

Explorative Data Analysis on Autism Project

Yunfei Yu, 1005976918

Feb.10, 2022

Introduction

Autism is a neuro-cognitive condition that influences social interaction, communication and behaviours (Botha et al., 2020). While autistic people think of autism as a value-neutral trait, the stereotypes of autism are predominantly negative and result in the stigmatization of autism and autistic people. A commonly-used strategy for autistic people to manage stigma is stigma concealment, or the attempts to hide autistic traits from others (Botha et al., 2020; Jackson & Mohr, 2016). Research shows that the concealment of autistic traits negatively impacts mental health among autistic people (Botha et al., 2020). In addition, connectedness to the autistic community is found to be a buffer against minority stress. Thus, our collaborator is interested in the relationship between autistic traits concealment, autistic community connectedness, stigma experience and mental health (specifically depression) among autistic people. She hypothesized a serial mediating relationship among the variables, where autistic trait concealment influences the severity of depressive symptoms both directly and indirectly via (i) autistic community connectedness and (ii) stigma experience.

Our collaborator collected data from 409 participants aged between 18 and 67. The participants filled out a survey consisting of 7 questionnaires, each with up to 23 items, and the demographics survey. The 4 questionnaires corresponding to the variables of interest are:

- Camouflaging Autistic Traits Questionnaire (CATQ), measuring concealment of autistic identity,
- Autistic Community Connectedness Measure (ACC),
- Stigma Experience (SE), and
- Center for Epidemiological Studies Depression Scale (CES-D), measuring depressive symptoms.

Each item has pre-determined options, and each option is assigned a score. A participant's total score across all items in a questionnaire quantifies their level on the corresponding variable.

Data cleaning

Removing Redundant Columns

The raw data contains 96 columns and 409 observations. We first removed the “administrative” columns because there is no variability and they do not contain information on the measures of interest or the participants. The “administrative” variables include the columns on

- (i) completion of the survey (every participant fully completed),
- (ii) the date of the survey (every participant did the survey on 2022/8/24),
- (iii) anonymous distribution channel,
- (iv) language (all participants are English speakers),
- (v) whether participants paid attention (every participant paid attention and answered carefully)
- (vi) consent statement (all participants signed the consent), and
- (vii) `ExternalReference`, which is an empty column.

Checking Missing Values

There are 5 observations with missing values in questionnaire items and/or demographic variables. 4 of the missing values occur in the visible minority indicator variable, which we replaced with an “unknown”. One of the participants had no score recorded for the SICS6 item, and we removed this observation because we cannot obtain a total SICS score for analysis.

Coding CES-D

In the raw data, items for all questionnaires except CES-D are coded where the pre-determined options for each item are converted to the corresponding numerical score to aid statistical analysis. Thus, the next step is to code CES-D items according to the coding manual provided by our collaborator.

We coded the direct-scored CES-D items (all items except items 4, 8, 12 and 16) as follows:

- “Rarely or none of the time (less than 1 day)” is assigned a score of 0,
- “Some or a little of the time (1-2 days)” is assigned a score of 1,
- “Occasionally or a moderate amount of time (3-4 days)” is assigned a score of 2, and
- “Most or all of the time (5-7 days)” is assigned a score of 3.

The coding is reversed for reversed-scored CES-D items (items 4, 8, 12 and 16). For example, a “Rarely or none of the time (less than 1 day)” response is assigned a score of 3 and a “Most or all of the time (5-7 days)” is assigned a score of 0.

Reverse-Coding

The reverse-scored items in the questionnaires other than CES-D are not properly coded. We thus reverse-coded these items according to the coding manual. For example, participants who scored 4 on a reverse-scored item are now re-assigned a score of 0, and who had a score of 3 are re-assigned a 1, etc.

Calculating Total Scores for Each Questionnaire

All items in each questionnaire are properly coded by this step. For each participant, we added the scores for each item together to obtain a total score for each questionnaire.

Cleaning Demographic Variables

After coding the questionnaire items and calculating total scores, we cleaned demographic variables such as geographical regions where the participants come from, diagnosis method, employment and education.

We first grouped regions with keywords like “UK”, “United Kingdom”, “England”, “Wales”, and “Scotland” into a larger category of “UK”. We then abbreviated the regions and dropped irrelevant characters for simplicity. For example, “Western europe :)” is converted into “W-EUR”.

Next, we cleaned the diagnosis variable. The diagnosis variable originally had three values: (i) “I am self-diagnosed, and I am not seeking a formal diagnosis”, (ii) “I have a formal diagnosis of autism” and (iii) “I am self-diagnosed, and I am seeking a formal diagnosis” which contains information on both diagnosis method and whether a participant is seeking a formal diagnosis. Thus, we expanded the diagnosis variable into two indicator variables: `self_dx` indicates whether a participant is self-diagnosed (1) or formally diagnosed (0), and `seeking_formal` indicates whether a participant is seeking formal diagnosis (1) or not (0).

Similarly, the employment variable originally included information on (i) employment status, (ii) whether a participant is looking for work, and (iii) whether a participant is a student. For example, the values of the employment variable would be “Unemployed and looking for work” or “Employed full-time, Student”. We instead created three indicators where `employed` indicates whether a participant is employed, `student` indicates whether a participant is a student and `finding_job` indicates whether a participant is seeking a job.

Finally, we grouped participants into broader education groups to balance the sample size within each category. For example, someone with a Doctoral degree and someone with a Master’s degree are both categorized as having a “Higher than Bachelor” education rather than being categorized into two distinct groups.

Examination of Relevant Variables

The cleaned data contains responses from 408 participants aged between 18 and 67, with a mean age of 32.

Demographic Variables

Diagnosis Method

Table 1: Percentage of Self-Diagnosed vs. Formally Diagnosed Participants

Diagnosis	N	Percentage (%)	Seeking Formal Diagnosis (N)	Seeking Formal Diagnosis (%) Out of Group Size	Seeking Formal Diagnosis (%) Out of Total
Formal diagnosis	143	35	0	0.0	0.0
Self-diagnosed	266	65	146	54.9	35.7

Table 1 summarizes the number and percentage of self-diagnosed and formally diagnosed participants. The majority (65.2%) of the participants are self-diagnosed. 146 self-diagnosed participants (54.9% of self-diagnosed participants, 35.8% of all participants) are seeking a formal diagnosis. While being self-diagnosed with autism is common among the autistic community according to our collaborator, the relatively imbalanced sample sizes across the diagnosis method might influence the generalizability of our results to the entire autistic community.

Stigmatized or Minority Identities (Apart from Autistic Identity) As having multiple stigmatized or minority identities might have unique consequences to mental health, we examined participant’s gender, sexual orientation and visible minority identities.

Table 2: Gender Distribution

Gender	N	Percentage (%)
Man	175	42.8
Woman	193	47.2
Other	41	10.0

Table 3: Sexual Orientation Distribution

Sexual Orientation	N	Percentage (%)
Bisexual and/or Pansexual	94	23.0
Gay or Lesbian	36	8.8
Heterosexual	231	56.5
Other	48	11.7

Table 4: Visible Minority Status

Visible Minority	N	Percentage (%)
Yes	51	12.5
No	354	86.6
Unknown	4	1.0

Tables 2-4 summarize participants’ gender, sexual orientation and visible minority identity, respectively. 40 out of 408 participants (9.8%) identify as having a non-binary gender, 177 participants (43.4%) identify as non-heterosexual orientation, and 51 participants (12.5%) are members of a visible minority group.

Table 5: Gender, Sexual Orientation and Visible Minority Group Membership

Number of stigmatized or minority identity	N	Percentage (%)
At least one	206	50.4
At least two	58	14.2
Three	6	1.5

Table 5 outlines the number and percentage (out of the total) of participants who identify with at least one, at least two, and three stigmatized or minority groups other than the autistic community. Notably, about half of the participants have at least one minority identity that might interact with their autistic identity in influencing mental health.

Geographical Region

Table 6: Distribution of Sample by Geographical Region

Continent or Region	N	Percentage (%)
Asia	2	0.5
Europe	356	87.0
Latin America and the Caribbean	3	0.7
North America	16	3.9
Oceania	4	1.0
Sub-Saharan Africa	28	6.8

Table 6 summarizes the participants’ geographical regions. 355 out of 408 participants (87.0%) come from Europe, but only 33 are from non-Western regions (i.e. Asia, Latin America and the Caribbean, and Sub-Saharan Africa). Since the majority (86.5%) of the participants do not identify with a visible minority group (as shown in **Table 4**), the sample is relatively homogeneous in terms of region, race and ethnicity. Thus, the results might not be representative of the autistic community worldwide, especially in non-Western cultures.

Measure (Total Score) Distributions

Next, we examined the distributions of total scores on the four variables of interest (autistic traits concealment, depression, autism community connectedness, and stigma experience) and the pairwise correlations among these variables.

Table 7: Summary of Variables of Interest and Score Distributions

Variable	Scale (possible range of scores)	Sample Max	Sample Min	Sample Mean	Standard Deviation
Autistic community connectedness	Autistic Community Connectedness Measure (0-40)	40	0	14.08	8.84
Autistic traits concealment	Camouflaging Autistic Traits Questionnaire (0-92)	88	11	58.28	14.99
Depressive symptoms	Center for Epidemiological Studies Depression Scale (0-60)	56	1	28.91	12.14
Stigma experience	Stigma Experience (0-12)	12	0	5.21	3.69

Table 7 outlines the four variables of primary interest and the questionnaires (scale) by which they are measured. A summary of total score distributions is also provided.

CESD (Response)

Figure 1 shows the overall distribution of depressive symptoms, measured by total CES-D scores, and the distribution for both diagnosis groups (formal and self). Overall, most participants scored higher on the depression scale than the clinical cutoff (a total score of 16, as indicated by the red vertical line).

The depression score distribution among self-diagnosed participants is similar to the overall distribution, and the majority of self-diagnosed participants scored higher than 16.

The depression score distribution among formally diagnosed participants appears to be bimodal. One mode is at around the clinical cutoff, and the other occurs at a score of around 35 which is similar to the mode of the overall distribution. While the bimodal distribution may indicate idiosyncrasies of formally diagnosed participants, it might as well arise due to a relatively small sample.

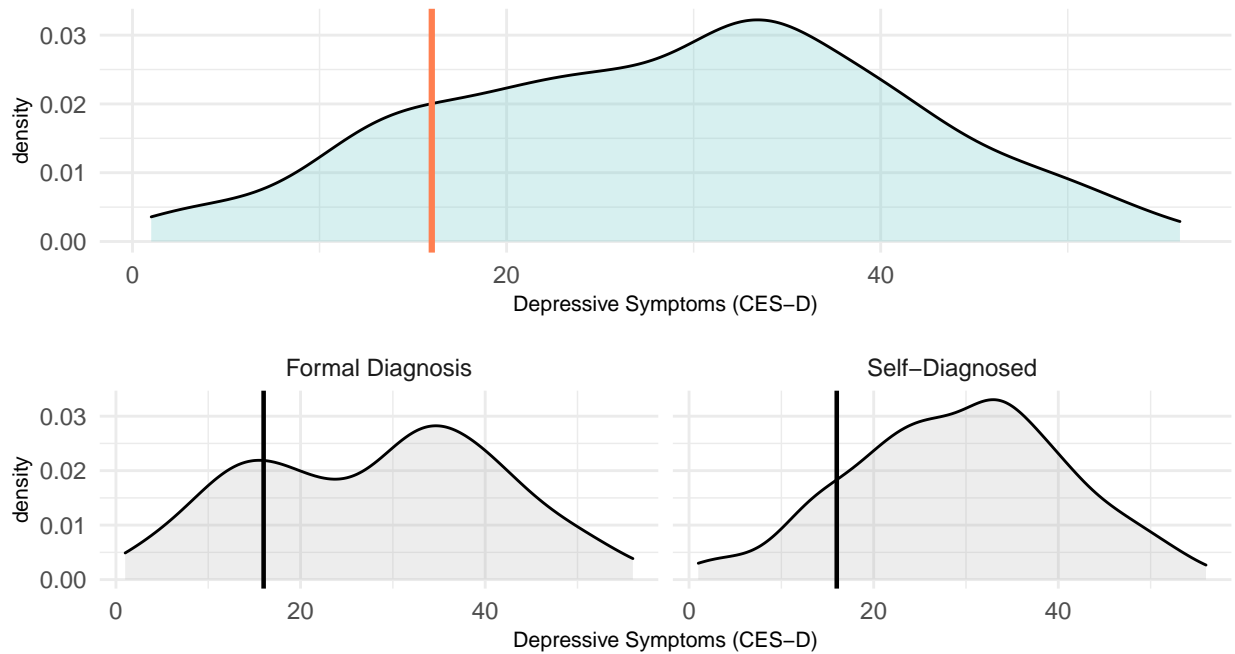


Figure 1: Overall depressive symptoms distribution, and distributions by diagnosis methods. CES-D scores > 16 indicates clinically significant depressive symptoms

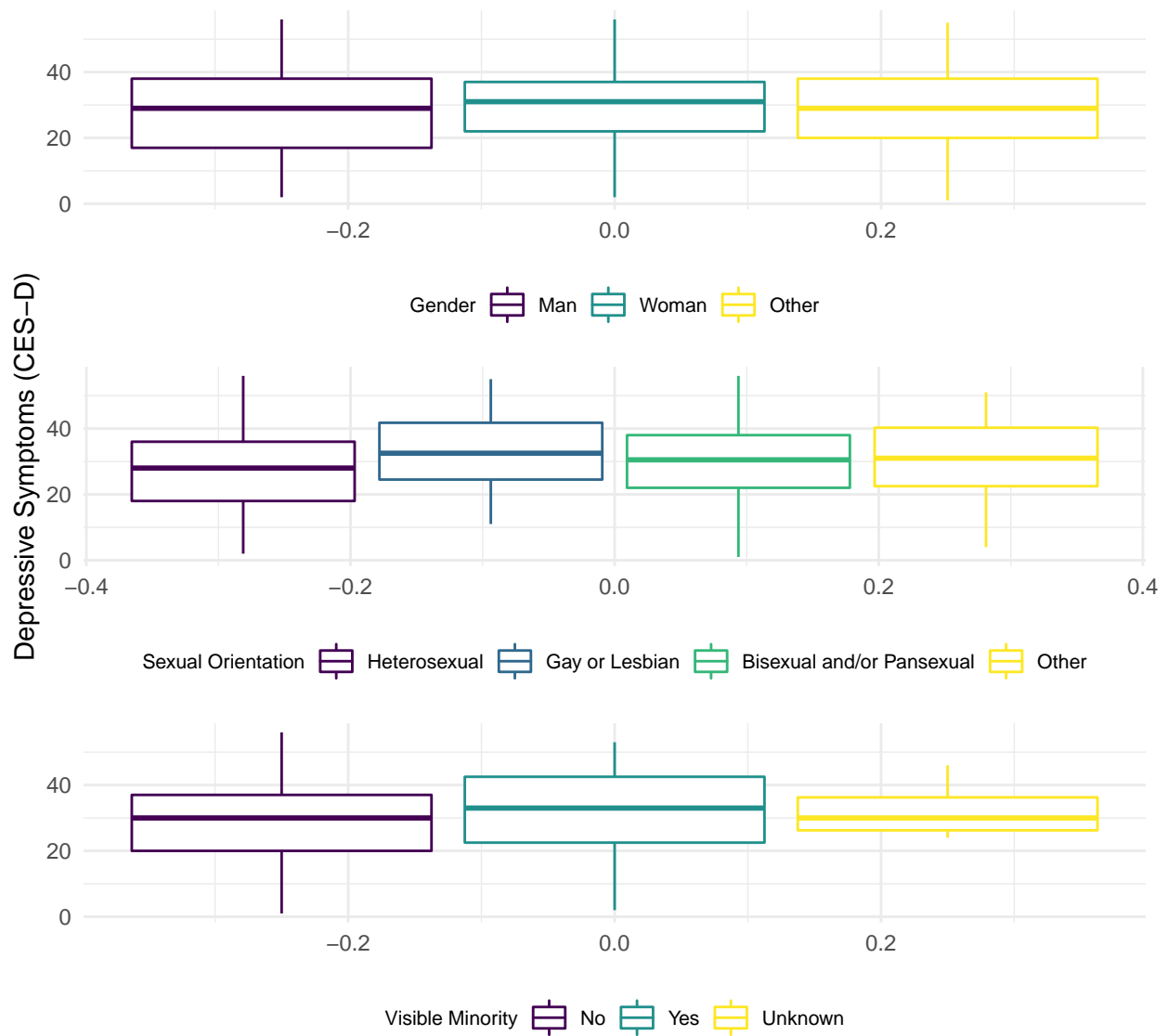


Figure 2: Depressive symptoms distribution by gender, sexual orientation, and visible minority groups

Figure 2 shows the distribution of depression scores over gender, sexual orientation, and visible minority groups. On average, participants who are male, heterosexual and/or not of visible minority groups have slightly fewer depressive symptoms than the other groups. But overall, there does not appear to be large differences in median depression scores across the sub-groups.

CATQ (Predictor)

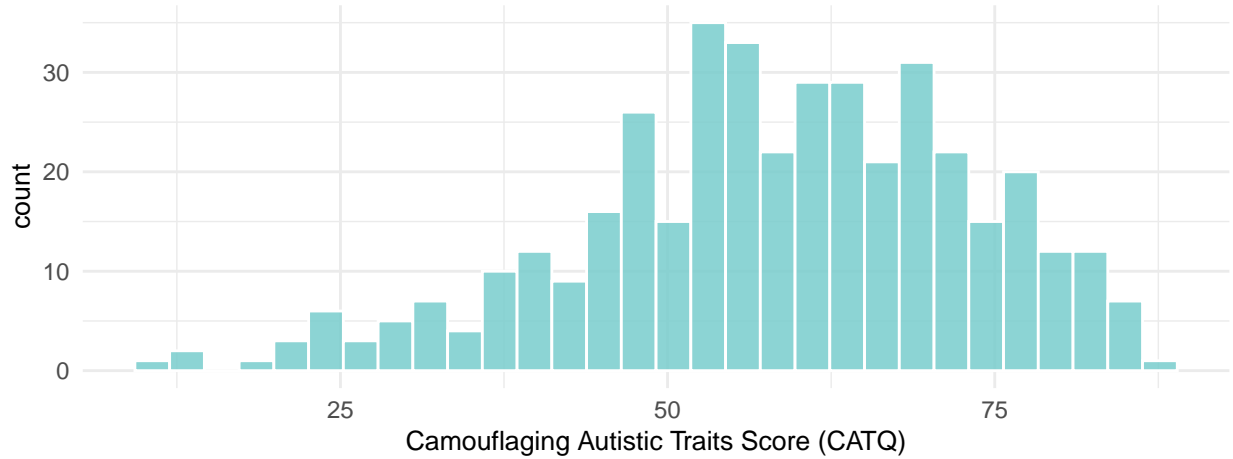


Figure 3: Overall autistic traits concealment score distribution. Higher scores indicate more autistic concealment behaviours.

Figure 3 shows the overall distribution of autistic traits concealment, measured by the total CATQ scores. Most participants scored higher than 50 on the CATQ questionnaire, indicating they display a relatively high level of autistic traits concealment behaviours.

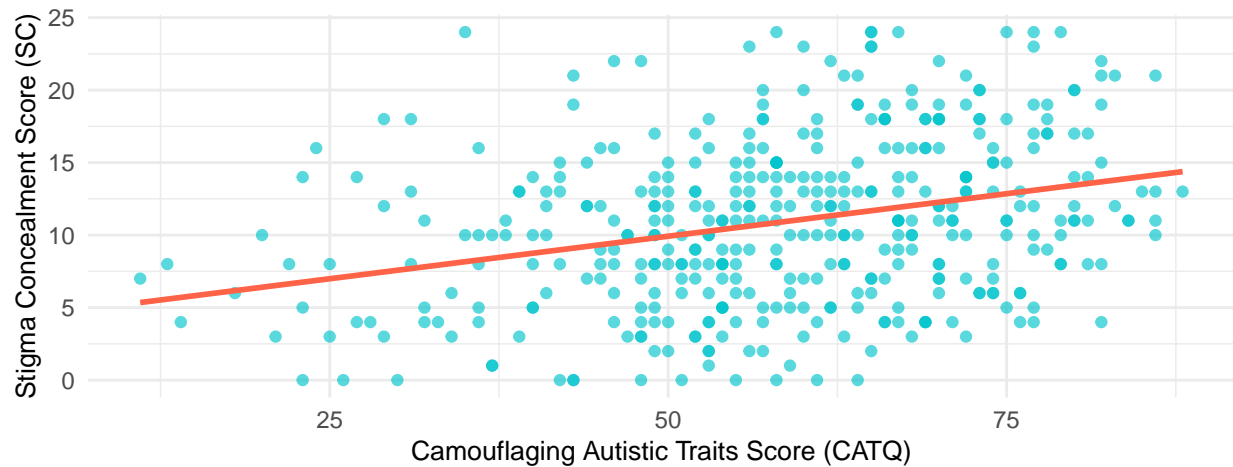


Figure 4: Correlation between overall stigma concealment and autistic traits concealment scores

Figure 4 shows the correlation between total scores on the Stigma Concealment scale (SC) and Camouflaging Autistic Traits Questionnaire (CATQ). SC is a more general measure of stigma concealment behaviours and beliefs, but CATQ is specific at measuring autistic trait concealment. SC and CATQ scores are moderately positively correlated, indicating that CATQ scores validly measure autistic trait concealment.

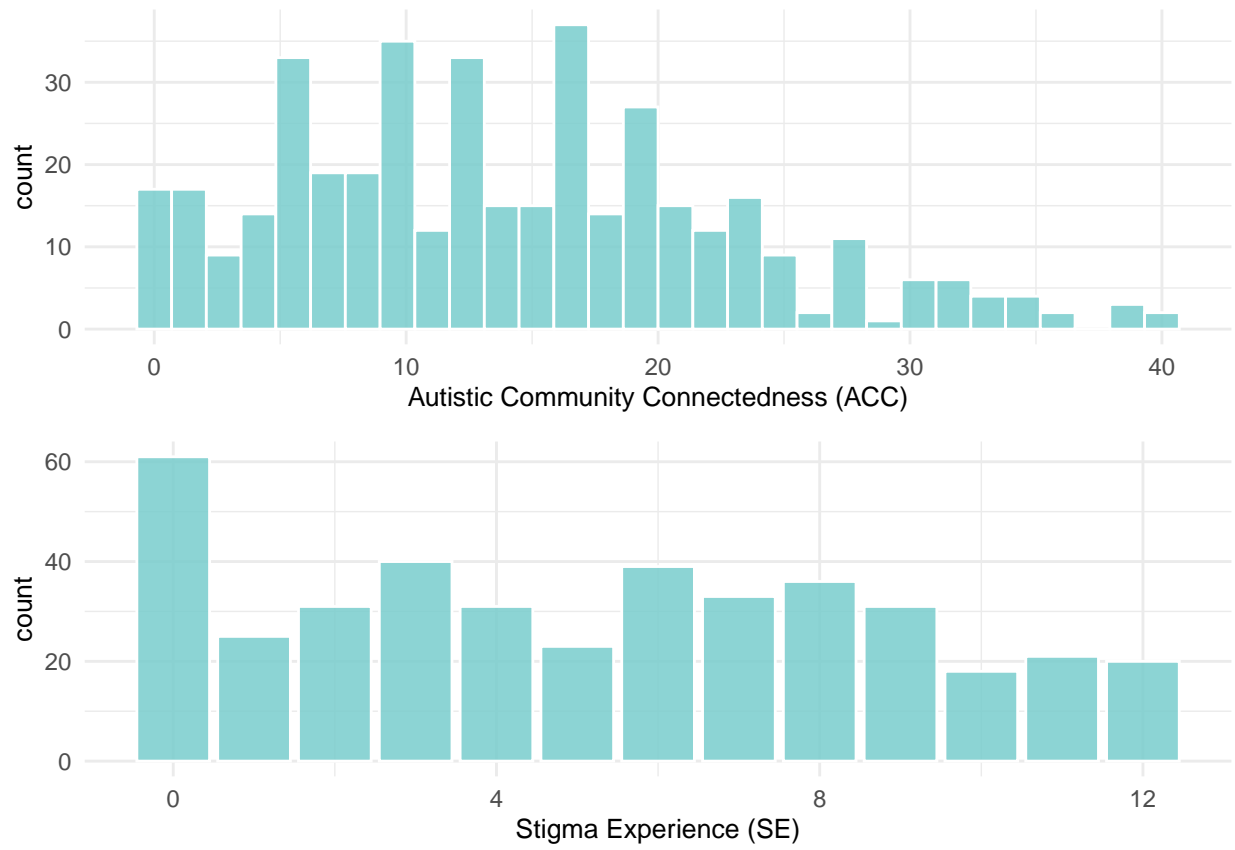


Figure 5: Overall distribution of autistic community connectedness scores and stigma experience scores. Higher scores indicate more connectedness or more experience with autism-related stigma

ACC and SE (Mediators)

Figure 5 shows the overall distributions of autistic community connectedness (measured by ACC total scores) and stigma experience (measured by SE total scores). Most participants are not very strongly connected to the autistic community and have experienced at least some stigma because of their autistic traits. However, it is surprising that about 60 participants scored 0 on the SE scale, which can have two explanations: (i) they truly have not had any experience with autism-related stigma and discrimination, or (ii) they have experienced other forms of autism-related stigma that are not listed in the SE items.

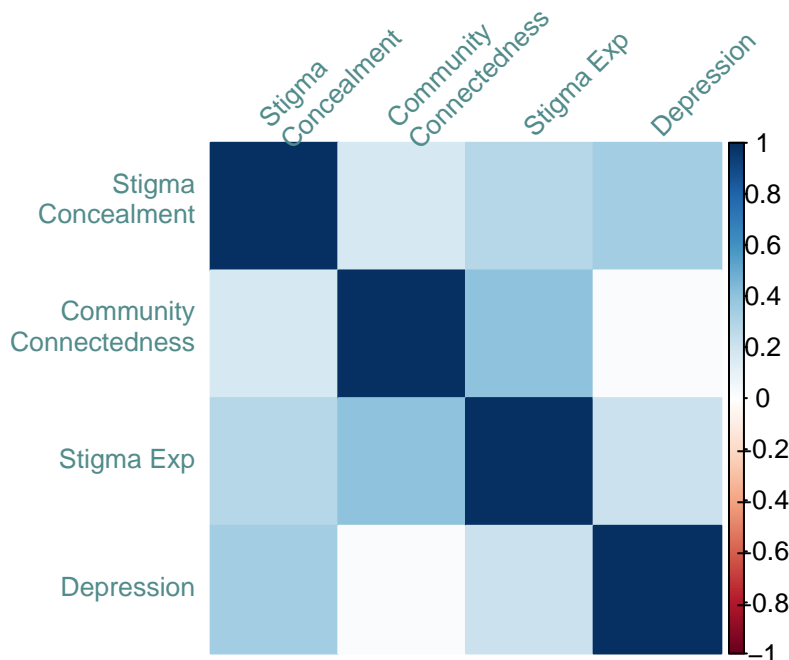


Figure 6: correlation among predictor, mediators and response

Correlation Among CATQ, ACC, SE and CESD Finally, we examined the pairwise correlations among depressive symptoms (CESD), autistic traits concealment (CATQ), autistic community connectedness (ACC) and stigma experience (SE). **Figure 6** show a weak to moderate positive correlation between each pair of variables. ACC and CESD are very weakly correlated, thus ACC may or may not have a mediating effect as our collaborator hypothesized.

References

Botha, M., Dibb, B., & Frost, D. (2020). “Autism is me”: An investigation of how autistic individuals make sense of autism and stigma. *Disability & Society*, 37, 427–453. <https://doi.org/10.31219/osf.io/gv2mw>

Jackson, S. D., & Mohr, J. J. (2016). Conceptualizing the closet: Differentiating stigma concealment and nondisclosure processes. *Psychology of Sexual Orientation and Gender Diversity*, 3(1), 80–92 <https://doi.org/10.1037/sgd0000147>