Written Report on Factors that Help Diagnose Autism

The Repos - Jeffrey Bohrer, Alexandra Green, Anna Zhang, Kevin Lee

2025-03-20

I. Introduction

Autism Spectrum Disorder (ASD) remains a highly prevalent condition despite modern strides made in medical technology. It is reported that nearly 2.2% of adults are affected by ASD, and growing awareness has led to an uptick in diagnoses, particularly in adults who went undiagnosed early in life (Hirota 2023). However, ASD screening tests for all age groups currently contain significant inaccuracies. For example, the most widely used toddler screening test, CHAT-R/FAs, was recently found to produce false negatives in 25% of cases. In contrast, the most commonly used adult autism screening test – the Autism-Spectrum Quotient (AQ) – was found to have limited predictive value in certain populations (Aishworiya 2023; Curnow 2023). Therefore, it has become critical to identify stronger predictors or explore underlying relationships to have more accurate tests and models to predict ASD in adults.

In this study, we will focus on identifying the features that most greatly affect the probability of being encouraged to pursue a diagnosis within this particular questionaire, as created by Prof. Fadi Thabtau of the Manukau Institute of Technology. The data was sourced from users of his app, ASDTests, which screens its users for potential indicators of autism using a ten-question survey. The data set being used will contain ten characteristics along with the answers of each individual to this survey. Because ASD is difficult to identify and can significantly impair an individual's quality of life, understanding the relationship between demographics, certain behaviors, and their association with autism could encourage individuals to seek diagnosis and gain the self-understanding they need. These adults who receive a positive diagnosis can then access the necessary resources for support. Our goal is to identify which aspects of this questionnaire are most closely associated with being encouraged to pursue an autism diagnosis.

For these purposes, both univariate and exploratory data analyses will be pursued as a means of assessing relevant avenues for further investigation and model fitting. These can be appreciated in the following pages.

i. Univariate EDA

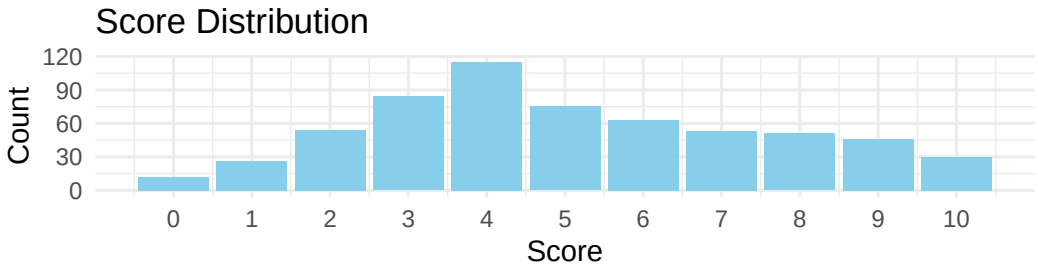


Figure 1: Distribution of total scores across dataset

As we can see in a quick summary of our data, the variable we are most interested in – the final score of users – ranges from 0 to 10, which makes sense as 10 questions can be answered either 'Yes' or 'No'. Generally, the scores are a bit right-skewed, with a single peak around 4 or 5 points. The mean score is 5.077, and the median is 5, both of which are relatively high considering that scores above 6 warrant further diagnostic evaluation. However, because suspecting a diagnosis is a reason for taking the test to begin with, these statistics are reasonable reflections of the test-taking population. We also observe that roughly 30% of test takers are encouraged to seek a diagnosis due to a score higher than 6, as calculated above. We also observe the IQR to be 4 points, as most test takers have a score between 3 and 7 points inclusive. In the context of the data, such a spread is reasonable, and no outliers exist which is understandable given the limited range of scores.

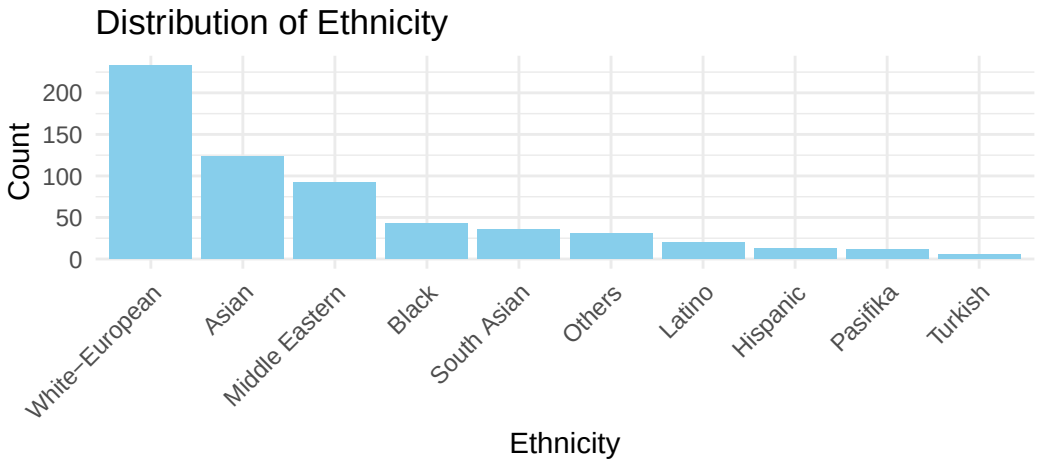


Figure 2: Distribution of ethnicities across dataset

We observe here that the most common ethnicity was White-Europeans, with over 200 observations in our dataset. Secondly and thirdly, we observe the Asian and Middle Eastern

populations to be above 100 and slightly below 100 observations, respectively. All other ethnicities have fewer than 50 observations in our data set, suggesting it may be more difficult to conclude about these populations. We can now check whether the proportion of individuals encouraged to seek a diagnosis varies by ethnicity:

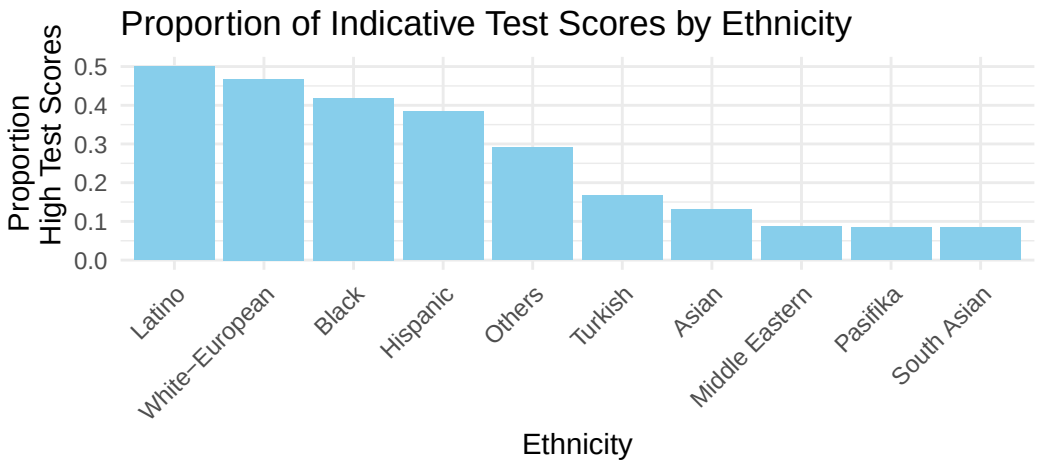


Figure 3: Proportion of indicative score across ethnicities

As we can see here, the population with the highest proportion of test takers receiving a high test score was the Latino population, with close to half achieving a score higher than a 6. Since this is a comparably uncommon population in our data set, it does raise questions of whether it is indicative of the greater Latino population or merely due to the smaller sample size in our data set. Additionally, the White-European, Black, and Hispanic populations also demonstrate a greater proportion of high test scores. The variation in these proportions according to ethnicity also warrants further exploration into how the distribution of test scores differs by ethnicity. Specifically, the questions remain of whether this means the rates of under- or over-diagnosis vary for different ethnic groups, and if so, what alterations would be necessary to ameliorate these errors.

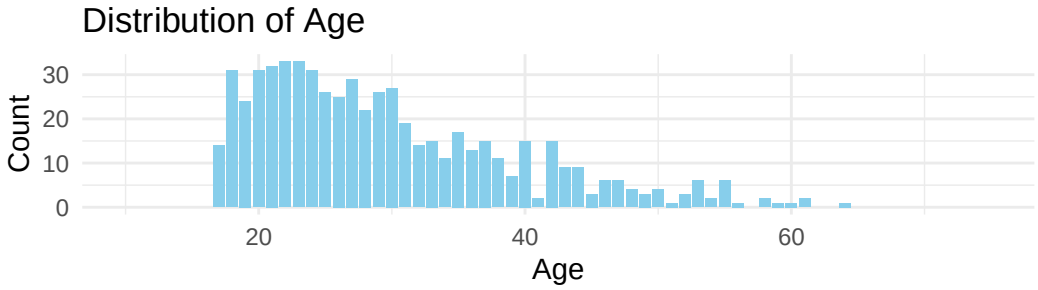


Figure 4: Age distribution across dataset

Generally, we observe the distribution of ages to possess a strong right-skewness, with a mean age of 30.22 and a median age of 27. There is a clear peak in the age distribution roughly around the early to mid-20s. The ages range from a minimum of 17 to a maximum of 383 – a false observation that should be filtered from the data. The IQR is 13 years, which is a fairly small spread given the range of ages, as we observe that the majority of test-takers are under 30 years old.

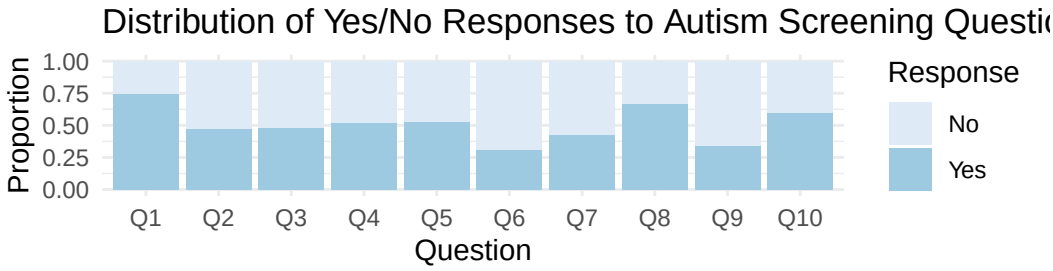


Figure 5: Distribution of Y/N responses for Q1-10. The questions are the following: Q1, noticing small sounds; Q2, finding it difficult to work out character intentions; Q3, finding it easy to read between lines; Q4, big picture-oriented; Q5, can tell if someone listening ot me is bored; Q6, can multitask; Q7, can tell feelings from faces; Q8, can go back to work when interrupted; Q9, enjoy collecting info on categories;Q10, find it difficult to work out people's intentions.

The two questions with the highest proportion of "Yes" responses were Q1 and Q8, suggesting that these behaviors are common among respondents. On the other hand, Q6 and Q9 had noticeably lower "Yes" responses, potentially indicating difficulties in such areas. Most questions, however, had roughly even proportions between "Yes" and "No" responses.

ii. Bivariate EDA

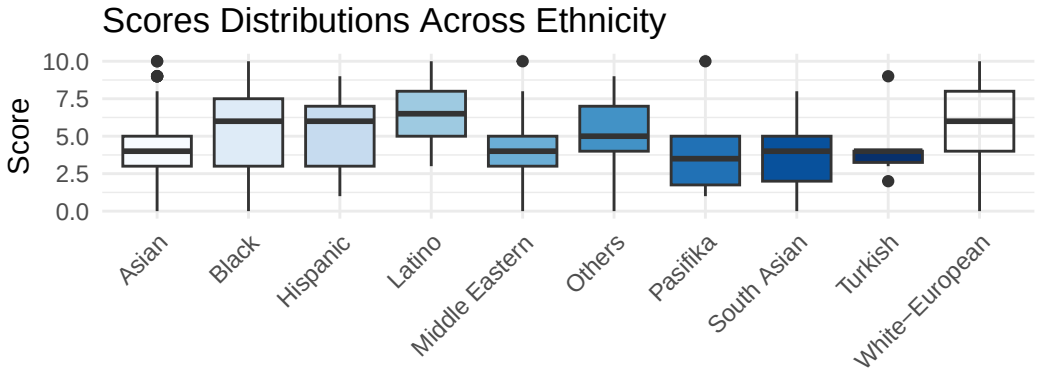


Figure 6: Score distribution across ethnicities

Here we present several graphs displaying bivariate relationships. With our first set of box plots, we find further evidence for our previous suspicion that test score distributions differ by ethnicity. Although most ethnicities have a median within the range of roughly 3 to 6, some ethnicities, like the White-European, Black, and Hispanic populations, demonstrate a greater spread through their larger IQRs, while the Turkish, Middle Eastern, and Asian populations are much more concentrated around their medians. However, most ethnicities appear to have values almost entirely across the range of 0 to 10 in their test scores.

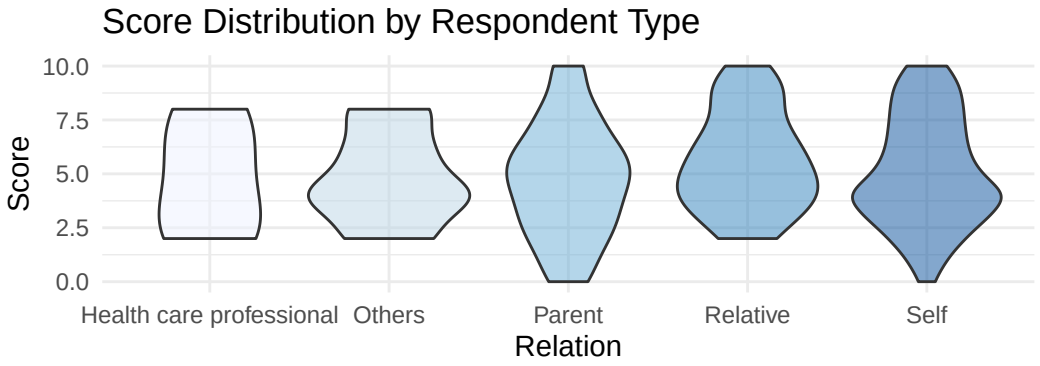


Figure 7: Distribution of scores by respondent type

Interestingly, we can also observe how the relationship between the test taker and the subject of the questions may lead to different distributions of test scores. In the case when it is filled out by a relative or health care professional for example, the observed test score is 2 or greater, while in the case of the test taker being the subject or the parent of the subject, some still received a score of 0. This could reflect how personal biases or relationships affect truthfulness during the test.

This initial exploration leads us to investigate one of these questions further, and particularly how our demographic data may impact the probability of being encouraged to seek a formal diagnosis. Though we cannot identify a definitive answer of whether it is due to social or cultural perceptions of autism within these subgroups or true difference in rates of its presence, it is worth identifying whether over- or under-diagnosis for certain subgroups is possible.

II. Methodology

i. Choosing Predictors

Firstly, a drop-in-deviance test between a logistic null model without predictor variables and a logistic model with a single predictor was systematically conducted across ethnicity, gender, presence of neonatal jaundice, and relationship as a means of assessing which predictors provide a statistically significant improvement in model fit against the null condition.These were chosen

first due to their "less direct" relationship with diagnosis. The hypothesis for this test can be observed below, where β_1 represents the coefficient for the associated predictor variable:

$$H_0 : \beta_1 = 0$$

$$H_a : \beta_1 \neq 0$$

The formulas for the models compared are the following:

$$Null : log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0$$

$$Alternative : log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0 + \beta_{predictor} x_{predictor}$$

The following table summarizes the results

Table 1: Non-obvious predictors drop-in-deviance test results

	Deviance	G statistic (respect to null)	p-value
Null	739.4	NA	NA
Ethnicity	645.3	94.1	0
Relation	738.194	1.206	0.877
Age	737.863	1.537	0.215
Gender	734.94	4.46	0.035
Jaundice	730.073	9.327	0.002

From the results of this test, it is evident that the parameters that are statistically significant for model fit are ethnicity, gender, and, quite surprisingly, neonatal jaundice when compared to a null model as the p-values are either significantly less than 0.05 (as in ethnicity) or slightly below it (gender and jaundice). We further investigate the correlation between answers to questions below, as a means of assessing if the test has any pitfalls regarding multicollinearity:

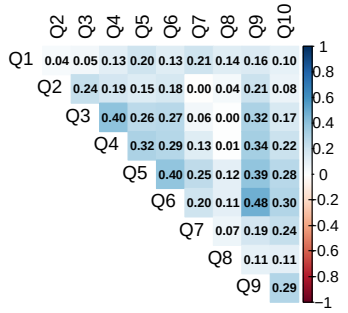


Figure 8: Correlation between AQ10 questions

By computing a correlation heat map among the ten behavior questions to check multicollinearity, we note that surprisingly, only a few questions have stronger correlations with one another. This is rather unexpected since we would expect many behaviors to be grouped together and more correlated (e.g., social behaviors). However, the highest correlation is between questions 6 and 9, which are the questions about multitasking and liking to collect information, respectively – this is also a rather surprising pairing. Due to this, Q3, Q5 and Q6 will be disregarded from modeling purposes as they showcase correlation with Q4, Q9, and Q10.

Considering this reality, principal component analysis was employed as a means of determining the question clusters that explain the highest variance and selecting the most significant in terms of loading for later inclusion in models (Szczęsna, 2022; finnstats, 2021).

Figure 9: Principal component analysis Q1-10 as a means of choosing questions that best explain the variance in the response

Ignoring questions 3, 5, and 6 due to issues with correlation, the questions that imply the most loading for principal components (i.e. the parameters that explain the most variance) are Q10, Q8, Q4, Q2, and Q1. Higher priority was given to questions that load the highest for the first 7 principal components, as these explain ~83% of variance in the response. Due to this, the answers to these questions will be employed as predictors (i.e finding it hard to figure out others intentions, ease of going back to work when interrupted, focus on whole picture, difficulty to understand characters in stories, and picking up on small sounds).

ii. Fitting the Model

Considering the two "groups" of variables we are working with: question answers and demogrpahic information. Two models will be fitted and compared in terms of a drop-in-deviance test: (a) a model that only accounts for the answers to the questions and (b) a model that accounts for demographic data on top of the previously selected questions. Subsequently, the models fitted will be of the forms:

$$Null : log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0 + \beta_{Q1}X_{Q1} + \beta_{Q2}X_{Q2} + \beta_{Q4}X_{Q4} + \beta_{Q8}X_{Q8} + \beta_{Q10}X_{Q10}$$

$$Alternative : log(\frac{\hat{\pi}}{1-\hat{\pi}}) = \beta_0 + \beta_{Q1}X_{Q1} + \beta_{Q2}X_{Q2} + \beta_{Q4}X_{Q4} + \beta_{Q8}X_{Q8} + \beta_{Q10}X_{Q10} +$$
$$\beta_{ethnicity}X_{ethnicity} + \beta_{gender}X_{gender} + \beta_{jaundice}X_{jaundice}$$

Subsequently, the hypotheses for the test will be the following:

$$H_0 : \beta_{ethnicity} = \beta_{gender} = \beta_{jaundice} = 0$$

$$H_a : \beta_{ethnicity}, \beta_{gender}, \beta_{jaundice} \neq 0, \text{for at least one} \beta_j$$

This is to assess whether demographic predictors significantly improve model fit with respect to predictors that most significantly explain the variability in final score and, ultimately, diagnosis. Additionally, AIC and BIC were included for additional comparison.

Table 2: Alternative and null model comparison in terms of deviance, AIC, and BIC

Model	Residual_Deviance	Df	AIC	BIC
Null Model	380.85	603	392.85	419.32
Alternative Model	358.48	592	392.48	467.48

It is evident that the alternative model (i.e. the one that account for demographic information on top of question data) performs significantly better exhibiting lower deviance, AIC. Nevrtheless, the BIC is higher. This is, however expected given the number of predictors as ethnicity is a categorical variable with 10 levels.

Table 3: Drop-in-deviance test result for alternative and null models

Test	G_stat	df	p_value
Drop-in-deviance (Chi-sq)	22.3746	11	0.0216

The drop-in-deviance tests corroborates that these results are statistically significant and that we can reject the null hypothesis since the p-value is less than 0.05. Hence, the alternative model will be further assessed.

Table 4: Output for alternative model

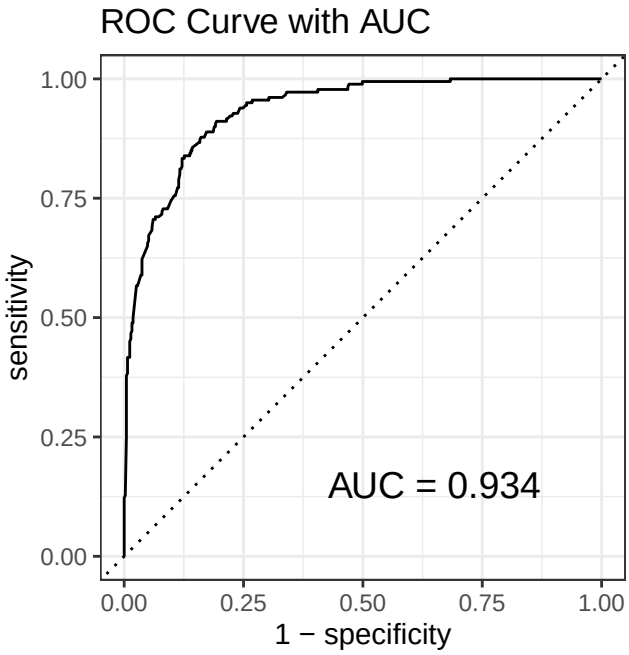
term	estimate	std.error	statistic	p.value
(Intercept)	-7.906	0.784	-10.085	0.000
small_sounds	2.263	0.419	5.399	0.000
difficult_to_understand_char	1.695	0.287	5.901	0.000
focus_on_whole_picture	2.636	0.327	8.070	0.000
i_can_go_back_to_work_when_interrupted	1.761	0.327	5.380	0.000
i_find_it_hard_to_figure_out_others_intentions	2.228	0.342	6.517	0.000
ethnicityAsian	-0.860	0.405	-2.124	0.034
ethnicityBlack	0.661	0.516	1.281	0.200
ethnicityHispanic	0.848	0.888	0.955	0.340
ethnicityLatino	-0.133	0.644	-0.207	0.836
ethnicityMiddle Eastern	-1.288	0.504	-2.559	0.011
ethnicityOthers	-0.473	0.585	-0.808	0.419
ethnicityPasifika	-1.620	1.528	-1.061	0.289
ethnicitySouth Asian	-1.568	0.763	-2.056	0.040
ethnicityTurkish	-1.093	1.733	-0.631	0.528
genderm	-0.162	0.278	-0.583	0.560
jundiceyes	0.343	0.435	0.789	0.430

The intercept for this model represents that, for an individual who answered no to all five of the AQ10 behavior questions relevant to the model, who identifies as White-European, is female and did not have neonatal jaundice, the odds of being classified as high probability for autism are 0.000368, calculated from exp(-7.906).

The estimates for the AQ10 showcase strong statistical associations with higher odds of ASD. Holding all else constant, individuals who reported picking up on small sounds have odds of being characterized as high probability for ASD multiply by a factor of 9.583, compared to when they report no to the question. Likewise, difficulty understanding character intentions (OR = 5.447), tendency to focus on whole picture (OR = 13.957), ease of going back to work when interrupted (OR = 5.818), and finding it hard to understand others' intentions (OR =9.281) all strongly correlate with high probability of ASD.

Among demographic predictors, people who identify as Asian, Middle Eastern, or South Asian are significantly less likely to be categorized as high probability for ASD with odd ratios (ORs) of approximately 0.42, 0.28, and 0.21 respectively, compared to the baseline category of White-European.

III. Results



The area under the curve is 0.934, which for a model fitted to demographic data and questionaire data. This is really impressive, but over-fitting is certainly a concern. Nevertheless, this indicates that further investigation into how these the AQ10 may over- or under-diagnose certain population demographics could still be worth pursuing. Thus, broader comparisons would be extremely interesting.

For model assumptions, logistic regression requires the assumption of log-odds possessing linearity, randomness, and independence of observations. It is worth noting that of these conditions, randomness is unlikely to be satisfied due to the nature of the data set – it was collected by people who already were likely to suspect an ASD diagnosis, as they are those who took the test voluntarily. However, our other two assumptions are likely to hold for this model, and we still determine our questions, model, and results worth further investigation. We do not need to check linearity as none of our predictors are quantitative. We can also assume independence as there are no anticipated temporal or spatial relationships between individuals taking the questionnaire.

Once more, a 0.934 AUC for a model that only accounts for 5 questions of the test and demographic data showcases that further analysis should be pursued to assess whether the test is biased towards certain populations. This is especially true given that the odds are notably high for individuals that identify as White-European.

IV. References

- Aishworiya, R., Kim, V., MA, Stewart, S., Hagerman, R., & Feldman, H. M. (2023). Meta-analysis of the Modified Checklist for Autism in Toddlers, Revised/Follow-up for Screening. *PEDIATRICS*, *151*(6). https://doi.org/10.1542/peds.2022-059393
- Curnow, E., Utley, I., Rutherford, M., Johnston, L., & Maciver, D. (2023). Diagnostic assessment of autism in adults – current considerations in neurodevelopmentally informed professional learning with reference to ADOS-2. *Frontiers in Psychiatry*, *14*. https://doi.org/10.3389/fpsyt.2023.1258204
- Hirota, T., & King, B. H. (2023). Autism spectrum Disorder. *JAMA*, *329*(2), 157. https://doi.org/10.1001/jama.2022.23661
- Marin, Z. (2021, April 26). GLM fit: Algorithm did not converge – How to fix it. Statology. https://www.statology.org/glm-fit-algorithm-did-not-converge/
- finnstats. (2021, May 14). Principal component analysis (PCA) in R. https://www.r-bloggers.com/2021/05/principal-component-analysis-pca-in-r/
- Szczęsna, K. (2022). PCA in R. RPubs. https://rpubs.com/KarolinaSzczesna/862710

::: Before you submit, make sure your code chunks are turned off with `echo: false` and there are no warnings or messages with `warning: false` and `message: false` in the YAML. :::

V. Additonal Materials