

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

3
4 Shui Jiang (MSc)¹, Keanna Wallace (BSc)^{1,2}, Esther Yang (MSc)^{1,2,3}, Leslie Roper (MSc)^{1,2},
5 Garima Aryal (MSc)^{1,2}, Dawon Lee (MD)^{1,2,4}, Rohit J Lodhi (MD, PhD, MRCPsych,
6 FRANZCP)^{2,5}, Rick Isenberg (MD)⁶, Bradley Green (PhD)⁷, David Wishart (PhD)⁸, Katherine J
7 Aitchison (BM BCh, PhD, FRCPsych)^{1,2,9,10*}

8
9 ¹Department of Medical Genetics, Faculty of Medicine and Dentistry, College of Health Sciences,
10 University of Alberta, Canada

11 ²Department of Psychiatry, Faculty of Medicine and Dentistry, College of Health Sciences, University of
12 Alberta, Canada

13 ³Department of Emergency Medicine, Faculty of Medicine and Dentistry, College of Health Sciences,
14 University of Alberta, Canada

15 ⁴Department of Medicine, Faculty of Medicine and Dentistry, College of Health Sciences, University of
16 Alberta, Canada

17 ⁵Department of Psychiatry, University of Western Ontario, Canada

18 ⁶Psychological Counselling Services, Ltd., Scottsdale, USA

19 ⁷Department of Psychology and Counseling, University of Texas at Tyler, USA

20 ⁸Department of Biological Sciences, Faculty of Science, College of Natural and Applied Sciences,
21 University of Alberta, Canada

22 ⁹Neuroscience and Mental Health Institute, University of Alberta, Canada

23 ¹⁰Division of Clinical Sciences, Psychiatry Section, Northern Ontario School of Medicine

24
25 *Corresponding author:

26 Email: kaitchis@ualberta.ca (KJA) Phone number: 780-492-4018

27
28 **Sources of support:** The study was funded by an Alberta Centennial Addiction and Mental
29 Health Research Chair and transitional funding (to KJA), and by the American Foundation for
30 Addiction Research (to KJA). Several scholarships, including a University of Alberta Dean's
31 Doctoral Student Award 2021-2022, a University of Alberta Medical Science Graduate Program
32 Scholarship 2020, a University of Alberta Doctoral Recruitment Scholarship 2017 and an

33 Alberta Graduate Excellence Scholarship 2020 provided additional funding for the first author
34 (SJ). A Fulbright-Canada-Palix Foundation grant to Dr. Patrick Carnes facilitated his
35 contributions to study design.

36 **Conflict of interest:** Dr. Carnes has a conflict of interest with the American Foundation for
37 Addiction Research, and for this reason, he is not an author on the paper. Dr. Aitchison has
38 received two research grants from Janssen Inc., Canada (fellowship grants for trainees). The
39 analysis and data interpretation were conducted independently.

40 **Abstract word count:** 250

41 **Manuscript word count:** 3124

42 **Number of references:** 40

43 **Abstract**

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

50 **Methods:** A Canadian university Office of the Registrar invited 68,846 eligible students and
51 postdoctoral fellows. Out of 4710 expressing interest in participating, 3359 completed on-line
52 questionnaires, and 1801 completed the Mini-International Neuropsychiatric Interview (MINI).
53 PSB was measured by combining those screening positive (≥ 6) on the Sexual Addiction
54 Screening Test-Revised (SAST-R) Core with those self-reporting PSB. Current mental health
55 condition(s) and childhood trauma were measured by self-report. OCD was assessed by a
56 combination of self-report and MINI data.

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

62 **Conclusions:** Factors that have previously been associated with RDS and OCD are both
63 associated with increased odds of PSB. The factors associated with RDS appear to contribute to a
64 separate data cluster from OCD, and to lie closer to PSB.

65

66 INTRODUCTION

67 A theory of hypersexuality with dependence was first proposed by Orford in 1978.¹ The term sex
68 addiction was used by Carnes in 1983, 1990, and 1991² and by Goodman in 1998.³ Continuation
69 of a sexual behavior despite adverse consequences and/or distress caused or worsened by the
70 sexual behavior has been a consistent theme in the literature,^{1,3,4} despite differences of opinion as
71 to whether the condition should be classified as an addiction or as an impulse control disorder.
72 Diagnostic criteria proposed by Carnes in 2005⁴ led to sex addiction being included in the
73 Diagnostic and Statistical Manual of Mental Disorders (DSM)-III-R (302.87). However, despite
74 draft proposed criteria for the inclusion of sex addiction and internet addiction in DSM-5,⁵ the
75 only behavioral addiction currently included in the main body of DSM-5 (American Psychiatric
76 Association 2013) is gambling disorder (312.31), which was first included in DSM as an impulse
77 control disorder.⁶ In the International Statistical Classification of Diseases (ICD), compulsive
78 sexual behavior disorder was proposed for inclusion as an impulse control disorder in 2014,⁷
79 which occurred in 2018.⁸ However, the scientific discussion about whether or not compulsive
80 sexual behavior disorder could constitute a manifestation of a behavioral addiction was
81 acknowledged,⁹ and it was predicted (as in the case of problematic gambling),⁶ that
82 understanding would evolve as research elucidated the phenomenology and neurobiological
83 underpinnings of the condition.

84 In this article, the term problematic sexual behavior (PSB) is used because it focuses on
85 the behavior rather than on the potentially heterogeneous etiology and pathology.¹⁰ PSB is
86 characterized by repetitive sexual behaviors and it is associated with uncontrolled sexual
87 urges/impulses and distress,¹¹ and social and functional impairments.^{8,11} It has been estimated that
88 the prevalence of PSB is 3-6% in adults,^{4,12-14} with higher frequencies in specific populations,

89 including those in post-secondary education. Reid (2011) reported that 19% of college men met
90 criteria for hypersexuality,¹⁵ while Giordano and Cecil (2014) found 11.1% of men and women
91 in an undergraduate sample met these criteria.¹⁶ Cashwell et al. (2015) reported that out of 379
92 undergraduates, 21.2% of men and 6.7% of women screened positive, i.e., scored in a range
93 indicating that they should be offered further assessment for sex addiction.¹⁷

94 Various vulnerability/risk factors have been associated with PSB, including those
95 previously associated with the construct known as reward deficiency syndrome (RDS)
96 (substance¹⁸ and non-substance use disorders,¹⁹ ADHD,²⁰ and personality disorder),²¹ as well as
97 other psychiatric disorders,²² and childhood trauma.²³ In terms of psychiatric disorders, owing to
98 the classification of PSB within ICD-10 as an impulse control disorder, we specifically included
99 obsessive-compulsive disorder. Mick and Hollander (2006) hypothesized that impulsivity
100 initiated the early stage of PSB, with compulsivity being involved in repetitive behaviors and
101 hence in the maintenance of PSB.²⁴ The magnitude of the contribution of the various different
102 postulated mechanisms, and whether PSB encompasses various syndromes with different
103 etiologies is not at present known. Moreover, systematic data on the prevalence of PSB and
104 associated sociodemographic factors across diverse populations, including non-treatment seekers
105 remain to be provided.⁸

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

112 **METHODS**

113 The study inclusion criteria were as follows: at least 18 years of age, undergraduates, graduates,
114 postdoctoral fellows, and recently convocated students registered in at least one course in the
115 preceding year, except for those who had completely withdrawn after registration without any
116 reasons to not return in the next academic term or not be on campus (such as suspension), and
117 being able to answer in English. A Canadian university Office of the Registrar invited 68,846
118 eligible students and postdoctoral fellows by email. Students interested in participating emailed
119 the study team. These students were then sent an email invitation to review the participant
120 information and complete consent on-line, followed by the screening measures (hosted by the
121 Qualtrics platform), with a subsequent email inviting them to complete the Mini-International
122 Neuropsychiatric Interview (MINI) version 7.0.2 on-line.

123 *Measures*

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

130 Measures used to screen for internet addiction, compulsive buying, personality disorders,
131 ADHD, nicotine dependence, and pathological gambling were: the Internet Addiction Test
132 (IAT),²⁶ the Richmond Compulsive Buying Scale (RCBS),²⁷ the 8-item Standardized Assessment
133 of Personality–Abbreviated Scale as a Self-Administered Screening Test (SA-SAPAS),²⁸ the full
134 Adult ADHD Self-Report Scale (ASRS-v1.1),²⁹ the Fagerström Test for Nicotine Dependence

135 (FTND),³⁰ and the DSM-V Pathological Gambling Diagnostic Form (PGDF),³¹ respectively.
136 Gender, sexual orientation, a family history of domestic violence or of sex addiction, and having
137 received a diagnosis of OCD, or any other mental health condition(s), childhood physical,
138 emotional/verbal, or sexual trauma were additionally collected by self-report. The MINI was
139 used to output psychiatric disorders by DSM-5 criteria.

140 ***Data analysis***

141 ***Demographic data and logistic regression***

142 Data were analyzed by STATA (Stata/SE 16) and R (Version 3.6.3) after dropping the missing
143 data. The PSB variable was created by combining those screening positive on the SAST-R Core
144 with those endorsing monthly self-reported PSB, after subtracting the two SAST-R Core
145 questions referred to below. Previous or current physical/verbal, emotional/sexual childhood
146 trauma and a family history of domestic violence (11.77% 387/3288) were regrouped as total
147 childhood trauma due to the relatively low rates of endorsement for the former group (physical
148 trauma: 2.98%, 99/3322; emotional trauma: 6.87%, 232/3322; and sexual trauma: 3.31%,
149 110/3212). The SAST-R Core question 1 (which enquires about a history of sexual abuse in
150 childhood or adolescence, SASTC1) was combined with total childhood trauma. Self-reported a
151 family history of sex addiction (“yes”/ “no”) data were combined with SAST-R Core question 2
152 (which enquires about parental trouble with sexual behavior, SASTC2) to create a new family
153 history of PSB variable. IAT was regrouped as “yes” (N=1174, combining mild, moderate, and
154 severe groups) and “no” (N=2160, the “normal” group).

155 Differences in data distributions between those with and without PSB were assessed by
156 Pearson’s χ^2 test for categorical variables and by non-parametric testing (Mann-Whitney U test)
157 for linear non-normally distributed variables. For correlational analyses, tetrachoric (for binary

158 variables) and Kendall's τ (for categorical variables) correlations were utilized. SA-SAPAS,
159 OCD, ASRS, IAT, childhood trauma, a family history of PSB and gender were used as
160 covariates in the regression analyses. All reported P values are unadjusted for multiple testing.
161 Multiple correspondence analysis (MCA) was used for data clustering.

162 *Machine learning approaches*

163 The “Haven” and “dplyr” packages were used for data input. The “forcats” and “creditmodel”
164 packages were used for recoding and splitting the dataset (into 80% training and 20% testing),
165 respectively. The Synthetic Minority Oversampling Technique (SMOTE) from the “Caret”
166 package was used to address data imbalance in the training set.³² Leave-one-out cross-validation
167 (LOOCV) in the “Caret” package was used for resampling.³³ Logistic regression (two-class
168 summary) in the “caret” package was used for regression. Accuracy, area under the receiver
169 operating characteristic (ROC) curve, sensitivity, specificity, and F score were used to evaluate
170 the models. The “pROC” package was used for the area under the ROC curve.

171 *Multiple correspondence analysis (MCA)*

172 Multiple correspondence analysis was conducted on OCD, RCBS, SA-SAPAS, ASRS, IAT, and
173 childhood trauma, stratifying by gender (women N=2084 and men N=1012). As a family history
174 of PSB was correlated with the combined childhood trauma variable (tetrachoric $\rho=0.69$,
175 $P<0.001$, N=3019, Figure 1, Supplemental Digital Content, Table 3), these variables were
176 combined as a supplementary childhood trauma-related variable. The “FactoMineR,”
177 “factoextra,” and “GDAtools” packages in R were used for analysis and data visualization.

178 **RESULTS**

179 Out of 4710 interested in participating, 237 withdrew from the study, 3359 completed online
180 questionnaires using the Qualtrics platform, and 1801 completed the MINI (Supplemental Digital
181 Content, Figure 1).

182 *Demographic data*

183 The mean score on the SAST-R Core Scale among all participants was 2.23 (SD=2.73), with
184 12.18% (407/3341) of participants scoring at least six (the threshold for screening positive in the
185 general population.²⁵ Of the 20 items composing the SAST-R Core questions, item 11 (Do you
186 hide some of your sexual behavior from others?) was the most endorsed (1234/3341,
187 endorsement rate: 36.78%), while item 9 (“Are any of your sexual activities against law?”) was
188 the least endorsed (35/3342; endorsement rate: 1.05%) (Supplemental Digital Content, Table 1).
189 The frequency of participants included in the PSB variable was 16.53% (532/3219). There was a
190 significant difference in the distribution of the following variables by those with and without
191 PSB: gender, sexual orientation, OCD, RCBS, a family history of PSB, SA-SAPAS, current
192 mental health condition(s), childhood trauma, ASRS, and IAT (Table 1, and Supplemental
193 Digital Content, Table 2).

194 Correlational analyses in the whole dataset after dropping the missing values showed that
195 PSB was correlated with non-heterosexual sexual orientation (baseline: heterosexual), current
196 mental health condition(s), childhood trauma, RCBS, OCD, SA-SAPAS, a family history of
197 PSB, ASRS, IAT and gender (baseline: women) (Figure 1, Supplemental Digital Content, Table
198 3). Interestingly, a family history of PSB was shown to be strongly correlated with childhood
199 trauma in our dataset (tetrachoric $\rho=0.69$, $P<0.001$, $N=3019$), and both were correlated with
200 PSB. For example, of those responding “yes” to a family history of sex addiction, 32.65%
201 (16/49) screened positive on the SAST-R Core. Our data also showed that: (A) PSB was

202 correlated with unadjusted SAST-R Core (tetrachoric $\rho=0.97$, $P<0.001$, $N=3219$) and self-
203 reported monthly PSB (tetrachoric $\rho=0.70$, $P<0.001$, $N=3271$), (B) self-reported OCD was
204 correlated with OCD MINI diagnosis (tetrachoric $\rho=0.47$, $P<0.001$, $N=1785$) (Supplemental
205 Digital Content, Table 4).

206 ***Regression models***

207 ***Logistic regression***

208 Outliers were removed from the logistic regression ($N=47$, with 8 categorized as “yes” to having
209 PSB, Supplemental Digital Content, Figure 2) if their Pearson standardized residual had an
210 absolute value of more than 2,³⁴ or Hosmer-Lemeshow $\Delta\hat{\beta}$ more than four.³⁵ After outlier
211 removal as above, no further outliers needed to be removed on the basis of leverage values (a
212 measure of the distance between independent variable observations). The following variables
213 were associated with PSB (pseudo $R^2=0.16$): OCD ($P=0.015$), ASRS ($P<0.001$), IAT ($P<0.001$),
214 a family history of PSB ($P<0.001$), SA-SAPAS ($P<0.001$), childhood trauma ($P<0.001$), RCBS
215 ($P<0.001$) and gender (men) ($P<0.001$) (Figure 2).

216 ***Machine Learning (ML) logistic regression***

217 Our ML logistic regression showed that RCBS ($P=0.0030$), SA-SAPAS ($P<0.001$), OCD
218 ($P<0.001$), a family history of PSB ($P<0.001$), ASRS ($P<0.001$), childhood trauma ($P<0.001$),
219 IAT ($P<0.001$), and gender (men) ($P<0.001$) were all associated with increased odds of PSB
220 (Figure 2B). A validation of the existing model was also undertaken. The validation model was
221 trained with SAST-R Core (subtracting SASTC1 and SASTC2, $N=3141$) and tested with self-
222 reported PSB ($N=3145$, Figure 2C). Sensitivity analyses were performed using unadjusted
223 SAST-R Core and family history of sexual addiction by logistic regression and ML logistic
224 regression (Supplemental Digital Content, Figure 3-4).

225 ***Multiple correspondence analysis (MCA)***

226 The MCA results showed that in both women and men, the first two dimensions accounted for
227 the most explained variance (Supplemental Digital Content, Figure 5). PSB was strongly
228 represented in both men and women in dimensions 1, 2, 4 and 5, although showing slightly
229 different data clustering patterns by gender in these dimensions (Figure 3-4). In dimension 1,
230 PSB clustered with all variables in women and men. In dimension 2 (Figure 3 and Supplemental
231 Digital Content, Figure 5), PSB clustered with IAT and ASRS in men, and with RCBS, IAT and
232 ASRS in women. In women and men, dimensions 4 and 5 illustrate that PSB was clustered
233 relatively closely to the composite childhood trauma variable. (Figure 4, Supplemental Digital
234 Content, Figure 5). The quality of representation and contributions of each variable are shown in
235 Supplemental Digital Content, Table 6.

236 **DISCUSSION**

237 The frequency of problematic sexual behavior, as defined by screening positive on the SAST-R
238 Core, was 8.49% in the women and 19.7% in the men in this post-secondary education sample,
239 consistent with prior reports in this population.¹⁵⁻¹⁷ Factors consistently associated with increased
240 odds of PSB were gender, mental health conditions (ADHD, personality disorder, and OCD),
241 other problematic behaviors such as internet addiction (IAT) and compulsive buying (RCBS),
242 childhood trauma (including a family history of domestic violence), and a family history of PSB.
243 Our results also showed that the self-reported data were correlated with the screening tool and/or
244 MINI data (Supplemental Digital Content, Table 3).

245 Model validation (training with SAST-R Core subtracting the first two questions, and
246 testing with self-reported PSB) resulted in very similar model fitting statistics (Figures 2B and
247 2C). A sensitivity analysis (Supplemental Digital Content, Figure 3-4) was undertaken using the

248 original variables, namely SAST-R Core (without subtracting the first two question), childhood
249 trauma (without adding the first question of SAST-R Core or a family history of domestic
250 violence) and a family history of sex addiction (without incorporating SASTC2). The logistic
251 regression model using these variables (pseudo $R^2=0.13$) revealed that RCBS ($P=0.033$), ASRS
252 ($P<0.001$), IAT ($P<0.001$), OCD ($P<0.001$), childhood trauma ($P<0.001$), and gender (baseline:
253 women, $P<0.001$) remained significant. A family history of sex addiction was the only variable
254 that became non-significant ($P=0.17$), probably due to the lower frequency of endorsement of
255 this variable (N=49) compared to that of SASTC2. The combined variable (a family history of
256 sex addiction with SASTC2, 6.55% endorsement: 214/3268), was, however, significant in all
257 models. Using the new PSB variable, including childhood trauma and a family history of PSB
258 but excluding sexual orientation as well as current mental health condition(s) improved the
259 model fit slightly in the logistic regression models (e.g., sensitivity: unadjusted 77.20% vs
260 adjusted 79.70%, F score: unadjusted 0.42 vs. adjusted 0.45). In general, the ML logistic
261 regression models showed a higher sensitivity and F score compared to conventional logistic
262 regression.

263 The variables previously associated with the RDS construct, such as SA-SAPAS, ASRS,
264 and IAT, were associated with increased odds of PSB. In the final logistic (Figure 2) and ML
265 logistic regressions (Figure 2), RCBS was associated with increased odds of PSB. Of note, there
266 were gender differences regarding RCBS: RCBS was negatively correlated with men (Figure 1,
267 baseline: women), and PSB displayed a weaker association with RCBS in men on MCA (Figure
268 3).

269 In addition, interestingly, the distribution of PSB differed significantly between OCD-
270 positive and -negative groups ($P<0.001$, Table 1), and PSB was correlated with OCD ($\rho=0.27$,

271 $P < 0.001$, Figure 1, Supplemental Digital Content, Table 3). In our final regression models and
272 sensitivity analyses (by both conventional and ML logistic regression), OCD was associated with
273 increased odds of PSB (OR 1.79, $P = 0.015$, and OR 1.85, $P = 0.013$ for conventional and ML
274 logistic regression respectively; sensitivity models: OR 2.28, $P < 0.001$, and OR 1.73, $P = 0.033$,
275 respectively).

276 ***Limitations***

277 Limitations of this study include the following. Firstly, as respondents to the email from
278 the registrar knew the subject area of the study, there is a potential selection bias. However, of
279 note, the frequency of PSB in our dataset was comparable to that seen in other studies of adults
280 in post-secondary education as mentioned above.^{12,17} Subsample sizes were another limiting
281 factor. Specifically, as the frequency of endorsement of OCD was relatively low (120/3206,
282 3.74%), although it appears on the MCA that PSB clusters more closely to the measures of
283 reward deficiency than to OCD, this should not be overinterpreted. For sexual orientation and
284 gender, the smaller groups were combined in the former case and dropped in the latter. For the
285 former, bisexual (366/3371, 10.86%), homosexual (129/3371, 3.83%), and other (91/3371,
286 2.70%) were regrouped as “non-heterosexual sexual orientation.” Due to the small numbers in
287 the gender minority groups (trans and other: 47/3375, 1.39%), only men and women were
288 included in the logistic regression analysis. Another potential limitation is that we dropped
289 missing values (rather than using imputation). However, if the proportion of missing data is
290 relatively small (5% or less), this should not be problematic.³⁶ We also dropped two measures,
291 the FTND (0.45%, 15/3332) and PGDF (2.28%, 76/3332) due to very low endorsement rates
292 (which were lower than in another university student sample).³⁷ Although various methods may
293 be used to impute data, some such methods (e.g., monotone imputation, or chained equations)

294 may introduce bias, and perfect (inaccurate) predictions, as the methods themselves involve the
295 fitting of a regression model for a categorical outcome.³⁸ In our sample, the feature with the most
296 missing data in the SAST-R Core scale was question 14, which is about feeling depressed after
297 having sex, with only 16 out of 3341 (0.48%) endorsed (and out of those 16 participants, only 1
298 scored as SAST-R Core positive). Another limitation is that logistic regression is sensitive to
299 high correlations between predictor variables. Although several variables were correlated with
300 each other in our dataset, the correlation coefficients were less than 0.80,³⁹ and the mean variance
301 inflation factor (VIF) and the condition number were at acceptable levels (VIF<10, condition
302 number<10) in our final models,⁴⁰ indicating that multi-collinearity was not a significant
303 concern. A power calculation using STATA using the parameters for OCD (the weakest
304 predictor, P1=0.16, P2=0.24, pseudo R^2 =0.16) that for an α of 0.05, a sample size of only 264
305 would give a power of 0.8. Finally, the proportion of the variance (pseudo R^2 =0.16, Figure 2)
306 explained by our analysis was relatively low. Of note, the variables that we included in our
307 models were driven by the prior literature. Future research could include feature selection
308 techniques, alternative machine learning models, and consideration of genetic and other
309 biological variables. Validation through additional independent datasets with overlapping
310 variables would also advantageous.

311 **CONCLUSIONS**

312 In summary, we have identified associations between problematic sexual behavior in adults in
313 post-secondary education and the following: compulsive buying, personality disorder, OCD,
314 ADHD, childhood trauma, a family history of PSB, internet addiction, and male gender. Our
315 results not only suggest that measures associated with reward deficiency syndrome are associated
316 with PSB, but also that OCD is a vulnerability factor. Childhood trauma and a family history of

317 domestic violence also showed associations with PSB, implying that emotional regulation and
318 attachment style may be involved in the development of PSB. Our data are consistent with
319 different subgroups within PSB with slightly somewhat contrasting etiological mechanisms by
320 gender.

321 **Acknowledgments**

322 We would like to acknowledge the contribution of Dr. Patrick Carnes to the study design, and
323 volunteer input of Penny Carnes, Grace Li, Hana Graham, and Chanelle Martens to data
324 collection. SJ would like to thank Afia Anjum from DW's lab for her assistance with R. There
325 was no pharmaceutical or industry support for the work reported herein.

326 REFERENCES

- 327 1. Orford J. Hypersexuality: implications for a theory of dependence. *Br J Addict Alcohol*
328 *Other Drugs*. Sep 1978;73(3):299-10. doi:<http://dx.doi.org/10.1111/j.1360-0443.1978.tb00157.x>
- 329 2. Carnes P. *Don't call it love : recovery from sexual addiction*. Bantam Books; 1991:439.
- 330 3. Goodman A. Sexual addiction: terminology and theory. *Psychiatric Times*
331 1998;15(7):22-26.
- 332 4. Carnes P. *Sexual addiction*. 8 ed. vol I. Lippincott, Williams & Wilkins: A Wolters
333 Kluwer Company; 2005.
- 334 5. Carnes PJ, Hopkins TA, Green BA. Clinical relevance of the proposed sexual addiction
335 diagnostic criteria: relation to the Sexual Addiction Screening Test-Revised. *Journal of addiction*
336 *medicine*. Nov-Dec 2014;8(6):450-61. doi:<http://dx.doi.org/10.1097/ADM.0000000000000080>
- 337 6. Rosenthal RJ. Inclusion of pathological gambling in DSM-III, its classification as a
338 disorder of impulse control, and the role of Robert Custer. *International Gambling Studies*.
339 2020;20(1):151-170. doi:<https://doi.org/10.1080/14459795.2019.1638432>
- 340 7. Rehm J, Probst C, Kraus L, Lev-Ran S. *The addiction concept revisited*. Reframing
341 addiction: Policies, processes and pressures. The ALICE RAP project; 2014:103-117.
- 342 8. Kraus SW, Krueger RB, Briken P, et al. Compulsive sexual behaviour disorder in the
343 ICD-11. *World Psychiatry*. Feb 2018;17(1):109-110. doi:<https://doi.org/10.1002/wps.20499>
- 344 9. Griffiths MD. Compulsive sexual behaviour as a behavioural addiction: The impact of
345 the internet and other issues. *Addiction*. 2016;111(12):2107-2108.
346 doi:<http://dx.doi.org/10.1111/add.13315>
- 347 10. Joannides P. The challenging landscape of problematic sexual behaviors, including
348 “sexual addiction” and “hypersexuality.” *New directions in sex therapy: Innovations and*
349 *alternatives*. 2012:69-83. doi:<https://doi.org/10.4324/9780203122556>
- 350 11. Dickenson JA, Gleason N, Coleman E, Miner MH. Prevalence of Distress Associated
351 With Difficulty Controlling Sexual Urges, Feelings, and Behaviors in the United States. *JAMA*
352 *Netw Open*. Nov 2 2018;1(7):e184468. doi:<https://doi.org/10.1001/jamanetworkopen.2018.4468>
- 353 12. Carnes PJ, Green BA, Merlo LJ, Polles A, Carnes S, Gold MS. PATHOS: a brief
354 screening application for assessing sexual addiction. *J Addict Med*. Mar 2012;6(1):29-34.
355 doi:<https://doi.org/10.1097/ADM.0b013e3182251a28>
- 356 13. Carnes PJ. *Facing the Shadow: Starting Sexual and Relationship Recovery: a Gentle*
357 *Path to Beginning Recovery from Sex Addiction*. Gentle Path Press; 2015.
- 358 14. Klein V, Rettenberger M, Briken P. Self-reported indicators of hypersexuality and its
359 correlates in a female online sample. *J Sex Med*. Aug 2014;11(8):1974-81.
360 doi:<https://doi.org/10.1111/jsm.12602>
- 361 15. Reid RC, Li DS, Gilliland R, Stein JA, Fong T. Reliability, validity, and psychometric
362 development of the pornography consumption inventory in a sample of hypersexual men. *J Sex*
363 *Marital Ther*. 2011;37(5):359-85. doi:<https://doi.org/10.1080/0092623x.2011.607047>
- 364 16. Giordano AL, Cecil AL. Religious coping, spirituality, and hypersexual behavior among
365 college students. *Sexual Addiction & Compulsivity*. 2014;21(3):225-239.
366 doi:<https://doi.org/10.1080/10720162.2014.936542>
- 367 17. Cashwell C, L. Giordano A, Lewis T, A. Wachtel K, L. Bartley J. Using the PATHOS
368 Questionnaire for Screening Sexual Addiction Among College Students: A Preliminary

- 369 Exploration. *Sexual Addiction & Compulsivity*. 2015; 22:2, 154-166.
370 [doi:https://doi.org/10.1080/10720162.2015.1037481](https://doi.org/10.1080/10720162.2015.1037481)
- 371 18. Deneke E, Knepper C, Green BA, Carnes PJ. Comparative study of three levels of care in
372 a substance use disorder inpatient facility on risk for sexual addiction. *Sexual Addiction &*
373 *Compulsivity*. 2015;22(2):109-125. doi:<https://doi.org/10.1080/10720162.2014.979382>
- 374 19. Carnes PJ. The Sexual Addiction Screening Process. *Clinical Management of Sex*
375 *Addiction*. 2019:21-39.
- 376 20. Blankenship R, Laaser M. Sexual Addiction and ADHD: Is There A Connection? *Sexual*
377 *Addiction & Compulsivity*. 2004;11(1-2):7-20. doi:<https://doi.org/10.1080/10720160490458184>
- 378 21. Ballester-Arnal R, Castro-Calvo J, Giménez-García C, Gil-Juliá B, Gil-Llario M.
379 Psychiatric comorbidity in compulsive sexual behavior disorder (CSBD). *Addictive behaviors*.
380 2020;107:106384. doi:<https://doi.org/10.1016/j.addbeh.2020.106384>
- 381 22. Grant JE, Lust K, Chamberlain SR. Body dysmorphic disorder and its relationship to
382 sexuality, impulsivity, and addiction. *Psychiatry Research*. 2019;273:260-265.
383 doi:<https://doi.org/10.1016/j.psychres.2019.01.036>
- 384 23. London S, Quinn K, Scheidell JD, Frueh BC, Khan MR. Adverse Experiences in
385 Childhood and Sexually Transmitted Infection Risk From Adolescence Into Adulthood. *Sexually*
386 *transmitted diseases*. 2017;44(9):524-532. doi:<https://doi.org/10.1097/OLQ.0000000000000640>
- 387 24. Mick TM, Hollander E. Impulsive-compulsive sexual behavior. *CNS Spectr*.
388 2006;11(12):944-55. doi:<https://doi.org/10.1017/s1092852900015133>
- 389 25. Carnes P, Green B, Carnes S. The same yet different: Refocusing the Sexual Addiction
390 Screening Test (SAST) to reflect orientation and gender. *Sexual Addiction & Compulsivity*.
391 2010;17(1):7-30. doi:<https://doi.org/10.1080/10720161003604087>
- 392 26. Young KS. Internet addiction: The emergence of a new clinical disorder.
393 *Cyberpsychology & behavior*. 1998;1(3):237-244.
- 394 27. Ridgway NM, Kukar-Kinney M, Monroe KB. An expanded conceptualization and a new
395 measure of compulsive buying. *Journal of consumer Research*. 2008;35(4):622-639. doi:
396 <https://doi.org/10.1086/591108>
- 397 28. Merlhiot G, Mondillon L, Vermeulen N, Basu A, Mermillod M. Adaptation and
398 Validation of the Standardized Assessment of Personality–Abbreviated Scale as a Self-
399 Administered Screening Test (SA-SAPAS). *Journal of Psychology and Psychotherapy*.
400 2014;4:6. doi:<http://dx.doi.org/10.4172/2161-0487.1000164>
- 401 29. Kessler RC, Adler L, Ames M, *et al*. The World Health Organization Adult ADHD Self-
402 Report Scale (ASRS): a short screening scale for use in the general population. *Psychol Med*.
403 Feb 2005;35(2):245-56. doi:<https://doi.org/10.1017/s0033291704002892>
- 404 30. Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerström test for
405 nicotine dependence: a revision of the Fagerstrom Tolerance Questionnaire. *British journal of*
406 *addiction*. 1991;86(9):1119-1127.
- 407 31. Rennert L, Denis C, Peer K, Lynch KG, Gelernter J, Kranzler HR. DSM-5 gambling
408 disorder: prevalence and characteristics in a substance use disorder sample. *Experimental and*
409 *clinical psychopharmacology*. 2014;22(1):50-56. doi:<https://doi.org/10.1037/a0034518>
- 410 32. Benedan L, Monti GS. Predicting the Risk of Gambling Activities in Adolescence: A
411 Case Study. Springer; 2019:47-57. doi:https://dx.doi.org/10.1007/978-3-030-51222-4_5
- 412 33. Nwanganga F, Chapple M. *Practical Machine Learning in R*. John Wiley & Sons; 2020.

- 413 34. Chen CY, Yang HCP, Chen CW, Chen TH. Diagnosing and revising logistic regression
414 models: effect on internal solitary wave propagation. *Engineering Computations*.
415 2008;25(2):121-139. doi:<https://dx.doi.org/10.1108/02644400810855940>
- 416 35. Hosmer Jr DW, Lemeshow S, Sturdivant RX. *Applied logistic regression*. vol 398. John
417 Wiley & Sons; 2013.
- 418 36. Momeni A, Pincus M, Libien J. Imputation and Missing Data. *Introduction to Statistical*
419 *Methods in Pathology*. Springer International Publishing; 2018:185-200.
- 420 37. Öncel SY, Gebizlioglu ÖL, Alioglu FA. Risk factors for smoking behavior among
421 university students. *Turkish Journal of Medical Sciences*. 2011;41(6):1071-1080.
422 doi:<https://dx.doi.org/10.3906/sag-1009-1122>
- 423 38. White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple
424 imputation of incomplete categorical variables. *Comput Stat Data Anal*. 2010;54(10):2267-2275.
425 doi:<https://doi.org/10.1016/j.csda.2010.04.005>
- 426 39. Shrestha N. Detecting Multicollinearity in Regression Analysis. *American Journal of*
427 *Applied Mathematics and Statistics*. 2020/06/16 2020;8(2):39-42.
428 doi:<http://dx.doi.org/10.12691/ajams-8-2-1>
- 429 40. Kim JH. Multicollinearity and misleading statistical results. *Korean J Anesthesiol*.
430 2019;72(6):558-569. doi:<https://doi.org/10.4097/kja.19087>

431 **Figure 1. Correlation matrix for problematic sexual behavior (PSB) with demographic and**
 432 **clinical variables**

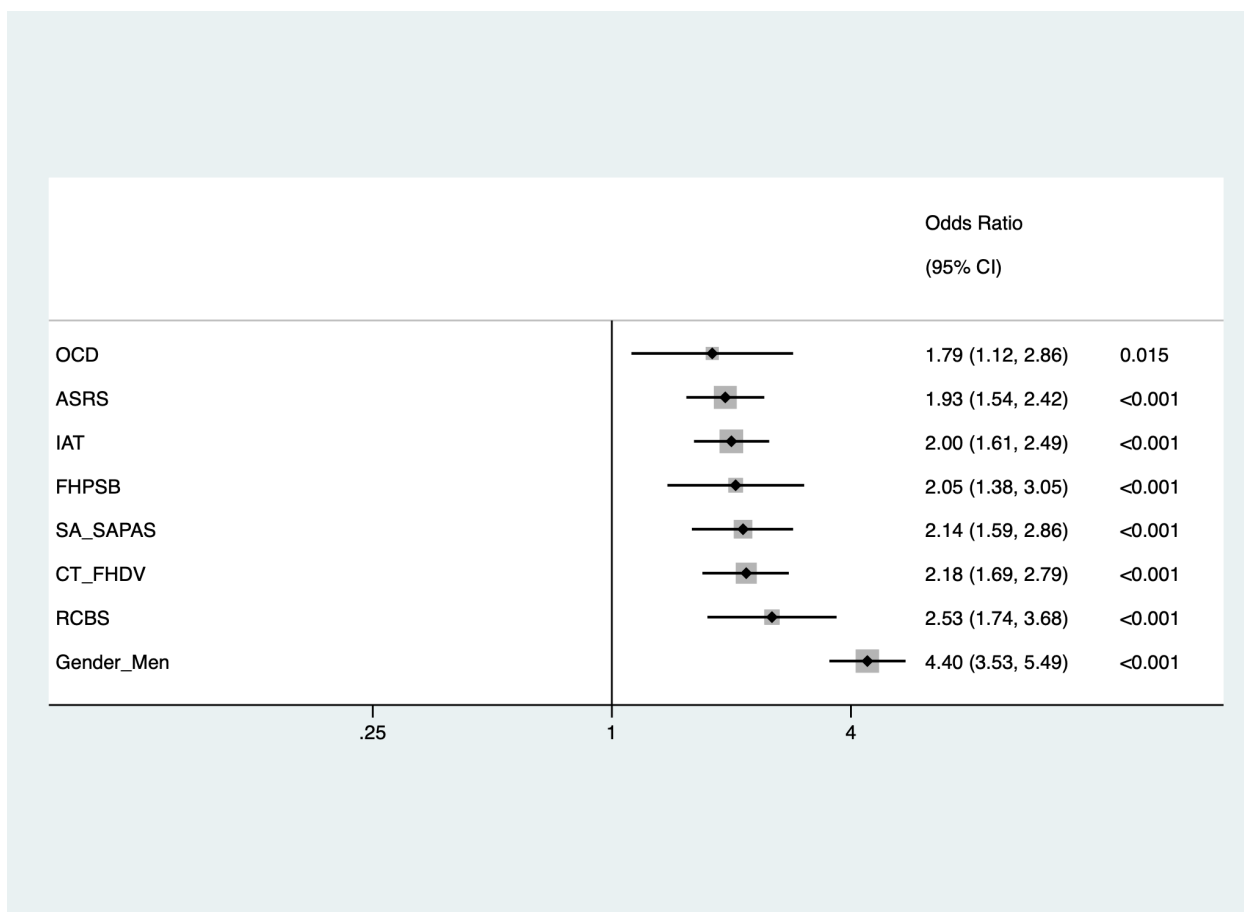


433
 434 *P* values are provided in Supplemental Digital Content, Table 3.
 435 MH_condition: mental health condition; CT_FHDV: childhood trauma and a family history of
 436 domestic violence (including physical, verbal/emotional and sexual childhood trauma and SAST-
 437 R Core question 1); RCBS: Richmond Compulsive Buying Scale; OCD: obsessive-compulsive
 438 disorder (including self-reported and MINI-diagnosis); SA-SAPAS: Standardized Assessment of
 439 Personality–Abbreviated Scale as a Self-Administered Screening Test; FHPSB: a family history
 440 of problematic sexual behavior (including a family history of sexual addiction and SAST-R Core
 441 question 2); ASRS: the Adult ADHD Self-Report Scale (v1.1); IAT: the Internet Addiction Test

442
 443

444 **Figure 2. Forest plots of problematic sexual behavior (PSB) logistic regression analysis**

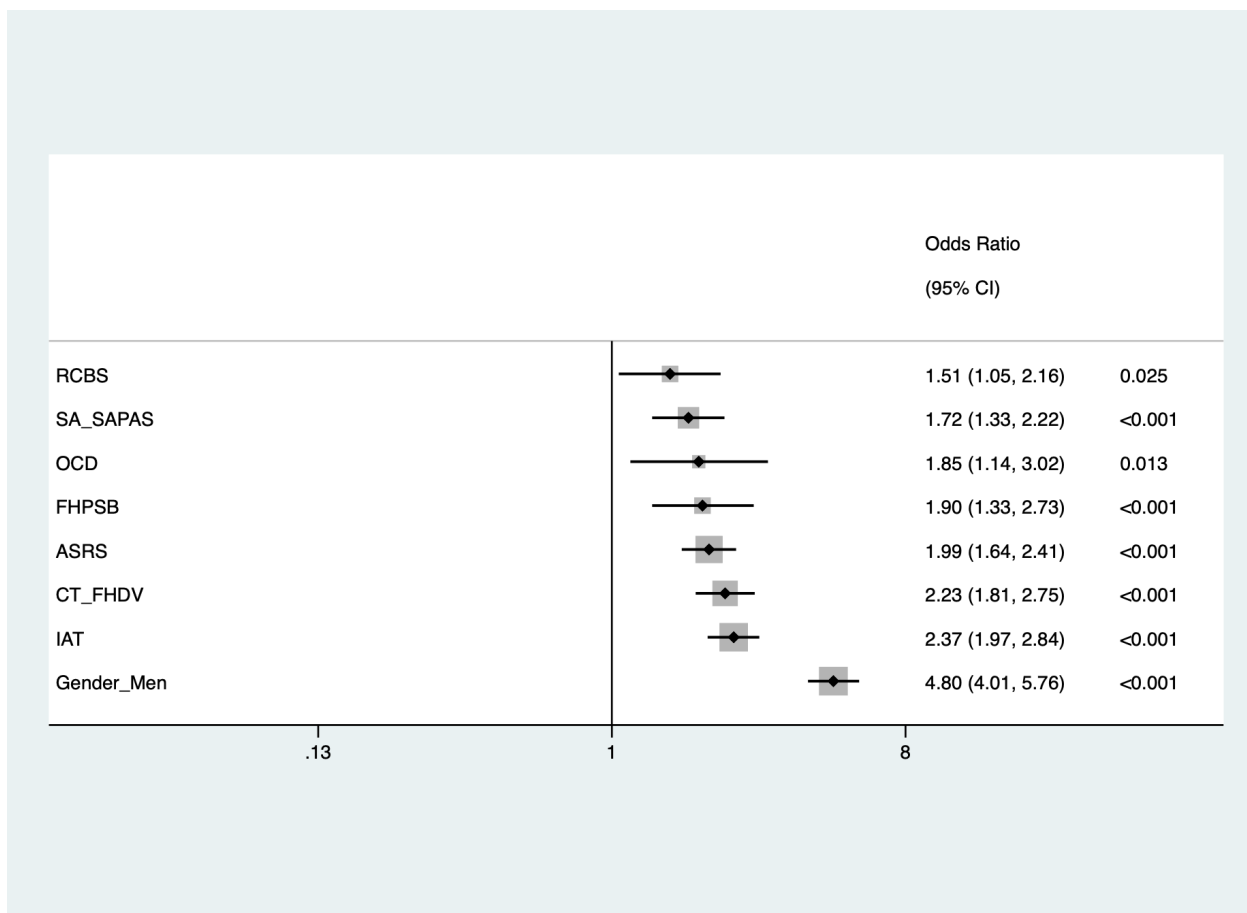
445 **A. By conventional logistic regression**



446

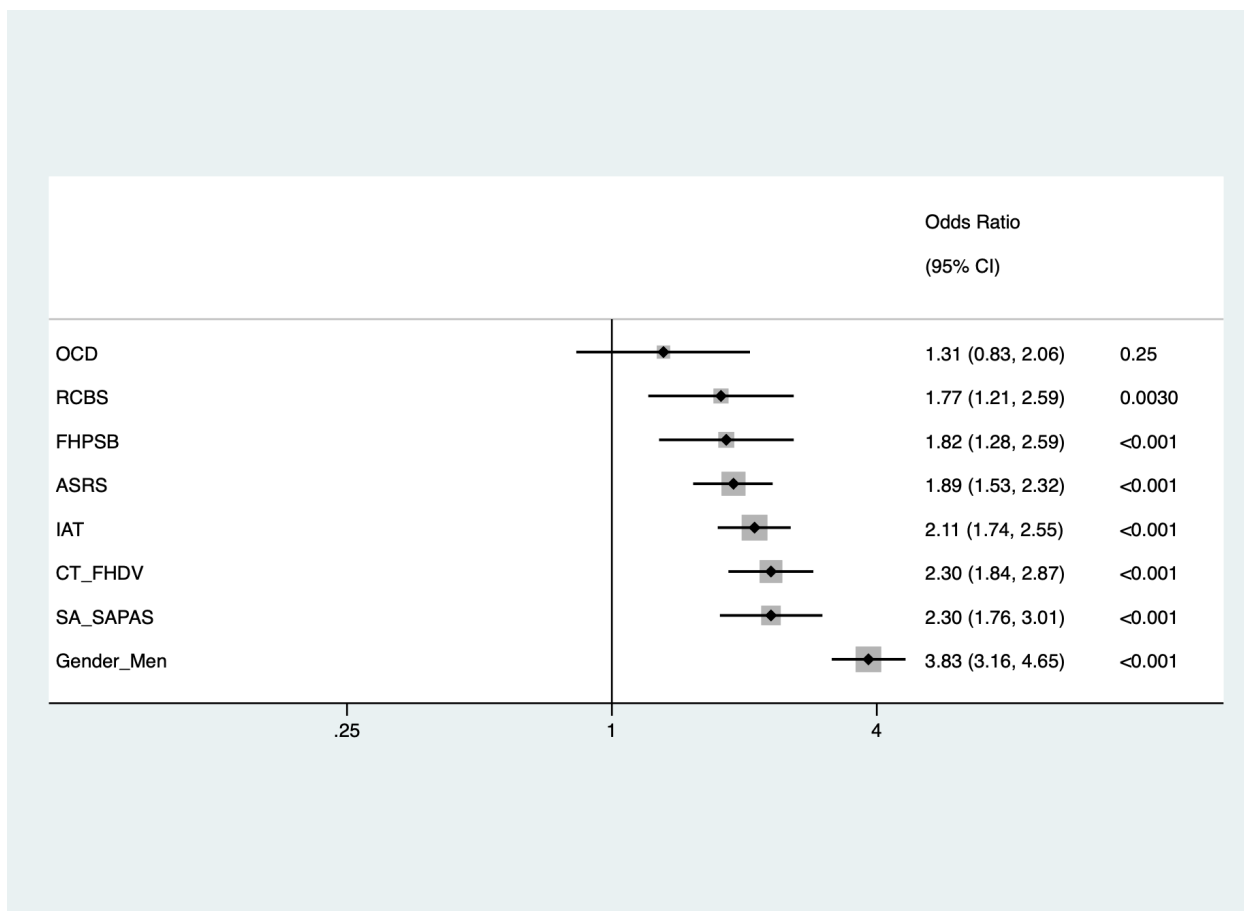
447

448 **B. By machine learning (ML) logistic regression**



449
450

451 **C. Validation ML logistic regression model**



452

453

454 A. The log likelihood of the model was -1149.98 (N=3049, after removing outliers, with a
 455 pseudo $R^2=0.16$, $P<0.00001$). The model passed the linktest with a significant $P(\text{hat})$
 456 ($P<0.001$) and a non-significant $P(\text{hatsq})$ ($P=0.12$). The probability of the Hosmer-
 457 Lemeshow χ^2 test was insignificant ($P=0.64$), suggesting a good model fit. The mean
 458 variance inflation factor (VIF) and the condition number were 1.08 and 6.95, respectively.
 459 The model accuracy, area under the receiver operating (ROC) curve, sensitivity, specificity,
 460 and F score were 84.32%, 77.37%, 14.46%, 98.19% and 0.23, respectively.

461 B. The machine learning logistic regression was conducted on the dataset without removing
 462 the outliers (training, N=2477; testing, N=619; train to test ratio: 0.80). The model

463 accuracy, area under the ROC curve, sensitivity, specificity and F score was

464 74.64%.73.90%, 53.55%, 79.76% and 0.45, respectively.

465 C. The validation model was trained by adjusted PSB, N=3141, and was tested by monthly

466 PSB, N=3145. The model accuracy, area under the ROC curve, sensitivity, specificity and F

467 score was 75.04%.74.26%, 59.11%, 76.97% and 0.34, respectively.

468

469 OCD: obsessive-compulsive disorder (including self-reported and MINI-diagnosis); ASRS: the

470 Adult ADHD Self-Report Scale (v1.1); IAT: the Internet Addiction Test; FHPSB: a family

471 history of problematic sexual behavior (including a family history of sex addiction and SAST-R

472 Core question 2); SA-SAPAS: Standardized Assessment of Personality–Abbreviated Scale as a

473 Self-Administered Screening Test; CT_FHDV: childhood trauma and a family history of

474 domestic violence (including physical, verbal/emotional and sexual childhood trauma and SAST-

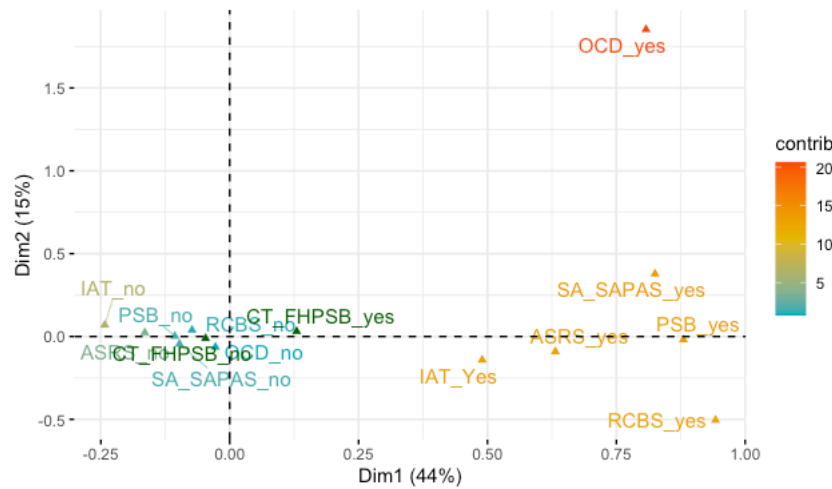
475 R Core question 1); RCBS: Richmond Compulsive Buying Scale

476

477

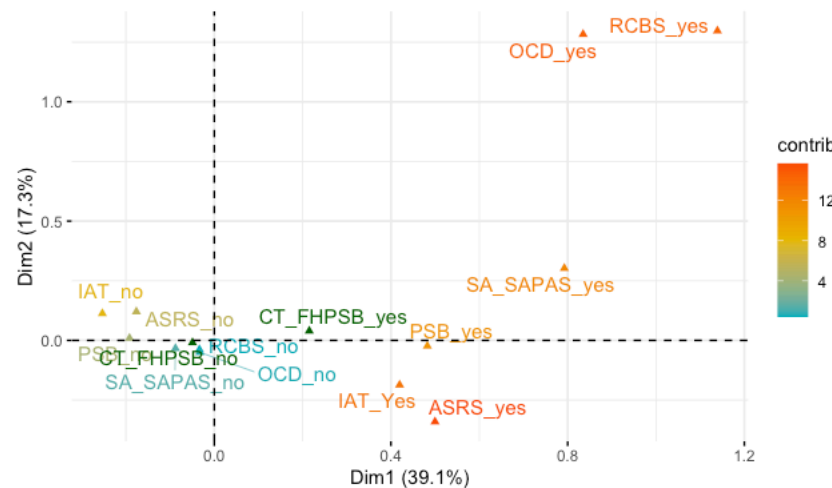
478 **Figure 3. Results of multiple correspondence analysis (dimensions 1 and 2) in women (A)**
479 **and men (B)**

480 **A. Women**



481

482 **B. Men**



483

484 The maps of data binarized into two categories (yes/no, for each variable) in dimensions 1 and 2

485 in women and men are shown. OCD: obsessive-compulsive disorder (including self-reported and

486 MINI-diagnosis); SA-SAPAS: Standardized Assessment of Personality–Abbreviated Scale as a

487 Self-Administered Screening Test; ASRS: the full Adult ADHD Self-Report Scale (v1.1); IAT:

488 the Internet Addiction Test; RCBS: Richmond Compulsive Buying Scale; CT_FHPSB:

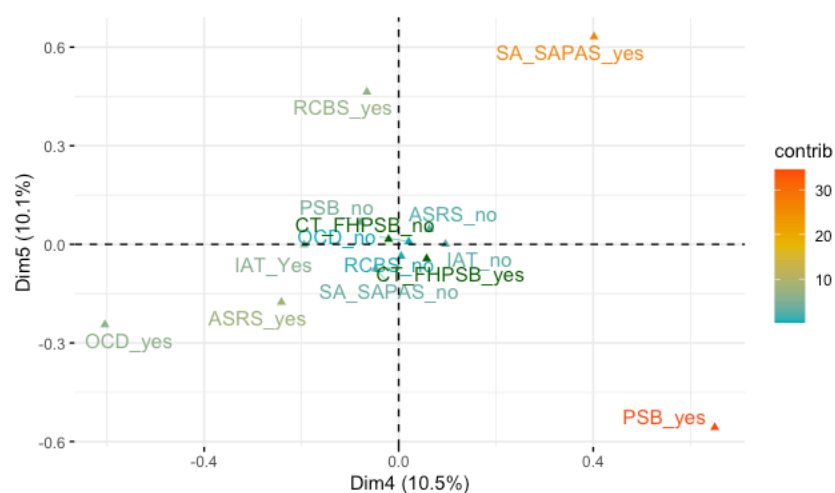
489 childhood trauma and a family history of problematic sexual behavior

490

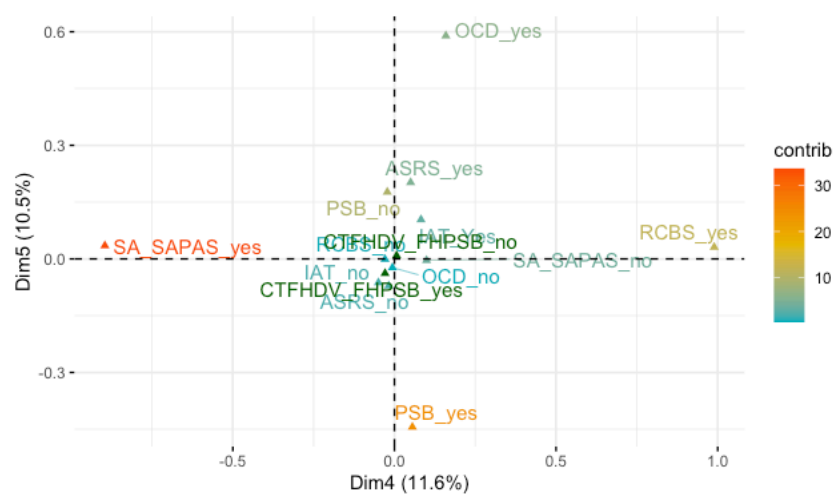
491

492 **Figure 4. Results of multiple correspondence analysis (dimensions 4 and 5) in women (A)**
493 **and men (B)**

494 **A. Women**



B. Men



498 The maps of data binarized into two categories (yes/no, for each variable) in dimensions 4 and 5
499 in women and men are shown. OCD: obsessive-compulsive disorder (including self-reported and
500 MINI-diagnosis); SA-SAPAS: Standardized Assessment of Personality–Abbreviated Scale as a
501 Self-Administered Screening Test; ASRS: the full Adult ADHD Self-Report Scale (v1.1); IAT:
502 the Internet Addiction Test; RCBS: Richmond Compulsive Buying Scale; CT_FHPSB:
503 childhood trauma and a family history of problematic sexual behavior
504

505 **Table 1. Distribution of demographic and clinical variables by problematic sexual behavior**
 506 **(PSB) group**

| Categories | | PSB “yes” 532/3219 (16.53%) | PSB “no” 2687/3219 (83.47%) | Total number | P value |
|--------------------|------------------|-----------------------------------|-----------------------------------|-----------------|------------------------------|
| SAST-R Core score | | 6.58±3.22 (N=532) | 1.38±1.56 (N=2687) | 3219 | <i>P</i> <0.001 ^a |
| Age | | 22.32±5.13 (N=532) | 22.34±5.05 (N=2686) | 3218 | <i>P</i> =0.81 ^a |
| Gender | Women | 232 (10.82%) | 1913 (89.18%) | 2145 | <i>P</i> <0.001 ^b |
| | Men | 290 (28.27%) | 736 (71.73%) | 1026 | |
| | Total | 522 (16.46%) | 2649 (83.54%) | 3171 | |
| Sexual Orientation | Heterosexual | 420 (15.85%) | 2230 (84.15%) | 2650 | <i>P</i> =0.034 ^b |
| | Non-heterosexual | 110 (19.50%) | 454 (80.50%) | 564 | |
| | Total | 530 (16.49%) | 2684 (83.51%) | 3214 | |
| OCD | No | 490 (15.88%) | 2596 (84.12%) | 3086 | <i>P</i> <0.001 ^b |
| | Yes | 40 (33.33%) | 80 (66.67%) | 120 | |
| | Total | 530 (16.53%) | 2676 (83.47%) | 3206 | |
| RCBS | No | 471 (15.54%) | 2559 (84.46%) | 3030 | <i>P</i> <0.001 ^b |
| | Yes | 61 (32.28%) | 128 (67.72%) | 189 | |
| | Total | 532 (16.53%) | 2687 (83.47%) | 3219 | |

| | | | | | |
|-----------------------------------|---------------------------------------|--------------|---------------|------|------------------------------|
| FHPSB | No | 453 (15.38%) | 2492 (84.62%) | 2945 | <i>P</i> <0.001 ^b |
| | Yes | 71 (34.63%) | 134 (65.37%) | 205 | |
| | Total | 524 (16.63%) | 2626 (83.37%) | 3150 | |
| SA-SAPAS | No | 425 (14.76%) | 2454 (85.24%) | 2879 | <i>P</i> <0.001 ^b |
| | Yes | 107 (31.47%) | 233 (68.53%) | 340 | |
| | Total | 532 (16.53%) | 2687 (83.47%) | 3219 | |
| Mental health condition(s) | No | 399 (15.35%) | 2200 (84.65%) | 2599 | <i>P</i> <0.001 ^b |
| | Yes | 131 (21.87%) | 468 (78.13%) | 599 | |
| | Total | 530 (16.57%) | 2668 (83.43%) | 3198 | |
| Childhood trauma | No | 339 (13.94%) | 2092 (86.06%) | 2431 | <i>P</i> <0.001 ^b |
| | Yes | 185 (25.45%) | 542 (74.55%) | 727 | |
| | Total | 524 (16.59%) | 2634 (83.41%) | 3158 | |
| ASRS | No | 321 (12.97%) | 2153 (87.03%) | 2474 | <i>P</i> <0.001 ^b |
| | Yes | 211 (28.32%) | 534 (71.68%) | 745 | |
| | Total | 532 (16.53%) | 2687 (83.47%) | 3219 | |
| IAT | Normal (No) | 241 (11.50%) | 1855 (88.50%) | 2096 | <i>P</i> <0.001 ^b |
| | Mild, moderate or severe (Yes) | 291 (25.91%) | 832 (74.09%) | 1123 | |
| | Total | 532 (16.53%) | 2687 (83.47%) | 3219 | |

507

508 ^aMann-Whitney U test

509 ^bPearson's χ^2 test

510

511 OCD: obsessive-compulsive disorder (including self-reported and MINI-diagnosis); RCBS:

512 Richmond Compulsive Buying Scale; FHPSB: a family history of problematic sexual behavior

513 (including a family history of sex addiction and SAST Core question 2); SA-SAPAS:

514 Standardized Assessment of Personality–Abbreviated Scale as a Self-Administered Screening

515 Test; Childhood trauma: childhood trauma (including physical, verbal/emotional and sexual
516 childhood trauma and SAST Core question 1) and a family history of domestic violence; ASRS:
517 the Adult ADHD Self-Report Scale (v1.1); IAT: the Internet Addiction Test