COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

Using Neural Networks to Show the Lack of Cultural Relativity in Autism Screeners

By Zachary Zivalich

Phone: 661-400-7890 | Email: zlz2001@columbia.edu

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

Aim

The purpose of this study is to address the potential bias within autism screeners as a result of lack of cultural relativity. I hypothesize that common autistic traits manifest differently across gender and sociocultural groups, which is a reality not accounted for by screeners at this moment.

Abstract

Autism Spectrum Disorder is a developmental disability that has a prevalence rate between .9 to 1.7% of the U.S. population with similar rates in developed countries. Epidemiological surveys show that the disorder is not associated with racial or genetic factors, yet this is in direct contrast with observed prevalence rates - where white children are significantly more likely to be diagnosed than Black or Hispanic children. Despite this, little has been done to explain how this imbalance came to be. I address the skewed prevalence rates by looking at methodological bias in Autism Screeners. Autism screeners are questionnaires acting as a first step in diagnosis – truncating individuals seeking a diagnosis by scoring autistic traits. I use an open source dataset, filled out by app takers from all over the world (n=2496), to address the issue of bias within the questions of autism screeners. These screeners are designed to gauge autistic traits. Yet, trait scores varied as a product of race and gender, with white men scoring higher overall. I use a logistic regression to show the likelihood of diagnosis, which is higher for white app users. Lastly, I build a neural network to predict whether a test taker is a white male or not, providing evidence that the screener's questions have been designed around more available white data. Autistic traits are the de facto symptom of diagnosis. It is possible that the dearth of nonwhite children in early autism studies lead to said traits being based on predominantly white male characteristics.

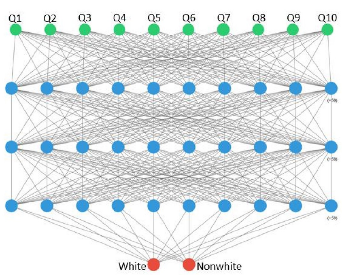


Fig. 1: Neural Network

Data and Methods

Two datasets were used for this study- one filled out by the guardians of toddlers and another by adults. These datasets are provided by Dr. Fadi Thabtah who collected it via smartphone app (ASD Tests) which is easily downloaded from the Android and Apple app store. The participants fill out varied autism screeners, which are based on the Q-CHAT and Autism Quotient surveys. Participants receive a score which determines whether they are recommended to pursue professional diagnosis. The dataset is prone to selection bias, as anyone can download the app and fill out the survey, meaning statistically it cannot be applied to a wider population. However, the prevalence rates for the samples (~70% and ~30%) are well above the national average of 1.5%. Therefore, it is likely that this dataset is comprised of participants who are much more likely to be autistic. I run a logistic regression to see which ethnicities and genders are more or less likely to reach diagnosis. Then, I compare the average scores of white and nonwhite test takers through a two-sample t-test. Finally, I run a neural network to see if a deep learning algorithm can pick up on a score pattern only white male participants display.

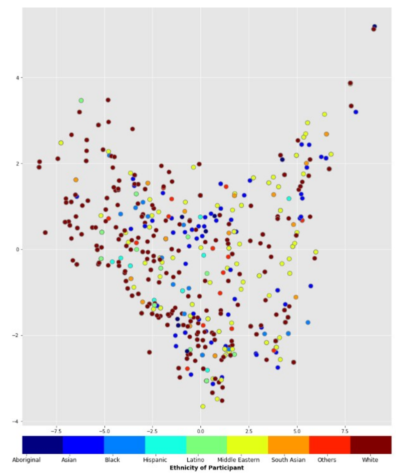


Fig. 2: Isometric Map of Adult Score Patterns

Results

Across both age groups, white toddlers and adults scored higher on each question then nonwhites. Scores were compared through a two-sample two-way T test. Every T-value was statistically significant but one question (A8), meaning there was below a 5% probability the difference in means were found by chance. However, this did not mean every ethnicity groups scored less than white test takers.

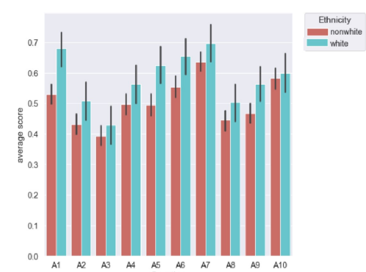


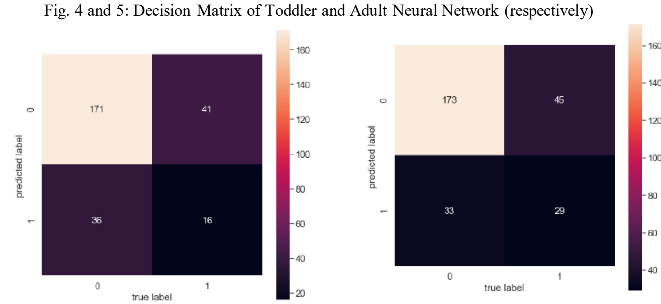
Fig. 3: Average scores for Adult White vs Nonwhite

Black and Latino Male test takers were more likely to receive a recommendation on both questionnaires then White males. However, these three groups were very likely to be diagnosed when compared with other groups. For example, Black and White Male toddlers were 2.05 and .85 log-odds more likely to be diagnosed then Middle Eastern Male toddlers.

	coef	std err	z	P>|z|
Intercept	0.4082	0.117	3.480	0.001
male:C(Ethnicity)[Hispanic]	0.8239	0.445	1.850	0.064
male:C(Ethnicity)[Latino]	1.4636	0.769	1.904	0.057
male:C(Ethnicity)[White European]	0.8048	0.193	4.165	0.000
male:C(Ethnicity)[asian]	0.4185	0.185	2.260	0.024
male:C(Ethnicity)[black]	2.0195	0.614	3.291	0.001
male:C(Ethnicity)[middle eastern]	-0.0335	0.228	-0.147	0.883
male:C(Ethnicity)[mixed]	-0.4082	1.007	-0.405	0.685
male:C(Ethnicity)[south asian]	0.1026	0.347	0.296	0.767

Table 1: Multivariate Logit Regression of Male Toddlers

Fig. 4 and 5: Decision Matrix of Toddler and Adult Neural Network (respectively)



	accuracy	f1_score	precision	recall
0	0.708333	0.554904	0.553394	0.557148

	accuracy	f1_score	precision	recall
0	0.721429	0.621254	0.615849	0.63066

A Neural network was then designed to classify test takers into nonwhite male and white male categories across both questionnaires based only on the participants score patterns. After some tuning, It was able to predict which class each participant belonged to with an accuracy of ~70% and an f1_score of ~60%. Just by training on how a participants answers the 10 questions, the model could guess the ethnicity to a moderate degree of accuracy. However, it more so excelled at classifying who was not a white male then who was.

Conclusion

•White scored higher than nonwhite test takers on average across every question on both screeners.

•Latino and Black male test takers were more likely to receive a diagnosis then white males. However, all three were much more likely to receive a diagnosis then Nonwhite, Black, or Hispanic males.

•The likelihood of diagnostic recommendation varying by ethnicity and gender in differing degree.

• The neural network was able to predict white male test takers to ~70% accuracy with ~60% f1_score for both datasets, signaling at least one measurable pattern that is dependent on race.

Discussion

As the awareness of autism continues to rise, psychiatric practitioners need to better understand how the behavioral symptomology of autism manifests itself as a product of socio-cultural factors. It is important that autistic children get access to the resources necessary to excel at an early age so that they do not develop the depression and suicidal tendencies shown to appear in higher rates in autistics diagnosed after childhood. Despite the selection bias in the sample, this sample's specific qualities do well in simulating the barriers of diagnosis for those who believe they are their children may be autistic. When the dataset is understood to be composed of participants seeking a diagnosis because of self realized symptomology, it is easy to see that certain groups of people struggle to receive the diagnosis they need. Going forward, follow up studies can select a true random sample to see the sociocultural differences in the wider population of autistic adults and toddlers to verify these findings.

Acknowledgements

Data comes from Dr. Fadi Thabhtah of creator of ASD tests and co-creator of Autism AI. His website can be found here https://fadifayez.com/

Using Neural Networks to Show the Lack of Cultural Relativity in Autism Screeners

By Zachary Zivalich

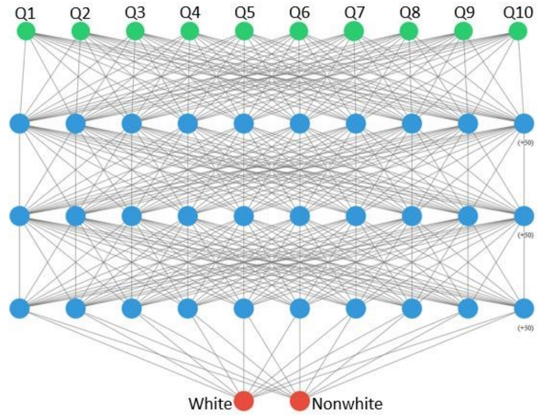
Phone: 661-400-7890
Email: zlz2001@Columbia.edu

Spring 2020

Autism Spectrum Disorder is a developmental disability that has a prevalence rate between .9 to 1.7% of the U.S. population with similar rates in developed countries. Epidemiological surveys show that the disorder is not associated with racial or genetic factors, yet this is in direct contrast with observed prevalence rates - where white children are significantly more likely to be diagnosed than Black or Hispanic children. Despite this, little has been done to explain how this imbalance came to be. I address the skewed prevalence rates by looking at methodological bias in Autism Screeners. Autism screeners are questionnaires acting as a first step in diagnosis – truncating individuals seeking a diagnosis by scoring autistic traits. I use an open source dataset, filled out by app takers from all over the world (n=2496), to address the issue of bias within the questions of autism screeners. These screeners are designed to gauge autistic traits. Yet, trait scores varied as a product of race and gender, with white men scoring higher overall. I use a logistic regression to show the likelihood of diagnosis, which is higher for white app users. Lastly, I build a neural network to predict whether a test taker is white or not, providing evidence that the screener's questions have been designed around more available white data. Autistic traits are the de facto symptom of diagnosis. It is possible that the dearth of nonwhite children in early autism studies lead to said traits being based on predominantly white male characteristics.

Zachary Zivalich is a University of Caliofornia San Diego Alumni where he majored in cognitive neuroscience and worked for the HEAL lab under Karen Dobkins. He also spent a year building and conducting research trails at Dexcom. He hopes to marry his neuroscience and data background to pursue of a PhD in neuroscience.

Neural Network predicting white versus nonwhite test takers



Adult White Versus Nonwhite Test Scores

Question	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Score
Nonwhite	0.688552	0.415825	0.429293	0.454545	0.478114	0.242424	0.415825	0.651515	0.286195	0.553872	4.616162
White	0.79771	0.566794	0.59542	0.633588	0.570611	0.419847	0.475191	0.645038	0.461832	0.666031	5.832061
T-Value	4.216574	5.091716	5.620109	6.098239	3.101777	6.37145	1.993907	-0.22609	6.134533	3.865248	8.081721
P-Value	0.000027	< .00001	< .00001	< .00001	0.001972	< .00001	0.046406	0.821243	< .00001	0.000117	< .00001

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

Data and Methods

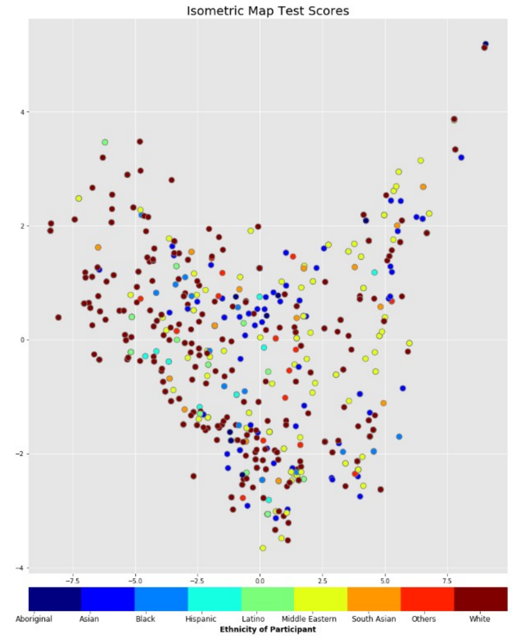
I look at one dataset filled out by the guardians of toddlers and another by adults for the study. These datasets are provided by Dr. Fadi Thabtah who collected it via smartphone app which is easily downloaded from the Android and Apple app store. The participants fill out varied autism screeners, which are based on the Q-CHAT and Autism Quotient surveys. Participants receive a score which determines whether they are recommended to pursue professional diagnosis. The dataset is prone to selection bias, as anyone can download the app and fill out the survey, meaning statistically it cannot be applied to a wider population. However, the prevalence rates for the samples (~70% and ~30%) are well above the national average of 1.5%. Therefore, it is likely that this dataset is comprised of participants who are much more likely to be autistic. I run a logistic regression to see which ethnicities and genders are more or less likely to reach diagnosis. Then, I compare the average scores of white and nonwhite test takers through a two-sample t-test. Finally, I run a neural network to see if a deep learning algorithm can pick up on a score pattern only white participants display.

Key Results

White's scored higher then test takers on average across every question on both screeners. Interestingly, Latino and Black male test takers were are more likely to receive a diagnosis then white males. However, all three were very likely to receive a diagnosis. Interestingly, the likelihood of diagnosis varied by ethnicity in varied degrees. With most Asian test takers scoring below the threshold of diagnosis. The neural network was able to predict white male test takers to 70% accuracy with 62% recall for both datasets, signaling a measurable pattern that is dependent on race.

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

Isometric Clusters of Adult Score Patterns



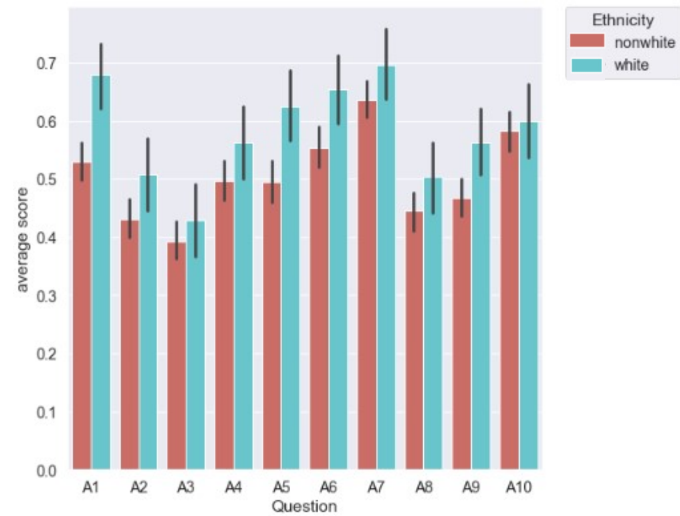
Likelihood of Male Toddlers of Varied Ethnicities Recieving Diagnostic Recommendation

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	0.4246	0.117	3.631	0.000	0.195	0.654
male:C(Ethnicity)[Hispanic]	0.8076	0.445	1.814	0.070	-0.065	1.680
male:C(Ethnicity)[Latino]	1.4472	0.769	1.883	0.060	-0.059	2.953
male:C(Ethnicity)[Pacifica]	0.9617	1.124	0.856	0.392	-1.242	3.165
male:C(Ethnicity)[White European]	0.7885	0.193	4.085	0.000	0.410	1.167
male:C(Ethnicity)[asian]	0.4021	0.185	2.174	0.030	0.040	0.765
male:C(Ethnicity)[black]	2.0032	0.614	3.265	0.001	0.801	3.206
male:C(Ethnicity)[middle eastern]	-0.0499	0.228	-0.219	0.827	-0.497	0.397
male:C(Ethnicity)[mixed]	-0.4246	1.007	-0.422	0.673	-2.398	1.549
male:C(Ethnicity)[south asian]	0.0863	0.347	0.249	0.804	-0.594	0.766

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

Key Results Cont.

Across both age groups, white toddlers and adults scored higher on each question then nonwhites. Scores were compared through a two-sample two-way T test. Every T-value was statistically significant but one question (A8), meaning there was below a 5% probability the difference in means were found by chance. However , this did not mean every ethnicity groups scored less than white test takers.



Key Results Cont

Black and Latino Male test takers were more likely to receive a recommendation on both questionnaires then White males. However, these three groups were very likely to be diagnosed when compared with other groups. For example, Black and White Male toddlers were 2.05 and .85 log-odds more likely to be diagnosed then Middle Eastern Male toddlers.

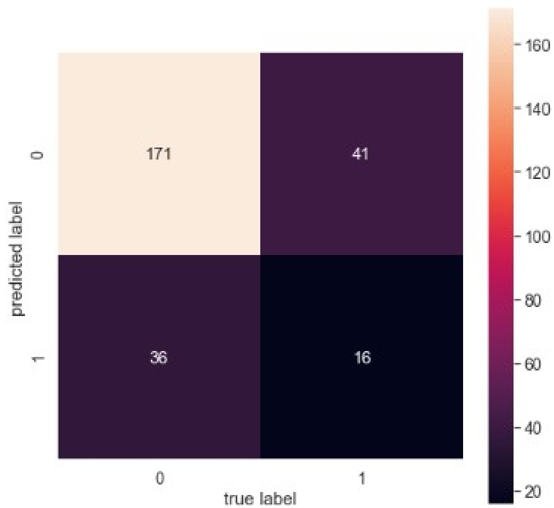
```
Adult Females
Intercept                                0.4100
female:C(Ethnicity)[aboriginal]          1.6260
female:C(Ethnicity)[asian]               0.4880
female:C(Ethnicity)[black]               2.8150
female:C(Ethnicity)[hispanic]            2.4390
female:C(Ethnicity)[latino]              3.9030
female:C(Ethnicity)[middle eastern]      0.3440
female:C(Ethnicity)[others ]             2.4390
female:C(Ethnicity)[south asians]        0.1220
female:C(Ethnicity)[white]               2.1060
dtype: float64
Adult Males
Intercept                              0.5490
male:C(Ethnicity)[aboriginal]          0.1820
male:C(Ethnicity)[asian]               0.3110
male:C(Ethnicity)[black]               1.3930
male:C(Ethnicity)[hispanic]            0.9940
male:C(Ethnicity)[latino]              1.6700
male:C(Ethnicity)[middle eastern]      0.2160
male:C(Ethnicity)[others ]             0.4050
male:C(Ethnicity)[south asians]        0.2280
male:C(Ethnicity)[white]               1.2930
dtype: float64
```

```
Toddler Females
Intercept                                2.5610
female:C(Ethnicity)[Hispanic]            0.7810
female:C(Ethnicity)[Latino]              0.6830
female:C(Ethnicity)[White European]      0.8750
female:C(Ethnicity)[asian]               1.1950
female:C(Ethnicity)[black]               0.1780
female:C(Ethnicity)[middle eastern]      0.2600
female:C(Ethnicity)[mixed]               1.1720
female:C(Ethnicity)[south asian]         0.5860
dtype: float64
Toddler Males
Intercept                              1.5040
male:C(Ethnicity)[Hispanic]            2.2790
male:C(Ethnicity)[Latino]              4.3210
male:C(Ethnicity)[White European]      2.2360
male:C(Ethnicity)[asian]               1.5200
male:C(Ethnicity)[black]               7.5350
male:C(Ethnicity)[middle eastern]      0.9670
male:C(Ethnicity)[mixed]               0.6650
male:C(Ethnicity)[south asian]         1.1080
dtype: float64
```

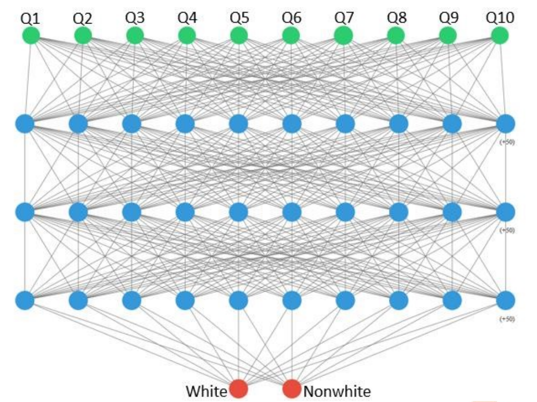

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

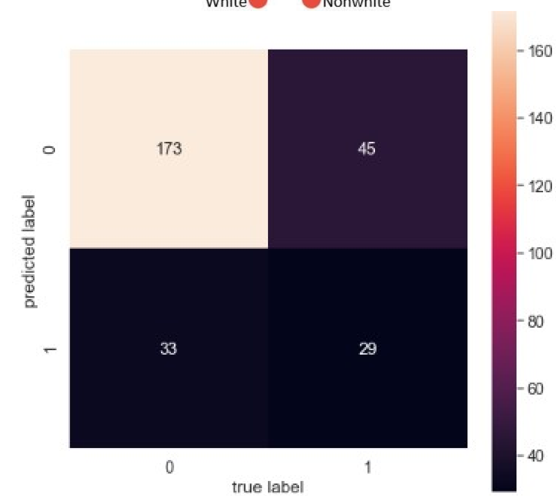
Key Results Cont.

A Neural network was then designed to classify test takers into nonwhite male and white male categories across both questionnaires based only on the participants score patterns. After some tuning, It was able to predict which class each participant belonged to with an accuracy of ~70% and an f1_score of ~60%. Just by training on how a participants answers the 10 questions, the model could guess the ethnicity to a moderate degree of accuracy. However, it more so excelled at classifying who was not a white male then who was.



	accuracy	f1_score	precision	recall
0	0.708333	0.554904	0.553394	0.557148





	accuracy	f1_score	precision	recall
0	0.721429	0.621254	0.615849	0.63086

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

Conclusion

•White scored higher than nonwhite test takers on average across every question on both screeners.

•Latino and Black male test takers were more likely to receive a diagnosis then white males. However, all three were much more likely to receive a diagnosis then Nonwhite, Black, or Hispanic males.

•The likelihood of diagnostic recommendation varying by ethnicity and gender in differing degree.

• The neural network was able to predict white male test takers to ~70% accuracy with ~60% f1_score for both datasets, signaling at least one measurable pattern that is dependent on race.

Conclusion

As the awareness of autism continues to rise, psychiatric practitioners need to better understand how the behavioral symptomology of autism manifests itself as a product of socio-cultural factors. It is important that autistic children get access to the resources necessary to excel at an early age so that they do not develop the depression and suicidal tendencies shown to appear in higher rates in autistics diagnosed after childhood. Despite the selection bias in the sample, this sample's specific qualities do well in simulating the barriers of diagnosis for those who believe they are their children may be autistic. When the dataset is understood to be composed of participants seeking a diagnosis because of self realized symptomology, it is easy to see that certain groups of people struggle to receive the diagnosis they need. Going forward, follow up studies can select a true random sample to see the sociocultural differences in the wider population of autistic adults and toddlers to verify these findings.

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

Using Neural Networks to Show the Lack of Cultural Relativity in Autism Screeners

By Zachary Zivalich

Phone: 661-400-7890 | Email: zlz2001@columbia.edu

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

COLUMBIA UNIVERSITY | QMSS
QUANTITATIVE METHODS IN THE SOCIAL SCIENCES
MASTERS PROGRAM IN THE INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY

Aim

The purpose of this study is to address the potential bias within autism screeners as a result of lack of cultural relativity. I hypothesize that common autistic traits manifest differently across gender and sociocultural groups, which is a reality not accounted for by screeners at this moment.

Abstract

Autism Spectrum Disorder is a developmental disability that has a prevalence rate between .9 to 1.7% of the U.S. population with similar rates in developed countries. Epidemiological surveys show that the disorder is not associated with racial or genetic factors, yet this is in direct contrast with observed prevalence rates - where white children are significantly more likely to be diagnosed than Black or Hispanic children. Despite this, little has been done to explain how this imbalance came to be. I address the skewed prevalence rates by looking at methodological bias in Autism Screeners. Autism screeners are questionnaires acting as a first step in diagnosis – truncating individuals seeking a diagnosis by scoring autistic traits. I use an open source dataset, filled out by app takers from all over the world (n=2496), to address the issue of bias within the questions of autism screeners. These screeners are designed to gauge autistic traits. Yet, trait scores varied as a product of race and gender, with white men scoring higher overall. I use a logistic regression to show the likelihood of diagnosis, which is higher for white app users. Lastly, I build a neural network to predict whether a test taker is a white male or not, providing evidence that the screener's questions have been designed around more available white data. Autistic traits are the de facto symptom of diagnosis. It is possible that the dearth of nonwhite children in early autism studies lead to said traits being based on predominantly white male characteristics.

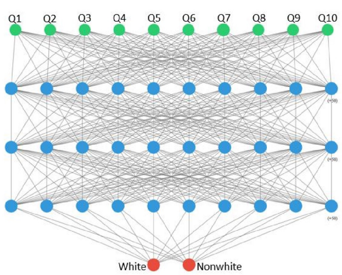


Fig. 1: Neural Network

Data and Methods

Two datasets were used for this study- one filled out by the guardians of toddlers and another by adults. These datasets are provided by Dr. Fadi Thabtah who collected it via smartphone app (ASD Tests) which is easily downloaded from the Android and Apple app store. The participants fill out varied autism screeners, which are based on the Q-CHAT and Autism Quotient surveys. Participants receive a score which determines whether they are recommended to pursue professional diagnosis. The dataset is prone to selection bias, as anyone can download the app and fill out the survey, meaning statistically it cannot be applied to a wider population. However, the prevalence rates for the samples (~70% and ~30%) are well above the national average of 1.5%. Therefore, it is likely that this dataset is comprised of participants who are much more likely to be autistic. I run a logistic regression to see which ethnicities and genders are more or less likely to reach diagnosis. Then, I compare the average scores of white and nonwhite test takers through a two-sample t-test. Finally, I run a neural network to see if a deep learning algorithm can pick up on a score pattern only white male participants display.

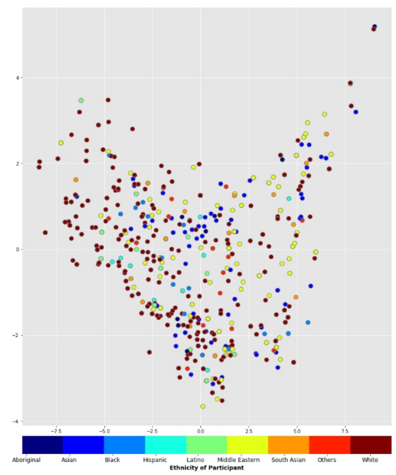


Fig. 2: Isometric Map of Adult Score Patterns

Results

Across both age groups, white toddlers and adults scored higher on each question then nonwhites. Scores were compared through a two-sample two-way T test. Every T-value was statistically significant but one question (A8), meaning there was below a 5% probability the difference in means were found by chance. However , this did not mean every ethnicity groups scored less than white test takers.

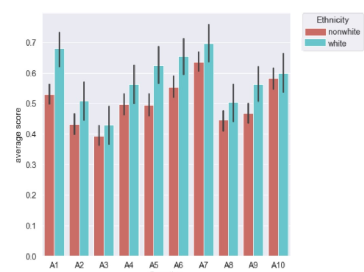


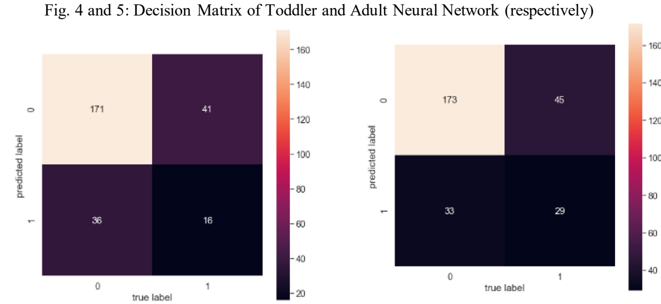
Fig. 3: Average scores for Adult White vs Nonwhite

Black and Latino Male test takers were more likely to receive a recommendation on both questionnaires then White males. However, these three groups were very likely to be diagnosed when compared with other groups. For example, Black and White Male toddlers were 2.05 and .85 log-odds more likely to be diagnosed then Middle Eastern Male toddlers.

	coef	std err	z	P>|z|
Intercept	0.4082	0.117	3.480	0.001
male:C(Ethnicity)[Hispanic]	0.8239	0.445	1.850	0.064
male:C(Ethnicity)[Latino]	1.4636	0.769	1.904	0.057
male:C(Ethnicity)[White European]	0.8048	0.193	4.165	0.000
male:C(Ethnicity)[asian]	0.4185	0.185	2.260	0.024
male:C(Ethnicity)[black]	2.0195	0.614	3.291	0.001
male:C(Ethnicity)[middle eastern]	-0.0335	0.228	-0.147	0.883
male:C(Ethnicity)[mixed]	-0.4082	1.007	-0.405	0.685
male:C(Ethnicity)[south asian]	0.1026	0.347	0.296	0.767

Table 1: Multivariate Logit Regression of Male Toddlers

Fig. 4 and 5: Decision Matrix of Toddler and Adult Neural Network (respectively)



	accuracy	f1_score	precision	recall
0	0.708333	0.554904	0.553394	0.557148

	accuracy	f1_score	precision	recall
0	0.721429	0.621254	0.615849	0.63066

A Neural network was then designed to classify test takers into nonwhite male and white male categories across both questionnaires based only on the participants score patterns. After some tuning, It was able to predict which class each participant belonged to with an accuracy of ~70% and an f1_score of ~60%. Just by training on how a participants answers the 10 questions, the model could guess the ethnicity to a moderate degree of accuracy. However, it more so excelled at classifying who was not a white male then who was.

Conclusion

•White scored higher than nonwhite test takers on average across every question on both screeners.

•Latino and Black male test takers were more likely to receive a diagnosis then white males. However, all three were much more likely to receive a diagnosis then Nonwhite, Black, or Hispanic males.

•The likelihood of diagnostic recommendation varying by ethnicity and gender in differing degree.

• The neural network was able to predict white male test takers to ~70% accuracy with ~60% f1_score for both datasets, signaling at least one measurable pattern that is dependent on race.

Discussion

As the awareness of autism continues to rise, psychiatric practitioners need to better understand how the behavioral symptomology of autism manifests itself as a product of socio-cultural factors. It is important that autistic children get access to the resources necessary to excel at an early age so that they do not develop the depression and suicidal tendencies shown to appear in higher rates in autistics diagnosed after childhood. Despite the selection bias in the sample, this sample's specific qualities do well in simulating the barriers of diagnosis for those who believe they are their children may be autistic. When the dataset is understood to be composed of participants seeking a diagnosis because of self realized symptomology, it is easy to see that certain groups of people struggle to receive the diagnosis they need. Going forward, follow up studies can select a true random sample to see the sociocultural differences in the wider population of autistic adults and toddlers to verify these findings.

Acknowledgements

Data comes from Dr. Fadi Thabhtah of creator of ASD tests and co-creator of Autism AI. His website can be found here https://fadifayez.com/