

1 categorical (discrete), as opposed to continuous. The dependent variable of interest in the
2 current study, namely physical activity level, was a binary variable which took on the values 1
3 for subjects belonging to the light/intensive activity group and 0 for subjects belonging to the
4 more sedentary group. A natural choice for modelling binary data as a function of a set of
5 covariates is the logistic regression model.

6 A second aspect was the challenge of how to appropriately select relevant variables to
7 answer the scientific question of interest while accounting for potential confounders. In
8 randomized experiments to investigate treatment effects, confounding is usually controlled by
9 design (4,5) as the randomization of subjects to either treatment or control does not depend on
10 any of the subject-specific characteristics, leading to an approximate balance in both observed
11 and unobserved covariates for both groups. Observational studies face a unique challenge in
12 estimating treatment effects in light of potential confounding, as distributions of observed and
13 unobserved variables are likely to be different among the treated and untreated individuals. It
14 is therefore of utmost importance to adjust for potentially confounding variables for valid
15 estimation of treatment effects (5). Failure to account for covariates that are associated with
16 the outcome variable or main predictor variable (or both) can result in biased parameter
17 estimates, ultimately causing incorrect inference (omitted variable bias; 6-9). In complex
18 survey data, there is usually an abundance of measured variables, leading to the challenge of
19 identifying important predictors to include in the model.

20 As was already mentioned in the descriptive analysis, initial variable selection was
21 based on the hypothesized predictive value of covariates for either the main outcome variable
22 (physical activity) or the main predictor of interest (cannabis use). Our approach to the
23 subsequent building of the final model was inspired by the general method of ‘double
24 selection’ (also referred to as post- double selection; 10,11). In this general framework,
25 important predictors are identified in three steps: in a first step, the main predictor variable of

1 interest is regressed on the set of candidate variables, and covariates are subsequently
2 identified for inclusion based on a selection procedure (e.g. stepwise selection, LASSO
3 regression, ...). In the second step, the same model is fit to the outcome variable, after which
4 the selected variables are once again retained. The final step involves fitting a regression
5 model to the outcome variable that contains as predictors the union of selected predictors from
6 the first two steps in addition to the main predictor variable of interest. Intuitively, this
7 approach leads to a parsimonious model that minimizes bias that could occur from wrongly
8 excluding important predictors (6). The procedure gives candidate variables ‘two chances’ so
9 to speak to be included in the model, while also allowing unimportant predictors to be
10 dropped. In the current study, we sought to keep model building to a minimum, hence instead
11 of applying the framework of double selection to the entirety of the HIS dataset to identify
12 potential covariates, we used a simplified version on the subset resulting from the initial
13 selection based on the literature.

14 As discussed in the previous section, observational studies pose an additional
15 challenge for causal inference about a specific treatment effect as the researcher has no
16 control over the allocation of treatment in the study population. In randomized controlled
17 studies, this is usually not a problem as the treatment allocation (in this case cannabis use)
18 does not depend on prior characteristics of the subjects by design. In theory, the only
19 difference between the treated and untreated would be the actual treatment they received,
20 making them comparable. In observational studies, this is usually not the case, as treated and
21 untreated individuals are likely to differ in terms of their baseline characteristics. It is evident
22 that comparison of treated and untreated individuals in the setting of observational studies is
23 only meaningful when both groups have a large enough overlap in their baseline covariates
24 (5,7). While adjusting for covariates in regression can help overcome confounding, it is hard
25 to evaluate if the treated and untreated groups have sufficient overlap in terms of their

1 baseline covariate distributions (especially if the number of important covariates is large as it
2 is nearly impossible to investigate differences in multivariate distributions for both groups).

3 One approach that allows researchers to evaluate overlap between treated and untreated
4 individuals is the use of propensity scores (5,7).

5 Propensity scores, as the name suggests, are the underlying probabilities (i.e.
6 propensities) of individual participants to belong to the treatment group (8,9). In the current
7 study, the propensity score was the unobserved probability of answering ‘yes’ to the question
8 of past-month cannabis use in the HIS dataset. The propensity score can be thought of a
9 single-value summary of an individual’s baseline covariates that can subsequently be used in
10 a number of adjustment techniques (e.g. matching, weighting, regression adjustment; (12,15-
11 18). One advantage of inspecting the distribution of estimated propensity scores per user
12 group is that it allows for efficient judgements of whether the groups have a sufficient overlap
13 in baseline covariates to allow for meaningful comparison. Additionally, Vansteelandt &
14 Daniel (10) showed that propensity adjustment in logistic regression makes the test of the
15 causal null-hypothesis more robust against misspecification of the outcome regression model,
16 by still delivering valid estimates, so long as the propensity score model is correctly specified.
17 Given these attractive properties of propensity scores, the current study also included
18 estimated propensities to adjust for confounding in the final regression model.

19 In addition to the scale of the outcome variable (categorical) and the issue of variable
20 selection and confounding, the complex sampling design of the HIS dataset also poses unique
21 challenges for data-analysis. Both stratification and multistage sampling are important to take
22 into account when analyzing survey data (2,11,12). The HIS dataset contained 12 unique
23 strata with predetermined sample sizes (5 provinces in Flanders, 5 provinces in Wallonia + 1
24 stratum corresponding to East Belgium and finally the region Brussels as the last stratum; 22).
25 In addition to information on strata, the HIS dataset also provided sampling weights for every

1 individual in the sample, with the weights denoting how many individuals in the target
2 population each respondent in the sample represents (the precise formula for the weights is
3 described in 22).

4 The nested cluster structure in the data obviously violated one of the key assumptions
5 in classical regression analysis, namely the assumption of independently distributed
6 responses. To account for correlated responses, two main approaches are usually identified,
7 namely a model-based approach and a design-based approach (13,14). The model-based
8 approach involves explicitly modelling the correlation structure between the responses by
9 specifying random effects, leading to so-called generalized linear mixed effects models. The
10 second set of methods, design-based methods, usually make adjustments to the variance
11 estimators of parameter estimates so that clustering information is taken into account. One
12 well known instance of design-based approaches is the sandwich estimator (which allows for
13 variance estimation that is robust against misspecification of the covariance structure, given a
14 correctly specified mean structure; 25). Another commonly used technique (and the one used
15 in the current study) is variance estimation based on a Taylor Series (TS) linearization
16 (discussed later in this paragraph; 24,26). Our motivation for using the TS approach over
17 multilevel models was based on both computational, and theoretical considerations. Firstly,
18 given that the interest of the current study is in the effect of level 1 covariates (i.e.
19 characteristics of the individual subjects) rather than higher-level characteristics or the
20 specific covariance structure of the data, there would be no added value in using multilevel
21 models (13). Secondly, several studies pointed out that -especially in large sample sizes- both
22 the model-based and design-based methods typically yield qualitatively similar results (15).
23 Finally, the use of TS approximation of estimator variance is well documented and easily
24 implemented in standard software (e.g., the surveylogistic procedure in SAS, 28).

25 As opposed to standard linear regression, where a closed form solution to the normal

1 equations exists to find the OLS estimates of regression parameters, the estimators of
2 regression parameters in a logistic regression model are non-linear, resulting in the need to
3 use iterative algorithms to find maximum likelihood estimates. In order to estimate the
4 variance of these non-linear estimators, a first-order Taylor expansion of the non-linear
5 estimator is calculated around the population parameter to yield a linear approximation of the
6 estimator that can subsequently be used in variance estimation. Specifically for the current
7 study, the TS linearization also incorporates important aspects of the sampling design when
8 approximating the non-linear estimator (e.g. clustering, strata and weights).

9 Taking together all of the aforementioned aspects of the HIS dataset and keeping in
10 mind our research question, the model that seemed most appropriate in this case is a survey-
11 weighted logistic regression model, where we aimed to adjust for confounding and try to
12 prevent extrapolation beyond the data at hand by including a set of baseline covariates into the
13 model, in addition to the estimated propensity scores. The (limited) model building followed
14 three phases, inspired by the double selection literature (16,6): first, the treatment variable
15 (cannabis use) was regressed on the set of covariates and type III tests were used to identify
16 important predictors at the .05 significance level. Second, the same approach was applied to
17 the outcome variable, by regressing physical activity on the same set of covariates in addition
18 to cannabis use. Finally, a final model was fit to the outcome variable with cannabis use and
19 the estimated propensity score as covariates. Both models (i.e. the propensity model and the
20 outcome model) were fit using the surveylogistic procedure in SAS (17), that uses sampling
21 weights, information about strata and clustering to estimate the regression parameters and
22 associated variance. Weights were used in the estimation of regression parameters by
23 including them into the formula for the pseudolikelihood and subsequently during the iterative
24 estimation of parameters, as well as in the Taylor linearization approach to estimating the
25 variance.

1 *Variables*

2 The initial set of covariates consisted of the following variables: Cannabis, Physical,
3 Education, Income, Age, Gender, Depression, Anxiety, Urbanization and the Global Activity
4 Limitations Indicator (GALI). Additionally, the variables Year and Province were included in
5 the models as fixed effects to account for potential heterogeneity in responses of different
6 years and geographical locations.

7 The binary variable ‘Physical activity’ was created from the original item ‘Past 12
8 months leisure time physical activity’ asking participants the question “What describes best
9 your leisure time activities in the past 12 months?”. The original question allowed for the
10 following answers: 1) ‘Hard training and competitive sport more than once a week’, 2)
11 ‘Jogging and other recreational sports or gardening, at least 4 hours per week’, 3) ‘Jogging
12 and other recreational sports or gardening, at the most 4 hours per week’, 4) ‘Walking,
13 bicycling or other light activities at least 4 hours a week’, 5) ‘Walking, bicycling or other light
14 activities at the most 4 hours a week’, and 6) ‘Reading, watching TV or other sedentary
15 activities’. From this variable, responses 1 and 2 were recoded as ‘Intensive training / sport \geq
16 4 hours per week’, responses 3 to 5 were recoded as ‘Sport < 4 hours / light activities’, and
17 response 6 was recoded as ‘Sedentary activities’. Finally, the former two categories were
18 recoded as ‘light/intensive physical activities’, resulting in the binary variable ‘Physical
19 activity’ that was used in subsequent analyses.

20 Education was not recorded at the individual level, but refers to the highest diploma
21 within the household of an individual. Similarly, income refers to quintiles of household
22 equivalized income. Furthermore, GALI refers to the presence or absence of long-term
23 functional limitations in typical activities in daily life (18). Depression was included as a
24 binary variable reflecting the presence of ‘any depression’. This variable was constructed
25 from two other variables in the HIS (Major Depressive Syndrome and Other Depression, both

1 were based on the Patient Health Questionnaire PHQ-9 (19). The other measure of mental
 2 health, anxiety, was also included as a binary variable. This variable assessed the presence of
 3 a Generalized Anxiety Disorder (GAD) based on the GAD-7 questionnaire (20). Finally,
 4 Urbanization refers to the level of morphologic and functional urbanization, and consisted of
 5 the categories 1) 'Big cities and dense agglomerations', 2) 'Suburban/banlieus', 3) 'Urbanized
 6 municipalities', and 4) 'Rural'.

7 *Model Building*

8 In a first step, we used the rationale of double variable selection to potentially refine
 9 our set of candidate predictors. To this end we first fit a survey-weighted logistic regression
 10 model to the 'treatment' variable, cannabis use, and retained the significant predictors. The
 11 model was defined as follows:

$$\log\left(\frac{P(\text{Cannabis}_i = 1|\mathbf{X}_i)}{1 - P(\text{Cannabis}_i = 1|\mathbf{X}_i)}\right) = \beta_1\text{Year}_i + \beta_2\text{Education}_i + \beta_3\text{Income}_i + \beta_4\text{Age}_i + \beta_5\text{Gender}_i \\ + \beta_6\text{Depression}_i + \beta_7\text{Anxiety}_i + \beta_8\text{Urbanization}_i + \beta_9\text{GALL}_i \\ + \beta_{10}\text{Province}_i$$

With

$P(\text{Cannabis}_i = 1|\mathbf{X}_i)$ The conditional probability of past-month cannabis use of subject i
 $\beta_1, \dots, \beta_{10}$ The regression coefficients

(1)

12 The model assumed that individual participant's responses were i.i.d Bernoulli random
 13 variables with conditional mean μ_i which is the 'risk' of observing past-month cannabis use,
 14 conditional on the set of specific covariate values. We assumed the logarithm of the odds of
 15 past-month cannabis use depended linearly on the covariates. To account for the survey
 16 design, information about which of the 12 strata each subject belonged to was also specified,
 17 together with clustering at the household level, and sampling weights at the individual level.
 18 We used households as the clustering variable, as information on the primary sampling units

1 (i.e. municipalities) was not available. We assumed that the use of households already
 2 captured most of the dependencies between responses, as it is the lowest form of clustering
 3 present in the study, which usually has the strongest intra-class correlation. To obtain the
 4 maximum likelihood estimates of the free model parameters, the Fisher scoring (iteratively
 5 reweighted least squares) algorithm was used to optimize the pseudo-likelihood function.
 6 Estimation of the variance of parameter estimates proceeded via TS linearization.

7 Inspection of Type III tests for the estimated effects revealed that all variables except
 8 for Education, Income, Depression and GALI significantly predicted past month cannabis use
 9 at the $\alpha = .05$ significance level (see Table S1.1 and see Table S1.2 for estimated odds ratios
 10 and confidence intervals)

Table S1.1

Results of the cannabis regression model (Type III tests)

Variable	F-value	Numerator Df	Denominator Df	P-value
Year	4.93	4	13809	<0.001 ***
Education	1.48	4	13809	0.207
Income	1.95	4	13809	0.099
Age	273.77	1	13812	<0.001 ***
Gender	103.86	1	13812	<0.001 ***
Depression	2.30	1	13812	0.129
Anxiety	10.17	1	13812	0.001 **
Urbanization	3.71	3	13810	0.011 *
GALI	1.35	1	13812	0.245
Province	3.36	10	13803	<0.001 ***

Note. $p < .05^*$, $p < .01^{**}$, $p < .001^{***}$. GALI = Global Activity Limitations Indicator

Table S1.2*Estimated odd ratios and 95% confidence intervals based on the cannabis regression model*

Variable	Effect	Odds Ratio	95% Confidence Bounds	
			Lower	Upper
Year	2001	1.00		
	2004	1.19	0.80	1.76
	2008	1.60	1.08	2.37
	2013	0.99	0.67	1.48
	2018	1.99	1.37	2.89
Education	Higher Education	1.00		
	Higher Secondary	1.29	0.97	1.72
	Lower Secondary	1.21	0.82	1.79
	No Diploma	0.83	0.46	1.48
	Other	1.86	0.82	4.21
Income	First	1.00		
	Second	1.12	0.74	1.68
	Third	0.68	0.45	0.98
	Fourth	1.03	0.70	1.51
	Fifth	1.01	0.68	1.53
Age	Age	0.91	0.90	0.98
Gender	Woman	1.00		
	Man	3.35	2.66	4.23
Depression	Yes	1.00		
	No	0.73	0.49	1.10
Anxiety	Yes	1.00		
	No	0.48	0.30	0.75
Urbanization	Rural	1.00		
	Suburban	0.86	0.51	1.46
	Urbanized	0.61	0.37	1.00
	Big Cities	1.20	0.75	1.92
GALI	Not Limited	1.00		
	Limited	1.23	0.87	1.74
Province	West-Vlaanderen	1.00		
	Antwerpen	1.62	0.90	2.93
	Waals Brabant	2.58	1.23	5.42
	Brussel	2.67	1.52	4.71
	Hainaut	1.26	0.68	2.35
	Limburg	0.93	0.44	2.00
	Liege	1.84	0.96	3.54
	Luxembourg	1.16	0.58	2.33
	Namur	0.87	0.40	1.91
	Oost-Vlaanderen	2.17	1.17	4.03
Vlaams Brabant	1.77	0.81	3.90	

Note. The degrees of freedom in computing the confidence limits is 13812. Effects for categorical variables represent a comparison with the reference level (first level of every variable). GALI = Global Activity Limitations Indicator

1 Next, a second survey-weighted logistic regression model was fit, but this time using
 2 the variable, physical activity level, as the outcome and including the same covariates as were
 3 used in the treatment regression model.

$$\log\left(\frac{P(\text{Active}_i = 1|\mathbf{X}_i)}{1 - P(\text{Active}_i = 1|\mathbf{X}_i)}\right) = \beta_1\text{Year}_i + \beta_2\text{Education}_i + \beta_3\text{Income}_i + \beta_4\text{Age}_i + \beta_5\text{Gender}_i \\ + \beta_6\text{Depression}_i + \beta_7\text{Anxiety}_i + \beta_8\text{Urbanization}_i + \beta_9\text{GALI}_i \\ + \beta_{10}\text{Province}_i$$

With

$P(\text{Active}_i = 1|\mathbf{X}_i)$ The conditional probability of light/intensive PA of subject i

$\beta_1, \dots, \beta_{10}$ The regression coefficients

(2)

4 Estimation and variance estimation proceeded identically to the previously fit model. Based
 5 on type III tests (Table S1.3, see Table S1.4 for estimated odds ratios and confidence
 6 intervals), all variables significantly predicted membership of the light/intensive physical
 7 activity group at the alpha = .05 significance level, except for cannabis use. Based on the
 8 results of this first step, we concluded that all variables that were identified based on subject-
 9 matter grounds are important predictors of either the physical activity level, cannabis use, or
 10 both, and should therefore be included in the final propensity score model and outcome
 11 regression model.

Table S1.3*First outcome regression model results (Type III tests)*

Variable	F-value	Numerator Df	Denominator Df	P-value
Year	5.26	4	13809	<0.001***
Education	41.19	4	13809	<0.001***
Cannabis	0.04	1	13812	0.836
Income	2.41	4	13809	0.047*
Age	17.95	1	13812	<0.001***
Gender	41.83	1	13812	<0.001***
Depression	35.53	1	13812	<0.001***
Anxiety	4.03	1	13812	0.045*
Urbanization	6.03	3	13810	<0.001***
GALI	93.29	1	13812	<0.001***

Note. $p < .05^*$, $p < .01^{**}$, $p < .001^{***}$. GALI = Global Activity Limitations Indicator

Table S1.4*Estimated odds ratios and 95% confidence intervals based on the first outcome regression model*

Variable	Effect	Odds Ratio	95% Confidence Bounds	
			Lower	Upper
Year	2001	1.00		
	2004	0.75	0.65	0.86
	2008	0.81	0.71	0.94
	2013	0.81	0.69	0.94
	2018	0.90	0.79	1.03
Education	Higher Education	1.00		
	Higher Secondary	1.59	1.42	1.79
	Lower Secondary	1.95	1.70	2.24
	No Diploma	2.84	2.39	3.37
	Other	1.97	1.30	2.99
Income	First	1.00		
	Second	0.91	0.78	1.07
	Third	0.85	0.73	0.99
	Fourth	0.86	0.73	1.00
	Fifth	0.79	0.68	0.92
Age	Age	0.99	0.99	1.00
Cannabis	Yes	1.00		
	No	1.03	0.78	1.36
Gender	Woman	1.00		
	Man	0.76	0.69	0.82
Depression	Yes	1.00		
	No	0.59	0.50	0.70
Anxiety	Yes	1.00		
	No	0.82	0.68	1.00
Urbanization	Rural	1.00		
	Suburban	1.00	0.84	1.19
	Urbanized	0.97	0.82	1.14
	Big Cities	1.26	1.07	1.47
GALI	Not Limited	1.00		
	Limited	1.77	1.57	1.98
Province	West-Vlaanderen	1.00		
	Antwerpen	0.78	0.63	0.96
	Waals Brabant	1.25	0.94	1.66
	Brussel	1.20	0.99	1.46
	Hainaut	1.71	1.42	2.07
	Limburg	1.06	0.83	1.35
	Liege	1.47	1.19	1.81
	Luxembourg	1.33	1.05	1.67
	Namur	1.19	0.94	1.52
	Oost-Vlaanderen	0.93	0.74	1.16
Vlaams Brabant	1.04	0.82	1.33	

Note. The degrees of freedom in computing the confidence limits is 13812. Effects for categorical variables represent a comparison with the reference level (first level of every variable). GALI = Global Activity Limitations Indicator

1 In a second step, we first used the cannabis model (as the relevant predictors were used
 2 already) to estimate individual participants' underlying propensities for past-month cannabis
 3 use to be used in the subsequent outcome regression model. Inspection of the estimated
 4 propensities per cannabis group allowed us to evaluate the degree of overlap between users
 5 and non-users, and to gain better insight in the danger of making extrapolations in the current
 6 study. A density plot (Figure S1.1) reveals three key insights. First, the distributions are
 7 highly left- skewed, confirming that cannabis use was a relatively rare event. Second, as
 8 would be expected, cannabis users had greater estimated propensities for cannabis use on
 9 average, based on their covariates. Third, and most importantly, though the distributions for
 10 users and non-users were clearly distinct, there was a good amount of overlap between both
 11 user levels.

12 With the set of covariates identified, we proceeded by fitting a final survey-weighted
 13 regression model to the outcome variable, with cannabis use as a main predictor in addition to
 14 the propensity score. Once again, clustering at the household level was specified, as well as
 15 information on the different strata and sampling weights. Estimation and variance estimation
 16 were carried out via the Fisher scoring algorithm to optimize the pseudo-likelihood function
 17 and TS linearization once more.

$$\log\left(\frac{P(\text{Active}_i = 1|\mathbf{X}_i)}{1 - P(\text{Active}_i = 1|\mathbf{X}_i)}\right) = \beta_1 \text{Cannabis}_i + \beta_2 \text{Propensity}_i$$

With

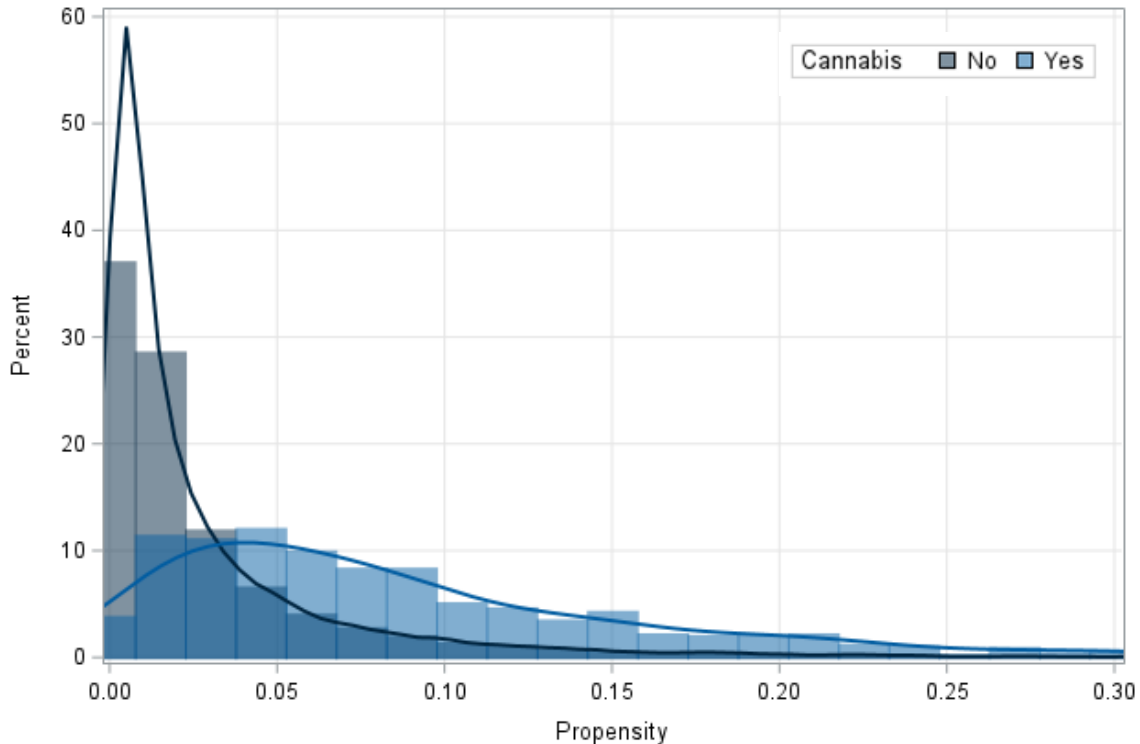
$P(\text{Active}_i = 1|\mathbf{X}_i)$ The conditional probability of light/intensive PA of subject i

β_1, β_2 The regression coefficients

(3)

Figure S1.1

Distribution of estimated propensity scores for users and non-users based on a survey-weighted logistic regression model



Note. Bars represent histograms of estimated propensity scores per group, whereas the smooth distributions represent estimated densities.

1 Propensity matching analysis

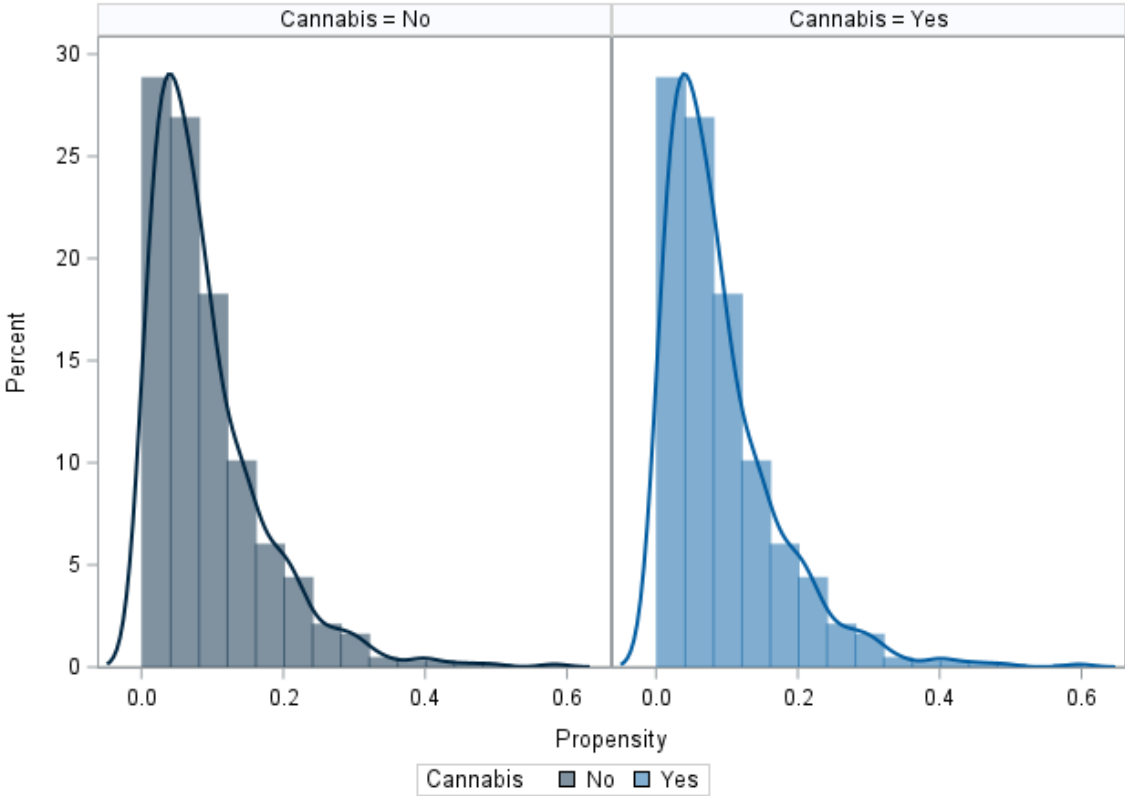
2 One potential drawback of the regression models above, especially with propensity
 3 score and covariate adjustment, is the rather complicated way of interpreting regression
 4 coefficients. The coefficients capture the natural logarithm of the estimated odds ratio of two
 5 levels of a covariate (e.g. user vs non-user in the categorical case). An alternative analysis that
 6 uses estimated propensity scores is propensity score matching. The general goal of this
 7 framework is to emulate the attractive property of balanced baseline covariates between
 8 treated and untreated individuals (usually present in randomized controlled trials), by
 9 matching the treatment groups on their estimated propensity scores. Using propensity scores

1 to match treated to untreated individuals allows researchers to estimate the average treatment
2 effect in the treated, i.e., the observed effect (risk ratio or risk difference) if the treated
3 individuals would have remained untreated. In the current study, this pertains to estimating
4 the difference between the proportion of active individuals in the users and the proportion of
5 active individuals in the users had they not used cannabis. The advantage of the current
6 approach is that the results are easily interpretable and clearly convey the effect of interest.

7 Propensity score matching, i.e., finding the non-user for every user with the most
8 similar estimated propensity for cannabis use, can be carried out in a lot of different ways.
9 Different algorithms have been developed, with each one yielding slightly different results.
10 According to a formal comparison based on simulations, Austin (21) found that nearest-
11 neighbor caliper matching performed the best. In this algorithm, applied in the current study,
12 users were selected one by one in a random fashion, each time defining an acceptable range
13 on the logit of the propensity score of the non-users and selecting a random non-user that
14 meets the criteria (if any). The use of an acceptable range around the estimated logit of each
15 user is necessary as estimated propensities are typically quite unique due to rounding for
16 example. The acceptable range is also commonly referred to as the caliper width as it allows
17 researchers to flexibly place a constraint on how close the estimated logits of the non-users
18 needs to be in order to be considered a valid candidate. The caliper width in this study was
19 determined as 20% of the estimated standard deviation of the users' logit of the propensity
20 score as shown to be optimal for estimating treatment effects in Austin (22). A suitable match
21 could not be found for 4 out of the 618 users in the sample. The resulting propensity
22 distributions for users and matched non-users are shown in Figure S1.2. As is evident from
23 the figure, the distributions were almost identical, suggesting we established a good balance in
24 baseline covariates between users and their matches.

Figure S1.2

Estimated propensities for users and matched non-users based on a survey-weighted logistic regression model



Note. Bars represent histograms of estimated propensity scores per group, whereas the smooth distributions represent estimated densities.

References

1. RStudio Team. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>. 2020.
2. Lumley T. Analysis of Complex Survey Samples. *Journal of Statistical Software*. 2004; 9(8).
3. Saylor J, Friedmann E, Lee HJ. Navigating Complex Sample Analysis Using National Survey Data. *Nursing Research*. 2012; 61(3): 231-237.
4. Jager KJ, Zoccali C, MacLeod AM, Dekker FW. Confounding: what it is and how to deal with it. *Kidney International*. 2008; 73(3): 256-260.
5. Hernán, M.A., Robins, J.M. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC; 2020.
6. Urminsky O, Hansen C, Chernozhukov V. Using Double-Lasso Regression for Principled Variable Selection. *SSRN Electronic Journal*. 2016.
7. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, et al. Results of Multivariable Logistic Regression, Propensity Matching, Propensity Adjustment, and Propensity-based Weighting under Conditions of Nonuniform Effect. *American Journal of Epidemiology*. 2006; 163(3): 262-270.
8. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*. 1984.

9. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70(1): 41-55.
10. Vansteelandt S, Daniel RM. On regression adjustment for the propensity score. *Statistics in Medicine*. 2014; 33(23): 4053-4072.
11. Demarest S, Van der Heyden J, Charafeddine R, Drieskens S, Gisle L, Tafforeau J. Methodological basics and evolution of the Belgian health interview survey 1997–2008. *Archives of Public Health*. 2013; 71(1): 24.
12. Royal K. Survey research methods: A guide for creating post-stratification weights to correct for sample bias. *Education in the Health Professions*. 2019; 2(1): 48.
13. McNeish DM, Haring JR. Clustered data with small sample sizes: Comparing the performance of model-based and design-based approaches. *Communications in Statistics - Simulation and Computation*. 2017; 46(2): 855-869.
14. Huang FL. Alternatives to Multilevel Modeling for the Analysis of Clustered Data. *The Journal of Experimental Education*. 2016; 84(1): 175-196.
15. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: What are the differences? *Statistics in Medicine*. 2009; 28(2): 221-239.
16. Belloni A, Chernozhukov V, Hansen C. Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*. 2014; 81(2): 608-650.
17. SAS Institute Inc. 2016. SAS/STAT® 14.2 User's Guide. Cary, NC: SAS Institute Inc. [Online].

18. Van Oyen H, Van Der Heyden J, Perenboom R, Jagger C. Monitoring Population Disability: Evaluation of a New Global Activity Limitation Indicator (GALI). *Sozial-und Praeventivmedizin*. 2006; 51(3): 153-161.
19. Spitzer R, Kroenke K, Williams J. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA*. 1999 Nov; 282(18): 1737-44.
20. Spitzer R, Kroenke K, Williams J, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006 May; 166(10): 1092-7.
21. Austin PC. A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*. 2014; 33(6): 1057-1069.
22. Austin PC. Some Methods of Propensity-Score Matching had Superior Performance to Others: Results of an Empirical Investigation and Monte Carlo simulations. *Biometrical Journal*. 2009; 51(1): 171-184.