Logistic regression with machine learning sheds light on the problematic sexual behavior phenotype

Shui Jiang[1,2], Keanna Wallace[1,2], Esther Yang[1,2], Leslie Roper[1,2], Garima Aryal[1,2], Dawon Lee[3], Rohit J Lodhi[4], Rick Isenberg[5], Bradley Green[6], David Wishart[7], Katherine J Aitchison[1,2,8*]

[1]Department of Medical Genetics, University of Alberta, Canada,

[2]Department of Psychiatry, University of Alberta, Canada,

[3]Department of Medicine, University of Alberta, Canada

[4]Department of Psychiatry, University of Western Ontario, Canada,

[5]Psychological Counselling Service, Ltd., Scottsdale, USA,

[6]Department of Psychology and Counseling, University of Texas at Tyler, USA,

[7]Department of Biological Sciences, University of Alberta, Canada,

[8]Neuroscience and Mental Health Institute, University of Alberta, Canada

*Corresponding author:

Email: kaitchis@ualberta.ca (KJA)

Abstract

There has been a longstanding debate about whether the mechanisms involved in problematic sexual behavior (PSB) are similar to those observed in addictive disorders, or related to impulse control or compulsivity. The aim of this report was therefore to contribute to this debate by investigating the association between PSB, addictive disorders (internet addiction, compulsive buying), other measures associated with the construct known as reward deficiency (attention deficit disorder and personality disorder), and obsessive-compulsive disorder (OCD) in young adults in post-secondary education.

A Canadian University Office of the Registrar invited 68846 eligible students and postdoctoral fellows. Out of 4710 expressing interest in participating, 3359 completed on-line questionnaires and 1801 completed the Mini-International Neuropsychiatric Interview (MINI). PSB was measured by combining those screening positive (≥6) on the Sexual Addiction Screening Test-Revised (SAST-R) Core with those self-reporting monthly PSB. Current mental health condition(s) and childhood trauma were measured by self-report. OCD was assessed by a combination of self-report and MINI data.

12.18% (407/3341) of participants screened positive on the SAST-R Core. On logistic regression analysis (conventional and by machine learning), OCD, attention deficit, internet addiction, a family history of PSB, childhood trauma, compulsive buying and male gender were associated with PSB. On multiple correspondence analysis (MCA), OCD appeared to cluster separately from the other measures, and the pattern of data differed by gender.

In summary, factors that have previously been associated with the construct known as RDS and OCD are both associated with increased odds of PSB. The factors associated with RDS appear to contribute to a separate data cluster from OCD, and to lie closer to PSB.

1. Introduction

A theory of hypersexuality with dependence was first proposed by Orford in 1978.[1] The term sex addiction was used by Carnes in 1983, 1990, and 1991[2] and by Goodman in 1998.[3] Continuation of a sexual behavior despite adverse consequences and/or distress caused or worsened by the sexual behavior has been a consistent theme in the literature,[1,3,4] despite differences of opinion as to whether the condition should be classified as an addiction or as an impulse control disorder. Diagnostic criteria proposed by Carnes in 2005[4] led to sex addiction being included in the Diagnostic and Statistical Manual of Mental Disorders (DSM)-III-R (302.87). However, despite draft proposed criteria for the inclusion of sexual addiction and internet addiction in DSM-5,[5] the only behavioral addiction currently included in DSM-5 (American Psychiatric Association 2013) is gambling disorder (312.31), which was first included in DSM as an impulse control disorder.[6] In the International Statistical Classification of Diseases (ICD), compulsive sexual behavior disorder was proposed for inclusion as an impulse control disorder in 2014,[7] which occurred in 2018.[8] However, the scientific discussion about whether or not compulsive sexual behavior disorder could constitute a manifestation of a behavioral addiction was acknowledged,[9] and it was predicted (as in the case of problematic gambling),[6] that understanding would evolve as research elucidated the phenomenology and neurobiological underpinnings of the condition.

In this article, the term problematic sexual behavior (PSB) is used because it focuses on the behavior rather than on the potentially heterogeneous etiology and pathology.[10] PSB is characterized by repetitive sexual behaviors and it is associated with uncontrolled sexual urges/impulses and distress,[11] and social and functional impairments.[8,11] It has been estimated that the prevalence of PSB is 3-6% in adults,[4,12-14] with higher frequencies in specific populations, including those in post-secondary education. Reid (2011) reported that 19% of college men met

criteria for hypersexuality,[15] while Giordano and Cecil (2014) found 11.1% of men and women in an undergraduate sample met these criteria.[16] Cashwell et al. (2015) reported that out of 379 undergraduates, 21.2% of men and 6.7% of women screened positive, i.e., scored in a range indicating that they should be offered further assessment for sexual addiction.[17]

Various vulnerability/risk factors have been associated with PSB, including those previously associated with the construct known as reward deficiency syndrome (RDS), such as substance[18] and non-substance use disorders,[19] ADHD,[20] personality disorders,[21] as well as other psychiatric disorders,[22] and childhood trauma.[23] In terms of psychiatric disorders, owing to the classification of PSB within ICD-10 as an impulse control disorder, we specifically included obsessive-compulsive disorder. Mick and Hollander (2006) hypothesized that impulsivity initiated the early stage of the PSB, with compulsivity being involved in repetitive behaviors and hence in the maintenance of PSB.[24] The magnitude of the contribution of the various different postulated mechanisms, and whether PSB encompasses various syndromes with different etiologies is not at present known. Moreover, systematic data on the prevalence of PSB and associated sociodemographic factors across diverse populations, including non-treatment seekers remain to be provided.[8]

In light of the above, this paper aims to further elucidate the prevalence, sociodemographic features, phenomenology, and neurobiological underpinnings of the construct known as PSB in a diverse sample of adults in post-secondary education. We hypothesized that in this sample, firstly screening positive for factors previously associated with RDS (such as internet addiction, compulsive buying, nicotine dependence, and pathological gambling) and secondly OCD and childhood trauma would be associated with increased odds of PSB.

2. Methods

The study inclusion criteria were as follows: at least 18 years of age, undergraduates, graduates, postdoctoral fellows, and recently convocated students registered in at least one course in the preceding year, except for those who had completely withdrawn after registration, without any reasons to not return in the next academic term or not be on campus (such as suspension), and being able to answer in English. A Canadian University Office of the Registrar invited 68,846 eligible students and postdoctoral fellows by email. Students interested in participating then emailed the study team. These students were then sent an email invitation to review the participant information and complete consent on-line, followed by the screening measures (hosted by the Qualtrics platform), with a subsequent email inviting them to complete the Mini-International Neuropsychiatric Interview (MINI).

2.1 Measures

The Sexual Addiction Screening Test-Revised (SAST-R)[25] is a 45-item screening tool comprising several subscales designed to detect potentially problematic sexual behavior. Each question is answered in a binary manner (yes/no=0/1). Sexual activities deemed to be problematic by participants in the past 30 days were also separately assessed by asking the following question: "Within the last 30 days, how often have you participated in activities of a sexual nature that you would regard as problematic?"

Measures used to screen for internet addiction, compulsive buying, personality disorders, ADHD, nicotine dependence, and pathological gambling were the Internet Addiction Test (IAT),[26] the Richmond Compulsive Buying Scale (RCBS),[27] the 8-item SA-SAPAS,[28] the full Adult ADHD Self-Report Scale (ASRS-v1.1),[29] the Fagerström Test for Nicotine Dependence (FTND)[30] and the DSM-V Pathological Gambling Diagnostic Form (PGDF),[31] respectively. The self-report online administered Mini-International Neuropsychiatric Interview (MINI) version

7.0.2 was used to output psychiatric disorders by DSM-5 criteria. Having received a diagnosis of OCD, any other mental health condition(s), childhood physical, emotional/verbal, or sexual trauma, and a family history of domestic violence or of sexual addiction, gender and sexual orientation were additionally collected by self-report.

2.2 Data analysis

2.2.1 Demographic data and logistic regression

Data were analyzed by STATA (Stata/SE 16) and R (Version 3.6.3) after dropping all of the missing data.

A "total PSB" variable was created by combining those screening positive on the SAST-R Core with those endorsing monthly self-reported PSB. Previous or current physical/emotional/sexual childhood trauma and a family history of domestic violence (11.77% 387/3288) were regrouped as total childhood trauma due to the relatively low rates of endorsement for the former group (physical trauma: 2.98%, 99/3322; emotional trauma: 6.87%, 232/3322; sexual trauma: 3.31%, 110/3212). SAST Core question 1 (which enquires about a history of sexual abuse in childhood or adolescence, SASTC1) was combined with total childhood trauma. The self-reported family history of sexual addiction ("yes"/ "no") data was combined with SAST Core question 2 (which enquires about parental trouble with sexual behaviour, SASTC2) to create a new family history of PSB variable. IAT was regrouped as "yes" (N=1174, combining mild, moderate, and severe groups) and "no" (N=2160, the "normal" group). Owing to much lower endorsement rates (or lower than in another university student sample),[32] the FTND (0.45%, 15/3332) and PGDF (2.28%, 76/3332) scores were dropped. In addition, due to the small numbers in the gender and sexual orientation minority subgroups, only men and women were included in the logistic regression analysis (N=3091).

Two-sample Wilcoxon rank-sum tests were used to determine the age difference between the PSB- positive and -negative groups. Data distributions were assessed by Pearson χ^2 test. Tetrachoric (for binary variables) and Kendall's τ (for categorical variables) correlations were utilized. SA-SAPAS, OCD, ASRS, IAT, childhood trauma, a family history of PSB and gender were used as covariates in regression analyses (Figure 2, Table 2-4). All reported *P* values are nominal. Multiple correspondence analysis (MCA) was used for data clustering.

2.2.2 Machine learning approaches

The "Haven" and "dplyr" packages were used for data input. The "forcats" and "creditmodel" packages were used for recoding and splitting (80% training 20% testing), respectively. The Synthetic Minority Oversampling Technique (SMOTE) from the "Caret" package was used to address data imbalance in the training set.[33] The leave-one-out cross-validation (LOOCV) in the "Caret" package was used for resampling.[34] Logistic regression (two-class summary) in the "caret" package was used for regression. Accuracy, area under the receiver operating characteristic (ROC) curve, sensitivity, specificity, and F score were used to evaluate the models. The "pROC" package was used for the area under the ROC curve.

2.2.3 Multiple correspondence analysis (MCA)

Multiple correspondence analysis was conducted on OCD, RCBS, SA-SAPAS, ASRS, IAT, and childhood trauma stratifying by gender (women N=2084 and men N=1012). As a family history of PSB was correlated with the combined childhood trauma variable (tetrachoric ϱ=0.69, P<0.001, N=3019, Figure 1, Supplementary Table 3), these variables were combined as a quality supplementary childhood trauma-related variable. The "FactoMineR," "factoextra," and "GDAtools" packages in R were used for analysis and data visualization.

3. Results

Out of 4710 interested in participating, 237 withdrew from the study, 3359 completed online questionnaires using the Qualtrics platform, and 1801 completed the MINI (Supplementary Figure 1).

3.1 Demographic data

The mean score on the SAST-R Core Scale among all participants was 2.23 (SD=2.73), with 12.18% (407/3341) of participants scoring at least six (the threshold for screening positive in the general population.[25] Of the 20 items composing the SAST-R Core questions, item 11 (Do you hide some of your sexual behavior from others?) was the most endorsed (1234/3341, endorsement rate: 36.78%), while item 9 ("Are any of your sexual activities against law?") was the least endorsed (35/3342; endorsement rate: 1.05%) (Supplementary Table 1). The frequency of participants included in the total PSB variable was 16.53% (532/3219). There was a significant difference in the distribution of the following variables by those with and without PSB: gender, sexual orientation, OCD, RCBS, a family history of PSB, SA-SAPAS, current mental health condition(s), childhood trauma, ASRS, and IAT (Table 1 and Supplementary Table 2).

Data correlation showed that PSB was correlated with non-heterosexual sexual orientation (baseline: heterosexual), current mental health condition(s), childhood trauma, RCBS, OCD, SA-SAPAS, a family history of PSB, ASRS, IAT and gender (baseline: women) (Figure 1, Supplementary Table 3). Multiple variables were correlated with one another. For example, sexual orientation was correlated with current mental health condition(s). Current mental health condition(s) were correlated with childhood trauma, OCD, SA-SAPAS, ASRS; and were negatively correlated with gender (baseline: women). RCBS was negatively correlated with gender (baseline: women). In addition, our results indicated that that after subtracting

SASTC1 and SASTC2, the PSB variable still had a strong correlation with SAST-R Core; and SAST-R Core was correlated with self-reported PSB (Supplementary Table 4). Similarly, self-reported OCD was correlated with MINI-identified OCD (Supplementary Table 4).

3.2 Regression models

3.2.1 Logistic regression

Outliers were removed from logistic regression (N=47, with 8 responding "yes" to total PSB, Supplementary Figure 2) if their Pearson standardized residual had an absolute value of more than 2,[35] or Hosmer-Lemeshow $\Delta\hat{\beta}$ more than four.[36] After outlier removal as above, no further outliers needed to be removed on the basis of leverage values (a measure of the distance between independent variable observations). The following variables were associated with PSB (pseudo R^2=0.16): OCD (P=0.015), ASRS (P<0.001), IAT (P<0.001), a family history of PSB (P<0.001), SA-SAPAS (P<0.001), childhood trauma (P<0.001), RCBS (P<0.001) and gender (men) (P<0.001) (Figure 2, Table 2).

3.2.2 Machine Learning (ML) logistic regression

Our ML logistic regression showed that RCBS (P=0.0030), SA-SAPAS (P<0.001), OCD (P<0.001), a family history of PSB (P<0.001), ASRS (P<0.001), childhood trauma (P<0.001), IAT (P<0.001), and gender (men) (P<0.001) were all associated with increased odds of PSB (Figure 2B, Table 4). A validation of the existing model was also undertaken (Figure 2C, Table 3). The validation model was trained with SAST-R Core (subtracting SASTC1 and SASTC2, N=3141) and tested with self-reported PSB (N=3145).

3.3 Multiple correspondence analysis (MCA)

The MCA results showed that in both women and men, the first two dimensions accounted for the most explained variance (Supplementary Figure 5). PSB had a high degree of representation in all dimensions (Supplementary Table 7). PSB was strongly represented in both men and women in dimensions 1, 2, 4 and 5, although showing slightly different data clustering patterns by gender in these dimensions (Figure 3-4). In dimension 1 (44.00% of the explained variance in women and 39.10% of the explained variance in men, Figure 3 and Supplementary Figure 5), PSB was correlated with all variables in women and men. In dimension 2 (15.00% of the explained variance in women and 17.30% of the explained variance in men, Figure 3 and Supplementary Figure 5), PSB clustered with IAT and ASRS in men, and with RCBS, IAT and ASRS in women. In women, in dimension 4 (10.50% of the explained variance) and dimension 5 (10.10% of the explained variance), PSB was correlated with childhood trauma-related variables. In men, by contrast, PSB was correlated with childhood trauma-related variables just in dimension 5 (10.50% of the explained variance) but not in dimension 4 (11.60% of the explained variance) (Figure 4, Supplementary Figure 5). The contributions of variables to each dimension by gender is shown in Supplementary Figures 6-7 and Supplementary Table 8.

4. Discussion

The frequency of problematic sexual behavior, as defined by screening positive on the SAST-R Core, was 8.49% in the women and 19.70% in the men in this post-secondary education sample, consistent with prior reports in this population.[15-17]

Factors consistently associated with increased odds of PSB were gender, mental health conditions (ADHD, personality disorder, and OCD), other problematic behaviors such as internet addiction (IAT) and compulsive buying (RCBS), childhood trauma (including a family history of domestic violence), and a family history of PSB. Our results also showed that self-reported data

183 were correlated with the screening tool or MINI data. For example, the binary SAST-R core was
184 significantly correlated with self-reported "problematic sexual activities in the past 30 days"
185 (τ_b=0.44, *P*<0.001, N=2049) and "maximum length of time on any single occasion spent in
186 problematic sexual activities" (τ_b=0.32, *P*<0.001, N=2049). IAT is significantly associated with
187 self-reported "problematic internet use in the past 30 days" (four categories IAT: τ_b=0.42, *P*<
188 0.001; binary IAT: τ_b=0.40, *P*<0.001, N=1441). SA-SAPAS was significantly correlated with
189 antisocial personality order diagnosed by MINI (tetrachoric ϱ=0.24, *P*=0.026, N=1788) and with
190 self-reported borderline personality disorder (tetrachoric ϱ=0.44, *P*<0.001, N=1429).
191 Interestingly, a family history of PSB was shown to be strongly correlated with childhood trauma
192 in our dataset (tetrachoric ϱ=0.69, *P*<0.001, N=3019, Figure 1, Supplementary Table 3), and
193 both were associated with PSB. Of those responding "yes" to a family history of sexual
194 addiction, 32.65% (16/49) screened positive on the SAST-R Core. The combined variable,
195 including a family history of sexual addiction and SASTC2 (with 6.55% endorsement: 214/3268)
196 was significant in all models.
197 Our model validation did show a lower F score than the final model training and testing
198 with total PSB, probably due to the difference between screening and self-report data (training
199 with SAST-R Core and testing with self-reported PSB, Figure 2C, Table 4). A sensitivity
200 analysis (Supplementary Figure 3-4, Supplementary Table 5-6) was undertaken using the
201 original variables, namely SAST-R Core (without subtracting SASTC1 and SASTC2), childhood
202 trauma (without adding SASTC1 or a family history of domestic violence) and a family history
203 of sexual addiction (without incorporating SASTC2). The logistic regression model using these
204 variables (pseudo R^2=0.13) revealed that RCBS (*P*=0.033), ASRS (*P*<0.001), IAT (*P*<0.001),
205 OCD (*P*<0.001), childhood trauma (*P*<0.001), and gender (baseline: women, *P*<0.001) remained

206 significant, but not a family history of sexual addiction (*P*=0.17), probably due to the lower
207 frequency of endorsement of this variable (N=49) compared to SASTC2. Using the new total
208 PSB variable, childhood trauma, a family history of PSB and excluding sexual orientation as
209 well as current mental health condition(s) improved the sensitivity and F-score in logistic
210 regression. This strategy also improved the specificity and F score in ML logistic regression. ML
211 logistic regression in general showed a higher sensitivity and F score compared to logistic
212 regression.

213 The variables previously associated with the RDS construct, such as SA-SAPAS, ASRS,
214 and IAT, were associated with increased odds of PSB. In the final logistic (Figure 2, Table 2)
215 and ML logistic regression (Figure 2, Table 3), RCBS was associated with increased odds of
216 PSB. Of note, there were gender differences regarding RCBS: RCBS was negatively correlated
217 with men (Figure 1, baseline: women), and PSB displayed a weaker association with RCBS in
218 men on MCA (Figures 3 and 4).

219 In addition, interestingly, the distribution of PSB differed significantly between OCD-
220 positive and -negative groups (*P*<0.001, Table 2), and PSB was correlated with OCD (*r*=0.27,
221 *P*<0.001, Figure 1, Supplementary Table 3). In our final regression models and sensitivity
222 analyses (by both logistic and ML logistic regression), OCD was associated with increased odds
223 of PSB (OR 1.79, *P*=0.015; and OR 1.85, *P*=0.013, Sensitivity model: OR 2.28, *P*<0.001; and
224 OR 1.73, *P*=0.033, respectively). The MCA results also indicated that PSB had a less strong
225 association with OCD than with SA-SAPAS, IAT and ASRS in women and men, as it clustered
226 further away from PSB (Figure 3).

227 Limitations of this study include the following. Firstly, as respondents to the email from
228 the registrar knew the subject area of the study, there is a potential selection bias. However, the

229 frequency of PSB in our dataset was comparable to that seen in other studies of adults in post-
230 secondary education as mentioned above.[12,17] Subsample sizes were a limiting factor in sexual
231 orientation and in gender subgroups other than cis-men and cis-women, and therefore the smaller
232 groups were combined or dropped. Another potential limitation is that we dropped missing
233 values (rather than using imputation). Various methods may be used to deal with missing data.
234 Some such methods (e.g., multiple by monotone imputation, or imputation by chained equations)
235 may introduce bias, and perfect (inaccurate) predictions, as the methods themselves involve the
236 fitting of a regression model for a categorical outcome.[37] However, if the proportion of missing
237 data is relatively small (5% or less), this should not be problematic.[38] In our sample, the feature
238 with the most missing data in the SAST-R core scale was: "SastC14 -- Feel depressed after
239 having sex, " with only 16 out of 3341 (0.48%) missingness (and out of those 16 participants,
240 only 1 scored as SAST-R core positive). Another limitation is that logistic regression is sensitive
241 to high correlations among predictor variables. Although several variables were correlated with
242 each other in our dataset, correlation coefficients among these variables were less than 0.80,[39]
243 and the mean variance inflation factor (VIF) and the condition number were at acceptable levels
244 (VIF<10, condition number<10) for our final models,[40] indicating multicollinearity was not a
245 significant concern. A power calculation is provided in Supplementary Table 9. Using our model
246 parameters, a sample size of 361 would have a power of 0.90.
247 Genetic and other biological variables may also contribute to a greater likelihood of PSB
248 that is as yet unexplained by our models (pseudo R^2=0.16, Table 2). Future directions should
249 include incorporating such factors, and optimizing the model for enhanced predictive power
250 through the comparison and use of alternative machine learning models, feature selection

techniques, and methods for computing missing values, as well as expanding the validation by the use of additional independent datasets.

5. Conclusions

In summary, we have identified associations between problematic sexual behavior in adults in post-secondary education participating in a study on this topic and the following: compulsive buying, personality disorder, OCD, ADHD, childhood trauma, a family history of PSB, internet addiction, and gender (men). Our results not only suggest that measures associated with reward deficiency syndrome are associated with PSB, but also that OCD is a vulnerability factor with a separate contribution. Childhood trauma and a family history of domestic violence also showed associations with PSB, implying that emotional regulation and attachment style may be involved in the development of PSB. Our data are consistent with different subgroups within PSB with contrasting etiological mechanisms that differ by gender.

6. Acknowledgments

We would like to acknowledge the contribution of Dr. Patrick Carnes to the study design, and volunteer input of Penny Carnes, Grace Li, Hana Graham, and Chanelle Martens to data collection. SJ would like to thank Afia Anjum from DW's lab for her assistance with R.

Table 1. Distribution of obsessive-compulsive disorder (OCD), RCBS, a family history of problematic sexual behavior (PSB), SA-SAPAS, current mental health condition(s), childhood trauma, ASRS, and IAT data by total PSB group (subtracting SASTC1 and SASTC2)

Categories	Total PSB responding "yes" 532/3219 (16.53%)	Total PSB responding "no" 2687/3219 (83.47%)	Total number
SAST scores	6.58±3.22 (N=532)	1.38±1.56 (N=2687)	3219
Two-sample Wilcoxon rank-sum (Mann-Whitney) test, *P*<0.001			
Age	22.32±5.13 (N=532)	22.34±5.05 (N=2686)	3218
Two-sample Wilcoxon rank-sum (Mann-Whitney) test, *P*=0.81			
Gender			
Women (N=2145)	232 (10.82%)	1913 (89.18%)	2145
Men (N=1026)	290 (28.27%)	736 (71.73%)	1026
Other (N=47)	10 (21.28%)	37 (78.72%)	47
Pearson χ^2 (2) =153.92, *P*<0.001			
Sexual Orientation			
Heterosexual (N=2650)	420 (15.85%)	2230 (84.15%)	2650

Non-heterosexual (N=564)	110 (19.50%)	454 (80.50%)	564
Pearson χ^2 (1) =4.51, P=0.034			
OCD (self-reported current diagnoses and MINI)			
No	490 (15.88%)	2596 (84.12%)	3086
Yes	40 (33.33%)	80 (66.67%)	120
Total	530 (16.53%)	2676 (83.47%)	3206
Pearson χ^2 (1) =25.51, P<0.001			
RCBS			
No	471 (15.54%)	2559 (84.46%)	3030
Yes	61 (32.28%)	128 (67.72%)	189
Total	532 (16.53%)	2687 (83.47%)	3219
Pearson χ^2 (1) =36.10, P<0.001			
A family history of PSB (a family history of sexual addiction with SASTC2)			
No	453 (15.38%)	2492 (84.62%)	2945
Yes	71 (34.63%)	134 (65.37%)	205
Total	524 (16.63%)	2626 (83.37%)	3150
Pearson χ^2 (1) =51.22, P<0.001			
SA-SAPAS			
No	425 (14.76%)	2454 (85.24%)	2879
Yes	107 (31.47%)	233 (68.53%)	340
Total	532 (16.53%)	2687 (83.47%)	3219

Pearson χ^2 (1) =61.54, P<0.001			
Current mental health condition(s)			
No	399 (15.35%)	2200 (84.65%)	2599
Yes	131 (21.87%)	468 (78.13%)	599
Total	530 (16.57%)	2668 (83.43%)	3198
Pearson χ^2 (1) =14.96, P<0.001			
Childhood trauma (physical, verbal/emotional, sexual childhood trauma with SASTC1 and a family history of domestic violence)			
No	339 (13.94%)	2092 (86.06%)	2431
Yes	185 (25.45%)	542 (74.55%)	727
Total	524 (16.59%)	2634 (83.41%)	3158
Pearson χ^2 (1) =53.50, P<0.001			
ASRS			
No	321 (12.97%)	2153 (87.03%)	2474
Yes	211 (28.32%)	534 (71.68%)	745
Total	532 (16.53%%)	2687 (83.47%)	3219
Pearson χ^2 (1) =97.76, P<0.001			
IAT			
Normal (No)	241 (11.50%)	1855 (88.50%)	2096
Mild (Yes)	207 (22.21%)	725 (77.79%)	932
Moderate (Yes)	80 (43.48%)	104 (56.52%)	184
Severe (Yes)	4 (57.14%)	3 (42.86%)	7

Total	532 (16.53%)	2687 (83.47%)	3219
Pearson χ^2 (2) =165.50, P<0.001			

Table 2. Several variables were associated with increased odds of problematic sexual behavior (PSB) (subtracting SASTC1 and SASTC2) on logistic regression (N=3049, after removing outliers)

Log likelihood= -1149.98		Number of observations=3049, *P*<0.00001, pseudo R^2=0.16
Variable		Odds Ratio, 95% CI
OCD (self-reported current diagnoses with MINI data, baseline: no)	Yes	1.79, CI: 1.12-2.86, *P*=0.015
ASRS (bassline: no)	Yes	1.93, CI: 1.54-2.42, *P*<0.001
IAT (baseline: normal)	Yes	2.00, CI: 1.61-2.49, *P*<0.001
A family history of PSB (a family history of sexual addiction with SASTC2, baseline: no)	Yes	2.05, CI:1.38-3.05, *P*<0.001
SA-SAPAS (baseline: no)	Yes	2.14, CI: 1.59-2.86, *P*<0.001
Childhood trauma (physical, verbal/emotional and sexual childhood trauma, SASTC1, with a family history of domestic violence, baseline: no)	Yes	2.18, CI: 1.69-2.79, *P*<0.001
RCBS (baseline: no)	Yes	2.53, CI:1.74-3.68, *P*<0.001
Gender (baseline: women)	Men	4.40, CI:3.53-5.49, *P*<0.001

Model Accuracy	84.32%
Area under the ROC curve	77.37%
Sensitivity	14.46%
Specificity	98.19%
F-score (harmonic mean of Precision and Recall)	0.23

The model passed the linktest with a significant *P*(hat) (*P*<0.001) and a non-significant *P*(hatsq) (*P*=0.12). The probability of the Hosmer-Lemeshow χ^2 test was insignificant (*P*=0.64), suggesting a good model fit.

Table 3. Several variables were associated with increased odds of problematic sexual behavior (PSB) (subtracting SASTC1 and SASTC2) on machine learning (ML) regression (train ratio:0.8)

Variable	Logistic regression (training with total PSB – subtracting SASTC1 and SASTC2, N=2477, testing with total PSB, N=619)		Validation (training with SAST-R Core subtracting SASTC1 and SASTC2, N=3141, testing with monthly PSB, N=3145)
	Odds Ratio, 95% CI		Odds Ratio, 95% CI
RCBS (baseline: no)	Yes	1.51, CI:1.05-2.16, *P*=0.025	1.77, CI:1.21-2.59, *P*=0.0030
SA-SAPAS (baseline: no)	Yes	1.72, CI:1.33-2.22, *P*<0.001	2.30, CI:1.76-3.01, *P*<0.001
OCD (self-reported current diagnoses with MINI data, baseline: no)	Yes	1.85, CI:1.14-3.02, *P*=0.013	1.31, CI:0.83-2.06, *P*=0.25
A family history of PSB (a family history of sexual addiction with SASTC2, baseline: no)	Yes	1.90, CI:1.33-2.73, *P*<0.001	1.82, CI:1.28-2.59, *P*<0.001
ASRS (bassline: no)	Yes	1.99, CI:1.64-2.41, *P*<0.001	1.89, CI:1.53-2.32, *P*<0.001

Childhood trauma (physical, verbal/emotional and sexual childhood trauma, SASTC1, with a family history of domestic violence, baseline: no)	Yes	2.23 CI:1.81-2.75, *P*<0.001	2.30, CI: 1.84-2.87, *P*<0.001
IAT (baseline: normal)	Yes	2.37, CI:1.97-2.84, *P*<0.001	2.11, CI:1.74-2.55, *P*<0.001
Gender (baseline: women)	Men	4.80, CI:4.01-5.76, *P*<0.001	3.83, CI:3.16-4.65, *P*<0.001
Train performance	Train ROC=75.72 %, Sensitivity=79.70%, Specificity=58.02%		Train ROC=76.16%, Sensitivity=78.31%, Specificity=62.28%
Model Accuracy	74.64%		75.04%
Area under the ROC curve	73.90%		74.26%
Sensitivity	53.33%		59.11%
Specificity	79.76%		76.97%

F-score (harmonic mean of Precision and Recall)	0.45	0.34

ML logistic regression showed increased sensitivity (53.33% vs. 14.46%) and F score (0.45 vs. 0.23) compared to logistic regression.

Figure 1. The correlation matrix of problematic sexual behavior (PSB)

P values are provided in Supplementary Table 3.

Figure 2. Forest plot of logistic regression on problematic sexual behavior (PSB)

A. By logistic regression

B. By machine learning (ML) regression

C. Validation on PSB (train with SAST-R Core, test with monthly PSB) by ML logistic regression

More details, including model evaluations and *P* values, are shown in Table 2 and 3.

Figure 3. Categorical maps in women and men in dimensions 1 and 2

A. Women

B. Men

Biplots are provided in Supplementary Figure 8-9.

Figure 4. Categorical maps in women and men in dimensions 4 and 5

A. Women

B. Men

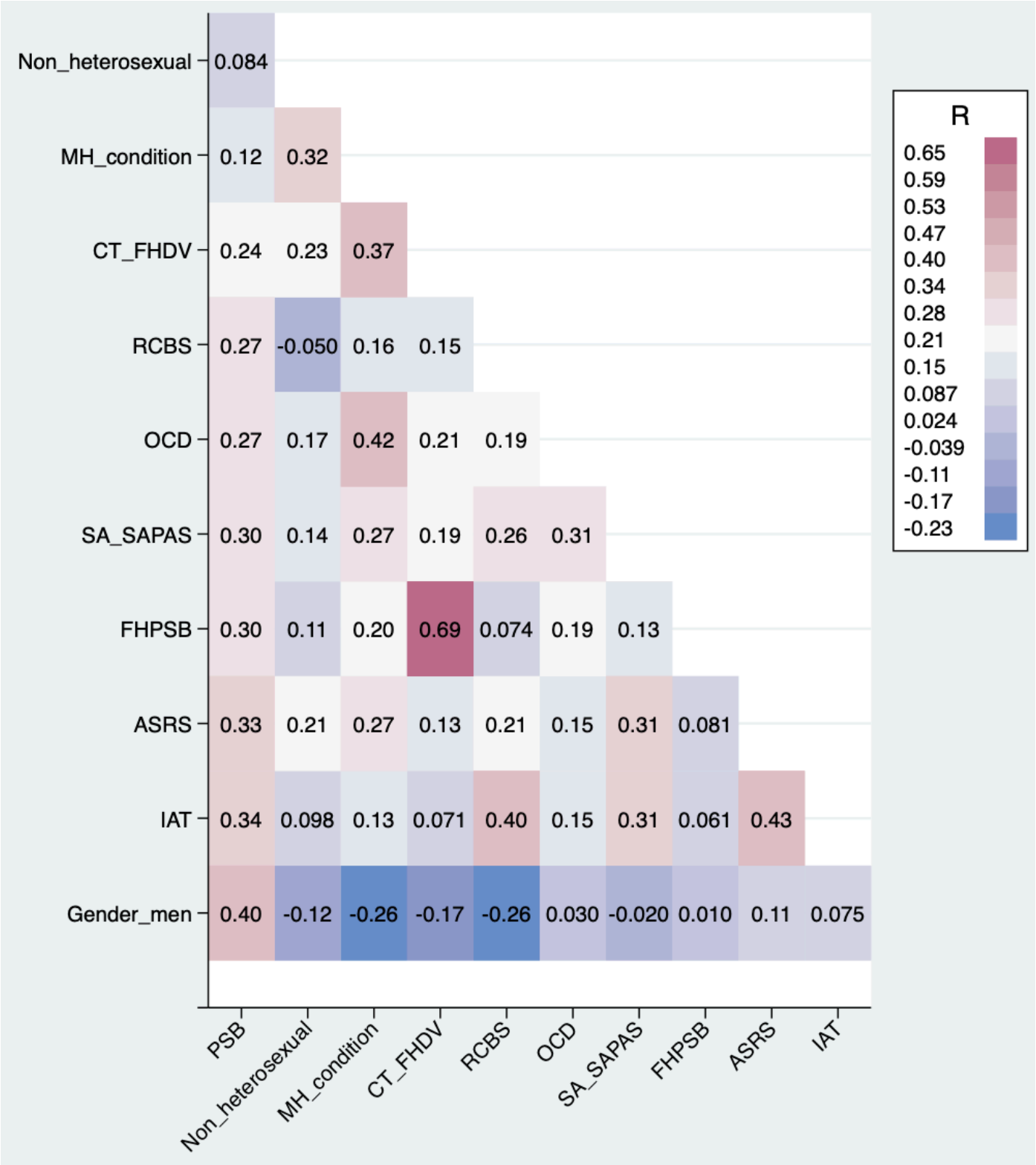
Biplots are provided in Supplementary Figure 8-9.

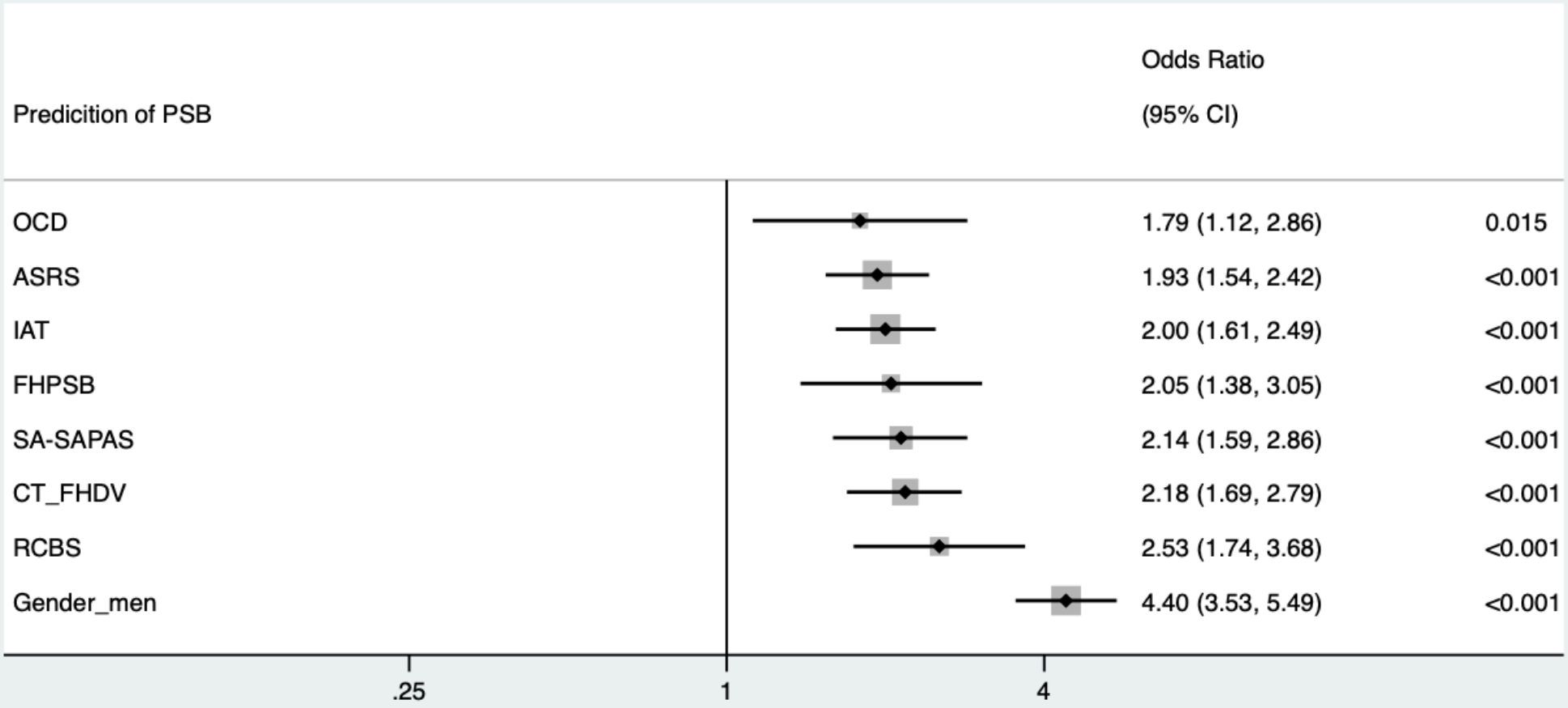
References

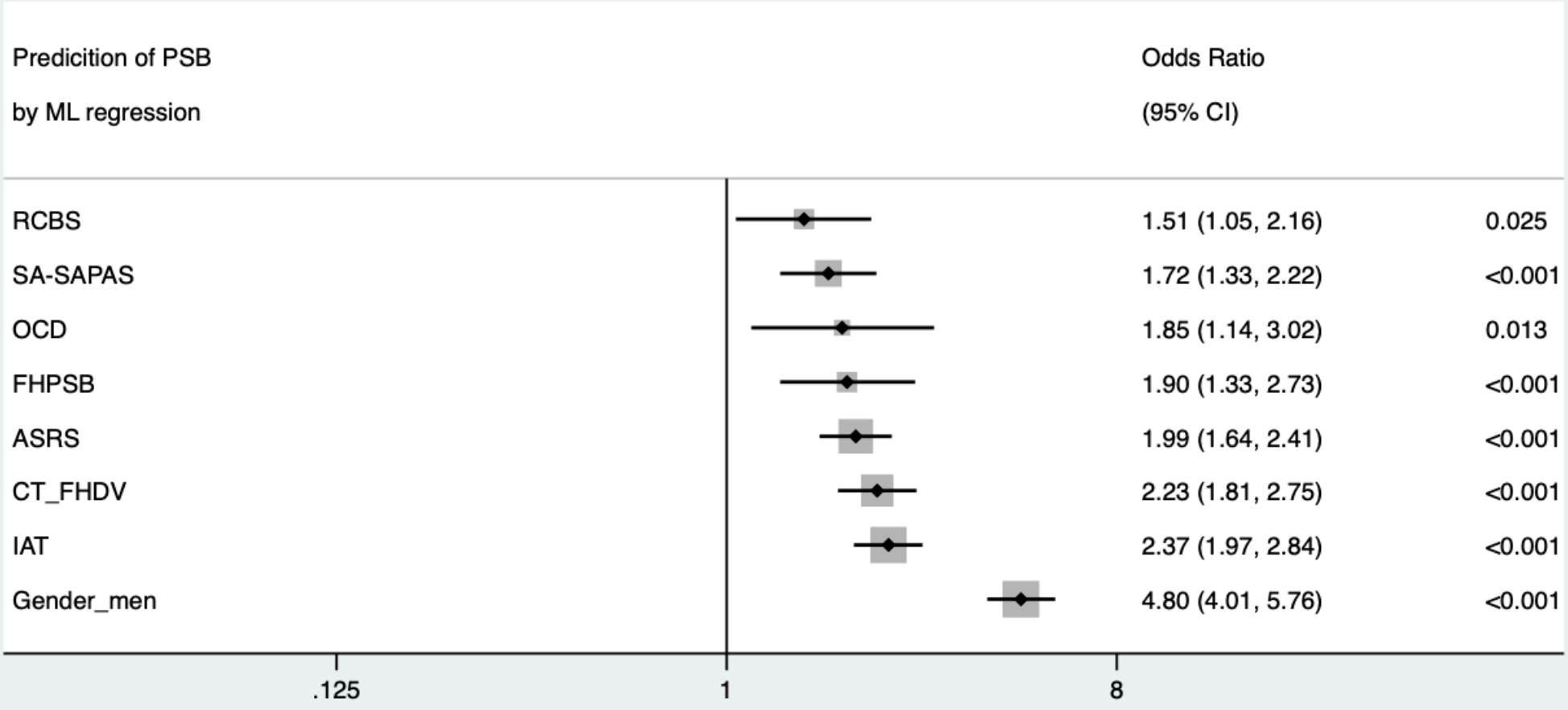
1. Orford J. Hypersexuality: implications for a theory of dependence. *Br J Addict Alcohol Other Drugs*. Sep 1978;73(3):299-10. doi:http://dx.doi.org/10.1111/j.1360-0443.1978.tb00157.x
2. Carnes P. *Don't call it love : recovery from sexual addiction*. Bantam Books; 1991:439 p.
3. Goodman A. Sexual addiction: terminology and theory. *Psychiatric Times* 1998;15(7):22-26.
4. Carnes P. *Sexual addiction*. 8 ed. vol I. Lippincott, Williams & Wilkins: A Wolters Kluwer Company; 2005.
5. Carnes PJ, Hopkins TA, Green BA. Clinical relevance of the proposed sexual addiction diagnostic criteria: relation to the Sexual Addiction Screening Test-Revised. *Journal of addiction medicine*. Nov-Dec 2014;8(6):450-61. doi:http://dx.doi.org/10.1097/ADM.0000000000000080
6. Rosenthal RJ. Inclusion of pathological gambling in DSM-III, its classification as a disorder of impulse control, and the role of Robert Custer. *International Gambling Studies*. 2020;20(1):151-170. doi:https://doi.org/10.1080/14459795.2019.1638432
7. Rehm J, Probst C, Kraus L, Lev-Ran S. *The addiction concept revisited*. Reframing addiction: Policies, processes and pressures. The ALICE RAP project; 2014:103-117.
8. Kraus SW, Krueger RB, Briken P, et al. Compulsive sexual behaviour disorder in the ICD-11. *World Psychiatry*. Feb *2018*;*17*(*1*):*109-110*. doi:10.1002/wps.20499
9. Griffiths MD. Compulsive sexual behaviour as a behavioural addiction: The impact of the internet and other issues. *Addiction*. 2016;111(12):2107-2108. doi:*http://dx.doi.org/10.1111/add.13315*
10. Joannides P. *The challenging landscape of problematic sexual behaviors, including "sexual addiction" and "hypersexuality."*. New directions in sex therapy: Innovations and alternatives. 2012:69-83.
11. Dickenson JA, Gleason N, Coleman E, Miner MH. Prevalence of Distress Associated With Difficulty Controlling Sexual Urges, Feelings, and Behaviors in the United States. *JAMA Netw Open*. Nov 2 *2018*;*1*(*7*):e184468. doi:10.1001/jamanetworkopen.2018.4468
12. Carnes PJ, Green BA, Merlo LJ, Polles A, Carnes S, Gold MS. *PATHOS: a brief screening application for assessing sexual addiction*. *J Addict Med*. Mar *2012*;*6*(*1*):*29-34*. doi:*10.1097/ADM.0b013e3182251a28*
13. Carnes PJ. *Facing the Shadow: Starting Sexual and Relationship Recovery: a Gentle Path to Beginning Recovery from Sex Addiction*. Gentle Path Press; 2015.
14. Klein V, Rettenberger M, Briken P. Self-reported indicators of hypersexuality and its correlates in a female online sample. *J Sex Med*. Aug 2014;11(8):1974-81. doi:10.1111/jsm.12602
15. Reid RC, Li DS, Gilliland R, Stein JA, Fong T. Reliability, validity, and psychometric development of the pornography consumption inventory in a sample of hypersexual men. *J Sex Marital Ther*. 2011;37(5):359-85. doi:10.1080/0092623x.2011.607047
16. Giordano AL, Cecil AL. Religious coping, spirituality, and hypersexual behavior among college students. *Sexual Addiction & Compulsivity*. 2014;21(3):225-239. doi:https://doi.org/10.1080/10720162.2014.936542
17. Cashwell C, L. Giordano A, Lewis T, A. Wachtel K, L. Bartley J. *Using the PATHOS Questionnaire for Screening Sexual Addiction Among College Students: A Preliminary Exploration*. vol 22. 2015.

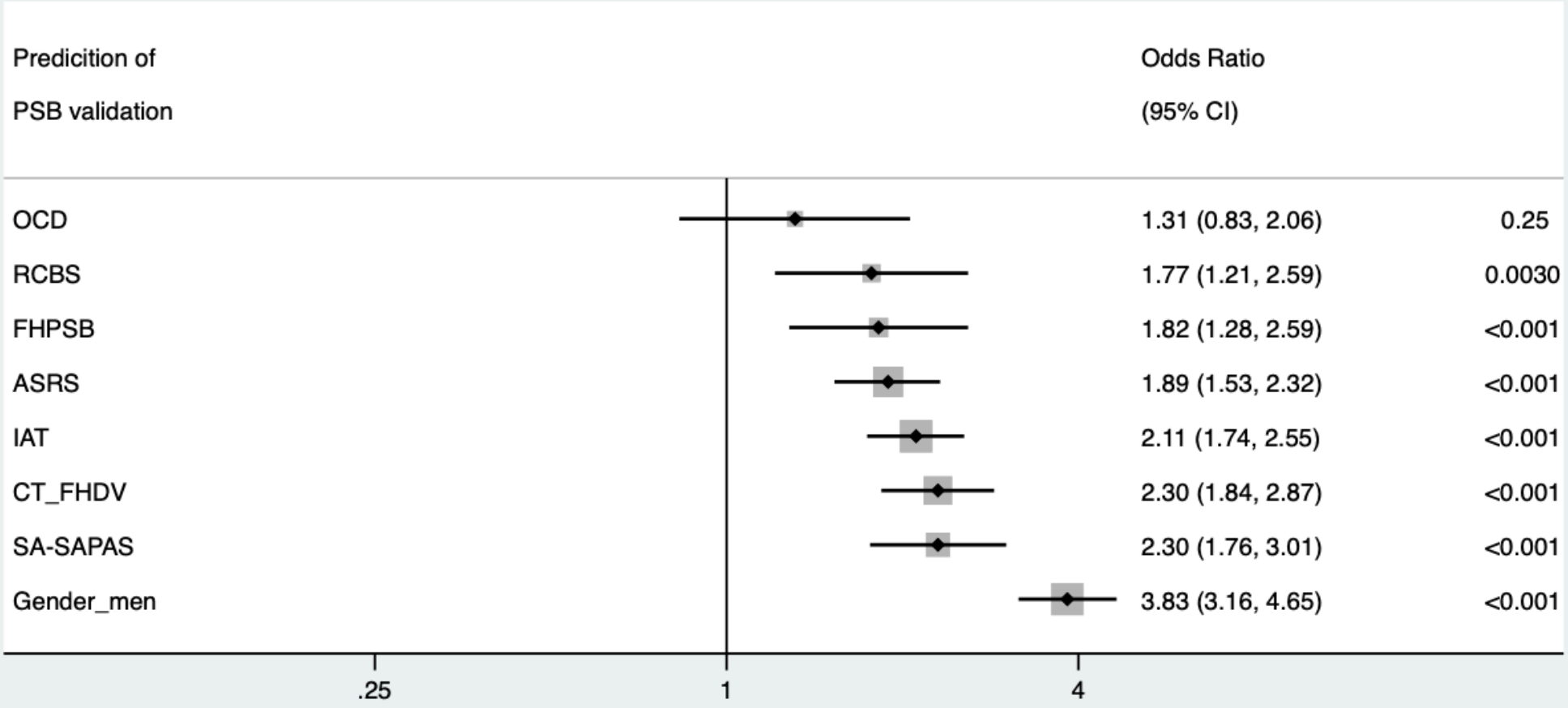
18. *Deneke E, Knepper C, Green BA, Carnes PJ. Comparative study of three levels of care in a substance use disorder inpatient facility on risk for sexual addiction. Sexual Addiction & Compulsivity. 2015;22(2):109-125.*
19. *Carnes PJ. The Sexual Addiction Screening Process. Clinical Management of Sex Addiction. 2019:21-39.*
20. Blankenship R, Laaser M. Sexual Addiction and ADHD: Is There A Connection? *Sexual Addiction & Compulsivity*. 2004/01/01 2004;11(1-2):7-20. doi:10.1080/10720160490458184
21. Ballester-Arnal R, Castro-Calvo J, Giménez-García C, Gil-Juliá B, Gil-Llario M. Psychiatric comorbidity in compulsive sexual behavior disorder (CSBD). *Addictive behaviors*. 2020;107:106384.
22. Grant JE, Lust K, Chamberlain SR. Body dysmorphic disorder and its relationship to sexuality, impulsivity, and addiction. *Psychiatry Research*. 2019/03/01/ 2019;273:260-265. doi:https://doi.org/10.1016/j.psychres.2019.01.036
23. *London S, Quinn K, Scheidell JD, Frueh BC, Khan MR. Adverse Experiences in Childhood and Sexually Transmitted Infection Risk From Adolescence Into Adulthood. Sexually transmitted diseases. Sep 2017;44(9):524-532.* doi:10.1097/OLQ.0000000000000640
24. Mick TM, Hollander E. *Impulsive-compulsive sexual behavior. CNS Spectr*. Dec *2006;11(12)*:944-55. doi:10.1017/s1092852900015133
25. *Carnes P, Green B, Carnes S. The same yet different: Refocusing the Sexual Addiction Screening Test (SAST) to reflect orientation and gender. Sexual Addiction & Compulsivity. 2010;17(1):7-30.*
26. Young KS. Internet addiction: The emergence of a new clinical disorder. *Cyberpsychology & behavior*. 1998;1(3):237-244.
27. *Ridgway NM, Kukar-Kinney M, Monroe KB. An expanded conceptualization and a new measure of compulsive buying. Journal of consumer Research. 2008;35(4):622-639.*
28. Merlhiot G, Mondillon L, Vermeulen N, Basu A, Mermillod M. Adaptation and Validation of the Standardized Assessment of Personality–Abbreviated Scale as a Self-Administered Screening Test (SA-SAPAS). *Journal of Psychology and Psychotherapy*. 2014;4:6. doi:http://dx.doi.org/10.4172/2161-0487.1000164
29. *Kessler RC, Adler L, Ames M*, et al. *The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. Psychol Med. Feb 2005;35(2):245-56.* doi:*10.1017/s0033291704002892*
30. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerström test for nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *British journal of addiction*. 1991;86(9):1119-1127.
31. Rennert L, Denis C, Peer K, Lynch KG, Gelernter J, Kranzler HR. DSM-5 gambling disorder: prevalence and characteristics in a substance use disorder sample. *Experimental and clinical psychopharmacology*. 2014;22(1):50-56. doi:10.1037/a0034518
32. Öncel SY, Gebizlioğlu ÖL, Alioğlu FA. Risk factors for smoking behavior among university students. *Turkish Journal of Medical Sciences*. 2011;41(6):1071-1080. doi:http://dx.doi.org/10.3906/sag-1009-1122
33. Benedan L, Monti GS. Predicting the Risk of Gambling Activities in Adolescence: A Case Study. Springer; 2019:47-57.
34. Nwanganga F, Chapple M. *Practical Machine Learning in R*. John Wiley & Sons; 2020.
35. Chen CY, Yang HCP, Chen CW, Chen TH. Diagnosing and revising logistic regression models: effect on internal solitary wave propagation. *Engineering Computations*. 2008;

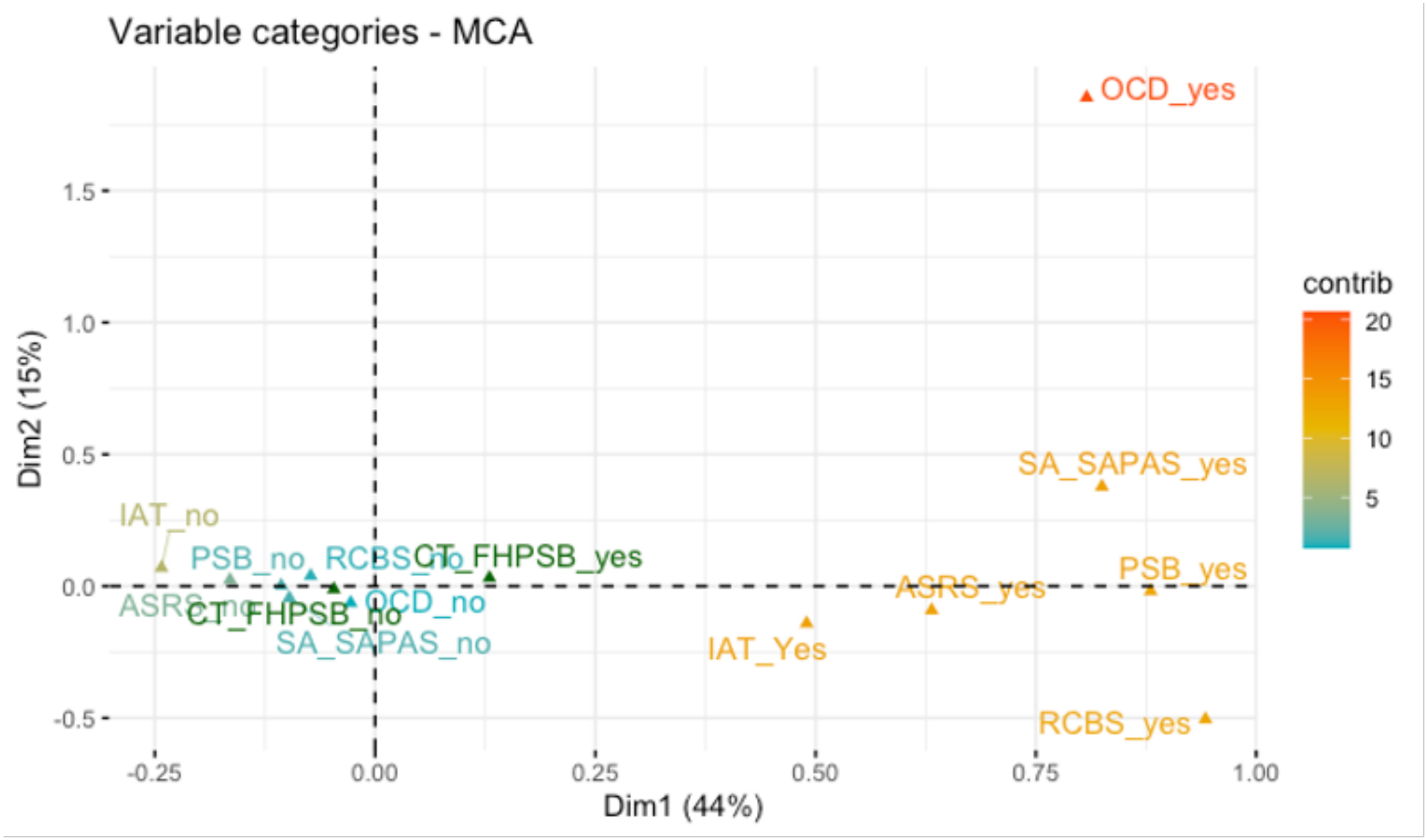
36. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. vol 398. John Wiley & Sons; 2013.
37. White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Comput Stat Data Anal*. 2010;54(10):2267-2275. doi:10.1016/j.csda.2010.04.005
38. Momeni A, Pincus M, Libien J. Imputation and Missing Data. *Introduction to Statistical Methods in Pathology*. Springer International Publishing; 2018:185-200.
39. Shrestha N. Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*. 2020/06/16 2020;8(2):39-42. doi:http://dx.doi.org/10.12691/ajams-8-2-1
40. Kim JH. *Multicollinearity and misleading statistical results*. *Korean J Anesthesiol*. Dec *2019*;*72*(*6*):*558-569*. doi:*10.4097/kja.19087*

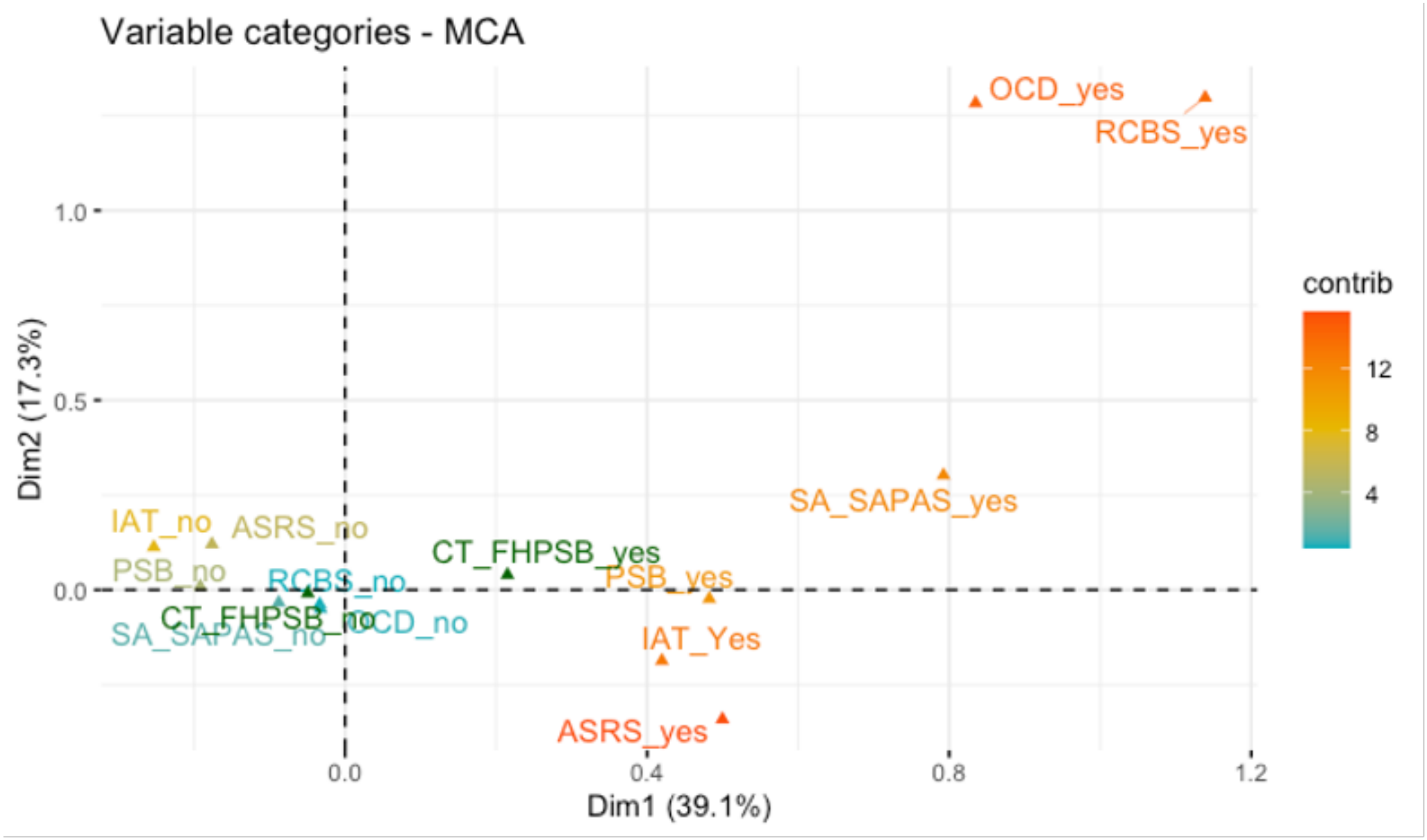

R
0.65
0.59
0.53
0.47
0.40
0.34
0.28
0.21
0.15
0.087
0.024
-0.039
-0.11
-0.17
-0.23
Non_heterosexual 0.084
MH_condition 0.12 0.32
CT_FHDV 0.24 0.23 0.37
RCBS 0.27 -0.050 0.16 0.15
OCD 0.27 0.17 0.42 0.21 0.19
SA_SAPAS 0.30 0.14 0.27 0.19 0.26 0.31
FHPSB 0.30 0.11 0.20 0.69 0.074 0.19 0.13
ASRS 0.33 0.21 0.27 0.13 0.21 0.15 0.31 0.081
IAT 0.34 0.098 0.13 0.071 0.40 0.15 0.31 0.061 0.43
Gender_men 0.40 -0.12 -0.26 -0.17 -0.26 0.030 -0.020 0.010 0.11 0.075
PSB
Non_heterosexual
MH_condition
CT_FHDV
RCBS
OCD
SA_SAPAS
FHPSB
ASRS
IAT

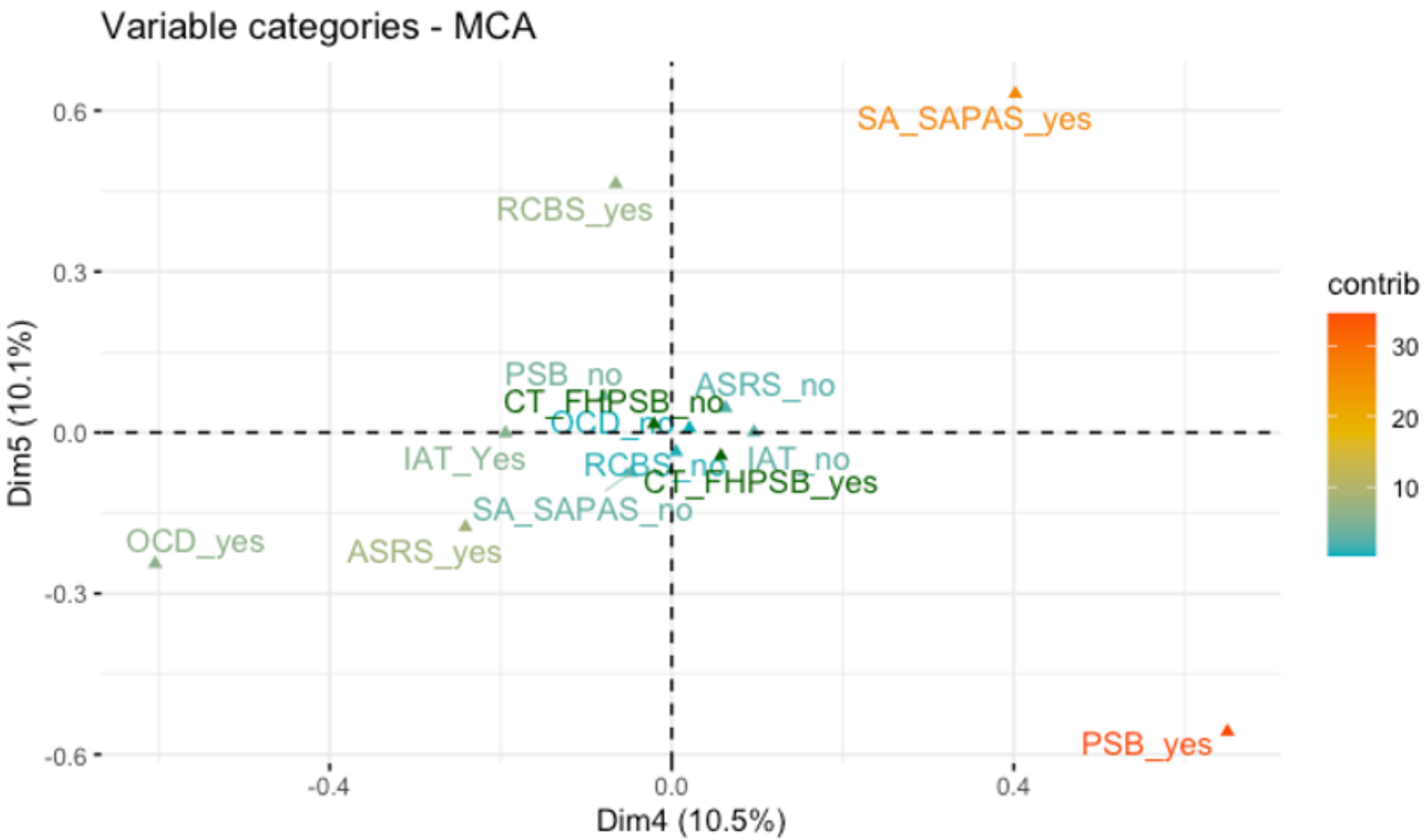

Predicition of PSB
Odds Ratio
(95% CI)
OCD
1.79 (1.12, 2.86)
0.015
ASRS
1.93 (1.54, 2.42)
<0.001
IAT
2.00 (1.61, 2.49)
<0.001
FHPSB
2.05 (1.38, 3.05)
<0.001
SA-SAPAS
2.14 (1.59, 2.86)
<0.001
CT_FHDV
2.18 (1.69, 2.79)
<0.001
RCBS
2.53 (1.74, 3.68)
<0.001
Gender_men
4.40 (3.53, 5.49)
<0.001
.25
1
4

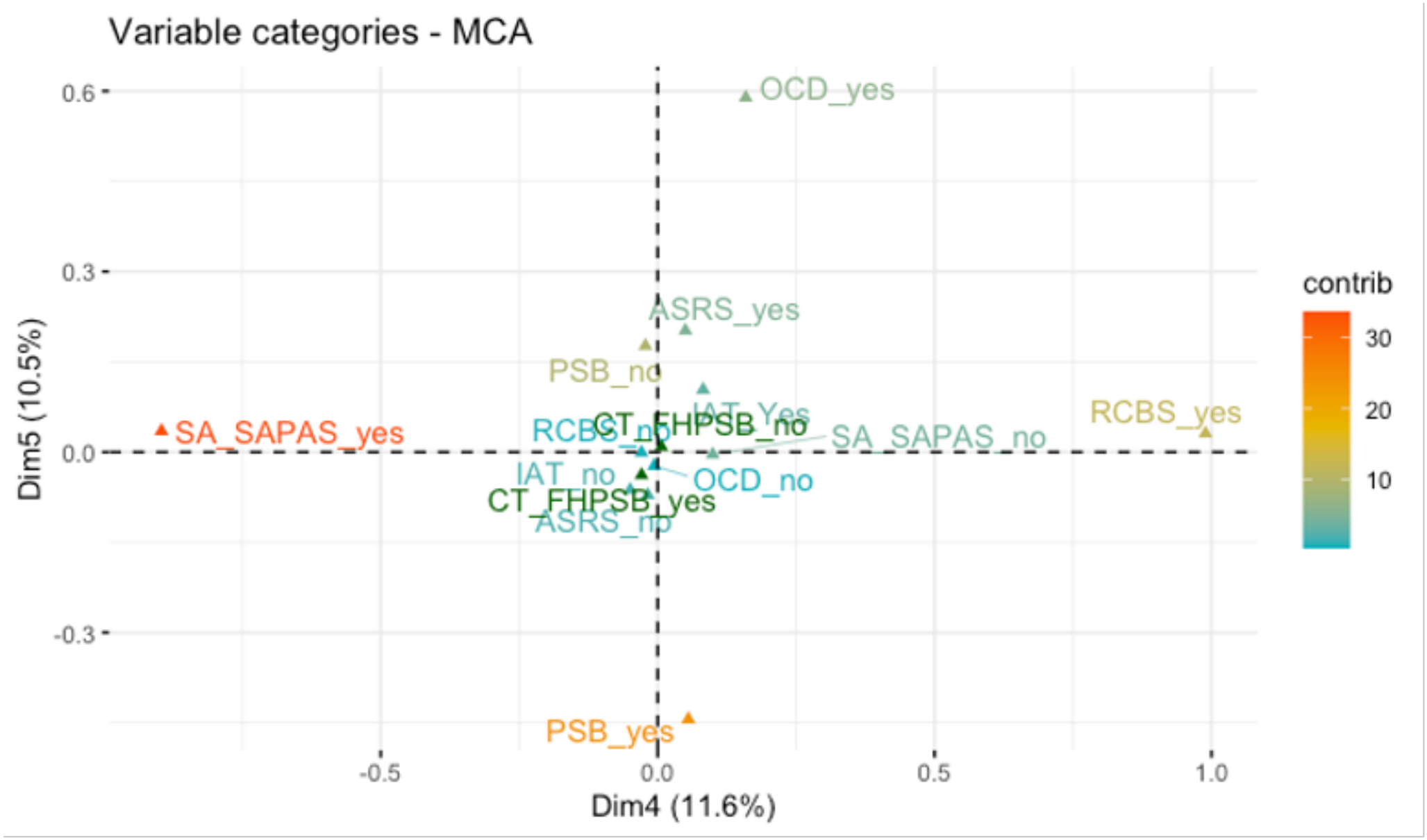

Predicition of PSB
by ML regression
Odds Ratio
(95% CI)
RCBS
1.51 (1.05, 2.16)
0.025
SA-SAPAS
1.72 (1.33, 2.22)
<0.001
OCD
1.85 (1.14, 3.02)
0.013
FHPSB
1.90 (1.33, 2.73)
<0.001
ASRS
1.99 (1.64, 2.41)
<0.001
CT_FHDV
2.23 (1.81, 2.75)
<0.001
IAT
2.37 (1.97, 2.84)
<0.001
Gender_men
4.80 (4.01, 5.76)
<0.001
.125
1
8


Predicition of
PSB validation
Odds Ratio
(95% CI)
OCD
1.31 (0.83, 2.06)
0.25
RCBS
1.77 (1.21, 2.59)
0.0030
FHPSB
1.82 (1.28, 2.59)
<0.001
ASRS
1.89 (1.53, 2.32)
<0.001
IAT
2.11 (1.74, 2.55)
<0.001
CT_FHDV
2.30 (1.84, 2.87)
<0.001
SA-SAPAS
2.30 (1.76, 3.01)
<0.001
Gender_men
3.83 (3.16, 4.65)
<0.001
.25
1
4


Variable categories - MCA
OCD_yes
SA_SAPAS_yes
PSB_yes
ASRS_yes
IAT_Yes
RCBS_yes
IAT_no
PSB_no
RCBS_no
CT_FHPSB_yes
ASRS_no
CT_FHPSB_no
OCD_no
SA_SAPAS_no
contrib
20
15
10
5
1.5
1.0
0.5
0.0
-0.5
Dim2 (15%)
-0.25
0.00
0.25
0.50
0.75
1.00
Dim1 (44%)


Variable categories - MCA
OCD_yes
RCBS_yes
SA_SAPAS_yes
IAT_no
ASRS_no
PSB_no
RCBS_no
CT_FHPSB_yes
PSB_yes
CT_FHPSB_no
OCD_no
SA_SAPAS_no
IAT_Yes
ASRS_yes
1.0
0.5
0.0
0.0
0.4
0.8
1.2
Dim2 (17.3%)
Dim1 (39.1%)
contrib
12
8
4


Variable categories - MCA
SA_SAPAS_yes
RCBS_yes
PSB_no
ASRS_no
CT_FHPSB_no
OCD_no
IAT_Yes
RCBS_no
IAT_no
CT_FHPSB_yes
SA_SAPAS_no
OCD_yes
ASRS_yes
PSB_yes
0.6
0.3
0.0
-0.3
-0.6
-0.4
0.0
0.4
Dim5 (10.1%)
Dim4 (10.5%)
contrib
30
20
10


Variable categories - MCA
OCD_yes
ASRS_yes
PSB_no
IAT_Yes
RCBS_yes
SA_SAPAS_yes
RCBS_no
CT_FHPSB_no
SA_SAPAS_no
IAT_no
OCD_no
CT_FHPSB_yes
ASRS_no
PSB_yes
0.6
0.3
0.0
-0.3
-0.5
0.0
0.5
1.0
Dim5 (10.5%)
Dim4 (11.6%)
contrib
30
20
10