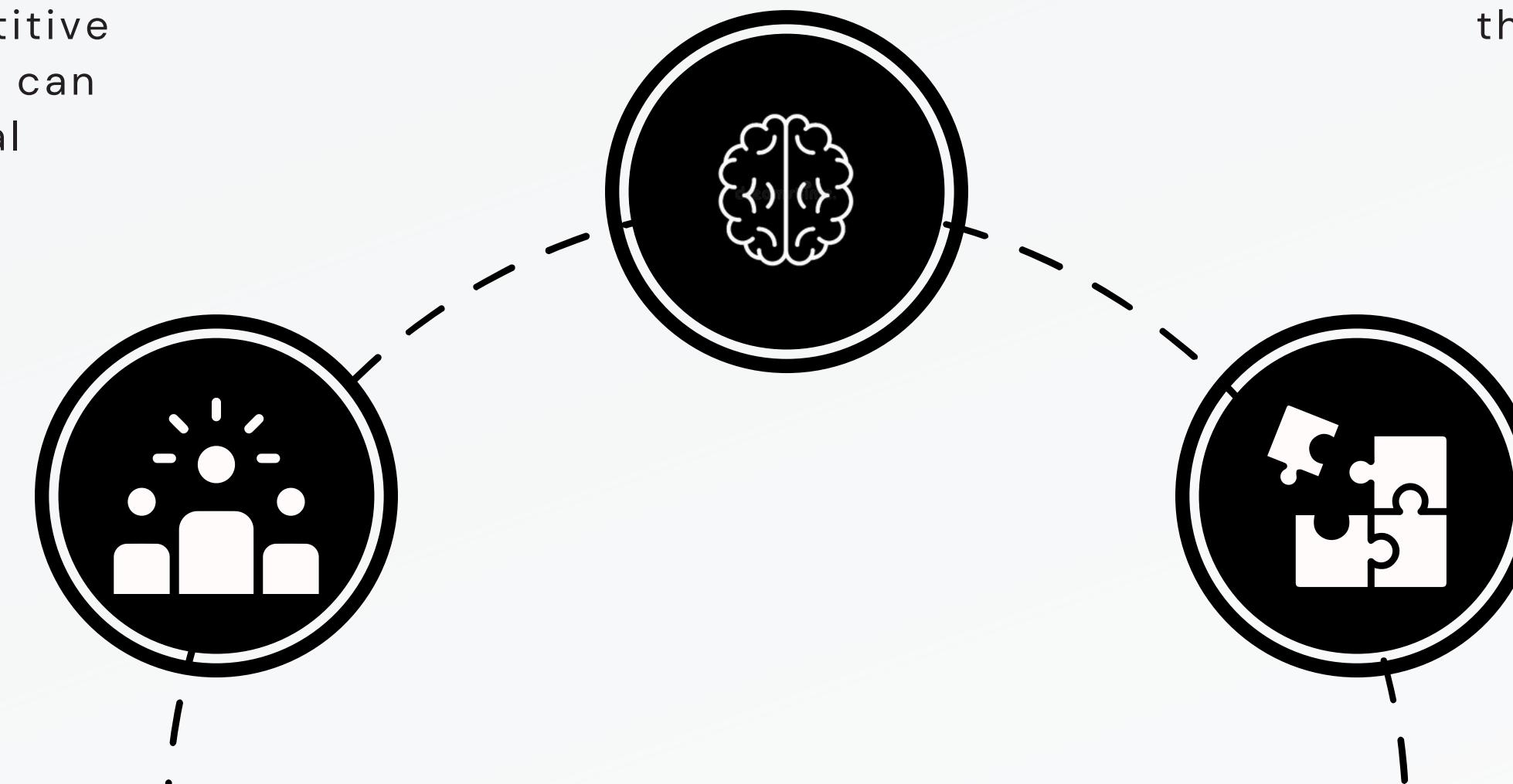
### What is it?

ASD is a Neurodevelopmental Disorder that (primarily) causes restricted/repetitive behaviours and can impair Social Functioning.

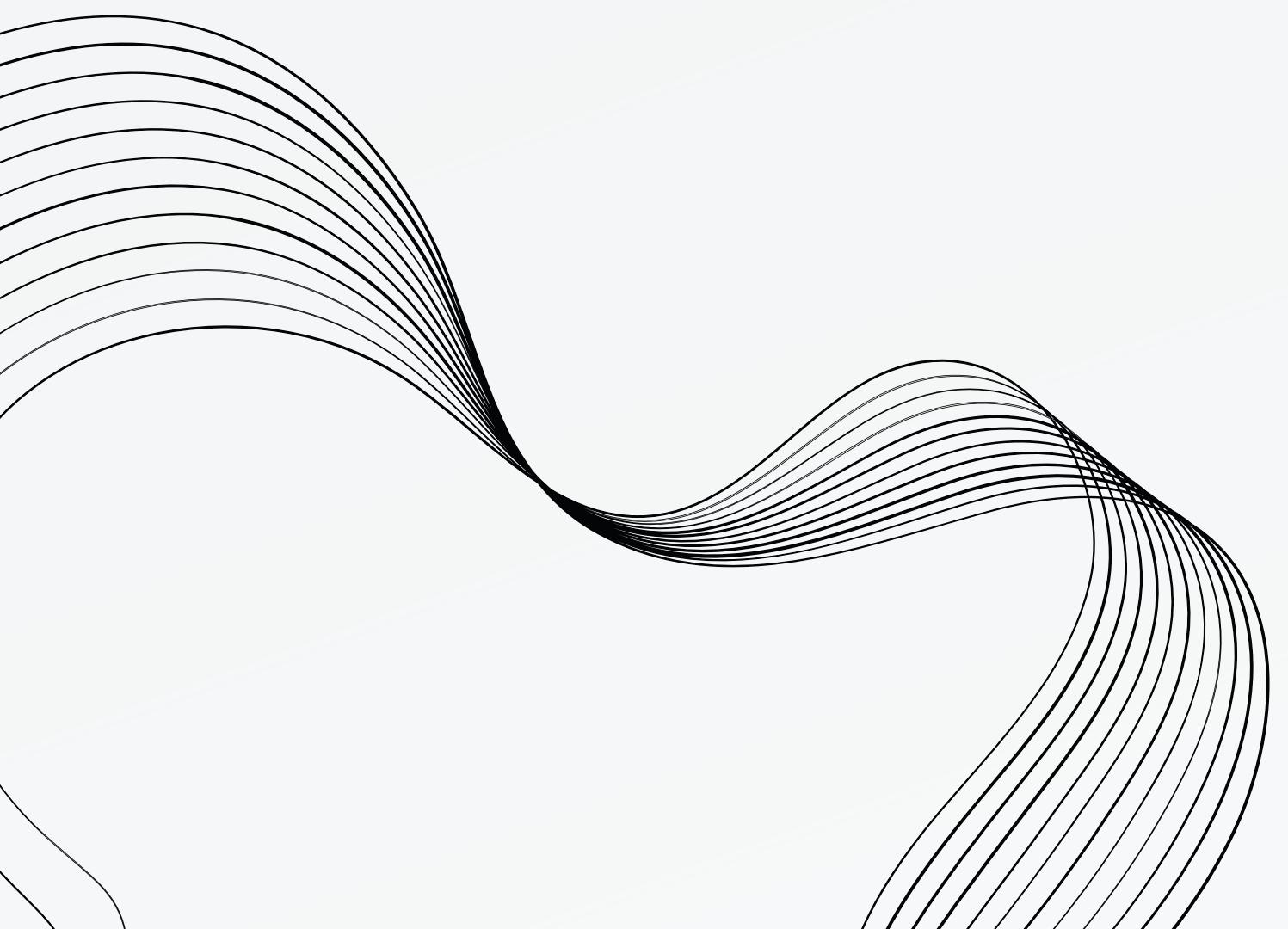
The Diagnostic and Statistical Manual of Mental Disorders (DSM) is a set of guidelines published by the American Psychiatric Association that seeks to provide a shared framework that clinicians can use to speak about Mental Disorders

### The spectrum

No 2 experiences of ASD are the same. This range of differences is implied by the use of the word “spectrum” in **Autism Spectrum** Disorder.



# HETEROGENEITY IN ASD



## SPIKE

The degree of variation or "heterogeneity" in ASD poses a problem. Former ASD researcher Uta Frith puts it succinctly stating that the diagnosis has been "stretched to breaking point" following a spike of 787% in the UK between 1998-2018. We can solve this problem with a better model of ASD

## SUBTYPE MODEL

The first type of model I investigated is a subtype model. This model would assume that ASD is best described by a set of related but distinct conditions that are similar enough to share 1 label: "ASD".

## SEVERITY MODEL

Severity models suggest that all experiences of ASD are largely the same, varying only in terms of severity.

## NEURO- BIOLOGY

These models can be defined behaviourally i.e. by behavioural outcomes, or neurobiologically i.e. by the brain of the person with ASD.

# MY APPROACH PT.1

## *The ABIDE Dataset*

- The ABIDE Dataset is a collection of data taken from various sites worldwide, intending to diminish the problem of small sample sizes in Autism research.
- It contains Phenotypic data (i.e. behavioural outcomes, and questionnaires). It also has Functional, Structural, and Diffusion MRI scans for each participant (Scans of brains). This data covers roughly 1000 participants

- To apply standard MRI preprocessing I used a prewritten pipeline provided by researchers at the Chinese Academy of Sciences known as the Connectome Computation System (CCS).
- After completing standard MRI preprocessing, and loading the volumes as arrays, I used PCA to reduce the dimensionality of the data in order to make analysis feasible.

## *Preprocessing with the CCS and PCA*

# MY APPROACH PT.2

## *ComBat and SRS to prepare for Analysis*

- (SRS) The Social Responsiveness Scale (SRS) is the key piece of data from the ABIDE's phenotypic dataset. The SRS is a questionnaire that attempts to operationalise the Severity of ASD.
- It reports on several categories such as Awareness, Cognition, and Mannerisms
- I also applied a technique called ComBat to the data. It reduces artefacts that may appear due to the fact that the scans were not all from the same machine.

- (Clustering) Clustering techniques find groups in disparate data. Following the example of papers that did similar investigations I decided to use clustering (KMeans) to investigate the best way to model ASD.
- I applied clustering to Neurobiological data (Structural MRI scans) and analysed patterns in the clusters using the Phenotypic data.

## *Applying Clustering*

# TABLES (BEFORE/AFTER AGE FILTER)

	Cluster 0 ( $N_0=57$ )	Cluster 1 ( $N_1=83$ )	Allistic Baseline ( $N_a=220$ )
Age	10.89 (2.64)*	26.27 (13.27)*	16.03 (11.56)
FIQ	104.79 (15.28)	107.95 (14.71)	114.74 (10.99)
Awareness (SRS)	13.33 (3.58)*	11.57 (4.08)*	4.51 (2.75)
Cognition (SRS)	17.09 (5.67)	18.34 (6.13)	3.19 (3.39)
Communication (SRS)	31.25 (9.89)	31.02 (11.07)	5.87 (5.46)
Motivation (SRS)	14.21 (6.6)	16.11 (6.65)	4.25 (3.72)
Mannerisms (SRS)	17.96 (6.74)	17.19 (6.77)	2.34 (2.96)
Medication Prevalence	0.61*	0.29*	0.13
Total SRS Score (RAW)	93.11 (28.24)	91.76 (30.37)	20.16 (15.07)

Table 1: Comparison of Clusters and Allistic Baseline

**Notes:**

- 1. \* indicates a statistically significant intergroup difference between Clusters 0 and 1
- 2. All differences between ASD scores and Allistic baseline scores showed statistically significant differences

	Cluster 0 ( $N_0=54$ )	Cluster 1 ( $N_1=14$ )	Allistic Baseline ( $N_a=161$ )
Age	10.4 (1.63)*	13.05 (1.65)*	10.58 (1.47)
FIQ	104.63 (15.64)	112.21 (13.26)	114.73 (11.14)
Awareness (SRS)	13.33 (3.58)	12.71 (3.3)	4.32 (2.72)
Cognition (SRS)	17.09 (5.67)	17.64 (6.31)	2.86 (3.02)
Communication (SRS)	32.28 (9.06)	34.29 (7.92)	5.1 (4.69)
Motivation (SRS)	14.74 (6.34)	16.14 (6.53)	3.57 (3.03)
Mannerisms (SRS)	18.56 (6.41)	14.79 (6.28)	1.75 (2.1)
Medication Prevalence	0.65*	0.14*	0.04
Total SRS Score (RAW)	96.0 (26.1)	95.57 (24.49)	17.59 (12.68)

Table 2: Comparison of Clusters and Allistic Baseline (Age filter applied)

**Notes:**

- 1. \* indicates a statistically significant intergroup difference between Clusters 0 and 1
- 2. Other than age, all differences between ASD scores and Allistic baseline scores showed statistically significant differences

# Outputs

01/02

03

04

05

## POST CCS SCANS

Datasets 1 and 2 are largely similar. They contain 376 volumes saved as python pkl files. Each is loaded as a numpy array in a python list. The first is Structural MRI, the second is Functional MRI

## PHENOTYPIC SUBSET

Dataset 3 is a subset of the Phenotypic data, that corresponds to the 376 subjects in datasets 1 and 2.

The 376 subjects are a subset of the full ABIDE dataset

## PCA SMRI

Dataset 4 is the initial Structural MRI dataset, after having PCA applied to it. 0 components have been removed.

## COMBAT SMRI

Dataset 5 is equivalent to dataset 4, but I have also applied ComBat to reduce the impact of cross site effects.

# LACK OF EVIDENCE FOR NBC'S

**Changing N:** After applying a post hoc age filter, the N of cluster 0 stays roughly the same, while the N of cluster 1 plummets to less than one-seventh of the original suggesting that age-related neurobiological differences were what the model discovered.

**Equivalent Severity:** As shown by the nearly identical severity levels between each cluster, my results fail to show neurobiological clusters with different levels of severity.

**Future potential:** With my dataset, it is still possible to discover neurobiological subtypes. You could mitigate for age differences by filtering out older subjects **before** clustering.

# MEDICATION-SEVERITY PARADOX

Other than the direct results, one interesting finding is that my 2 clusters have nearly identical severity (SRS scores) but massive statistically significant differences in medication usage. Even after the age filter is applied. This problem deserves investigating.

C0 and C1 show statistically significant IQ differences of a notable effect size

**PREMISE 1**

The higher IQ of Cluster C1 makes it easier for them to manage their symptoms

**PREMISE 2**

While the severity is equal for each cluster, C1 outwardly appears to be managing the symptoms better, which leads to lower levels of medication usage.

**CONCLUSION**

# FINAL COMMENTS



The DSM is concerned with behaviour, not neurobiology. My findings failing to show a clear link between neurobiology/behaviour support the choice to ignore neurobiology as it can be very complicated

**SUPPORTS BEHAVIOURAL DSM**



I have created multiple easy-to-use datasets. The PCA/CSV data in particular can be used within minutes.

**EASY TO USE**



Not only have I made a subset of the ABIDE dataset easier to use, but I have also made it easier to do ASD research in the future.

**CALL TO ACTION**

# **THANK'S FOR LISTENING**

*Any questions?*

