Appendix 1. Data dictionary for 61 variables included in the random forest models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Definition** | **Levels** | **Included in Analysis 1**  **(n = 882)a** | **Included in Analysis 2 (n = 344)a** |
| **Outcome** | | | |  |
| Cesd1\_score | Average CESD score at year 1 | Continuous score (1 to 4) | X |  |
| Diagnosis\_1 | Mental health diagnosis by healthcare professional at year 1 | 0 = No; 1 = Yes |  | X |
| **Socio-demographic** | | | |  |
| age | Age | Continuous (from 15 -29 years) | X | X |
| province | Canadian Province | 7= Ontario, 9= Quebec | X | X |
| curr\_orient | Current sexual orientation | 0= asexual, 1 = heteroflexible / straight or heterosexual, 2 = bisexual, 3 = gay, 4 = lesbian, 5 = pansexual, 6 = queer, 7 = questioning / not sure, 8 = Two-spiritb | X | Xb |
| gender | Current gender identity | 0= cisgender man, 1= cisgender woman, 2 = trans man, 3 = trans woman, 4 = genderfluid / genderqueer / gender non-binary / gender non-confirming / Two-spirit | X | X |
| ethnicity | Ethnicity | 0= White, 1= Latin American, 2 = Middle Eastern, 3 = East & Southeast Asian, 4 = South Asian, 5 = Black (African, Caribbean, North American), 6 = Indigenous (FN, Metis, Inuit), 7 = Mixed | X | X |
| residence | How long living in Canada | 1 = 2 years or less, 2 = 3 to 5 years, 3 = 6 to 10 years, 4 = 11 to 20 years, 5 = Canadian born | X | X |
| education | Highest level of education completed | 0= high school or below, 1= college or diploma or similar, 2= university degree | X | X |
| employ | Employment status | 0= unemployed (including student, working without pay and others), 1= employed (including self-employed); 2 = student only | X | X |
| house\_income | Combined income of everyone in the house for the past year | 0 = Less than $15,000  1 = $15,000 to $29,999  2 = $30,000 to $59,999  3 = $60,000 to $79,999  4 = $80,000 to $100,000  5 = More than $100,000 | X | X |
| ind\_income | Individual income for the past year | 0 = Less than $15,000  1 = $15,000 to $29,999  2 = $30,000 to $59,999  3 = $60,000 to $79,999  4 = $80,000 to $100,000  5 = More than $100,000 | X | X |
| Rural city | What best describes the environment you live in? | 1 = Large urban center (100,000+)  2 = Medium city / town (30,000 – 99,000)  3 = Small city / town  (1,000 – 29,999)  4 = Rural area ( < 1,000) | X | X |
| where\_live | Current living situation or accommodation | 0= not in own place (including shelter, supportive housing or homeless), 1= own place or rent own place with roommate or partner, 2 = live with parents / guardian | X | X |
| **General health** | | | |  |
| gen\_health | Health status ratings | 1 – 5 (poor to excellent) | X | X |
| fitness1 | Past 7 day, sports or physical activities for 10 continuous mins. | 0= no, 1= yes | X | X |
| treat\_comorbidities | Diagnosed with comorbid conditions: chronic lung disease, asthma, chronic pain, diabetes, heart disease, lung cancer, other cancers | 0= no, 1= yes. (Note: nobody has cancer, so we did not use it) c | X | X |
| Treatment\_lung | Diagnosed or treated with chronic lung disease (e.g. emphysema, chronic bronchitis) | 0= no, 1= yes. | X | -c |
| Treatment\_asthma | Diagnosed or treated with asthma | 0= no, 1= yes. | X | -c |
| Treatment\_chronic | Diagnosed or treated with chronic pain | 0= no, 1= yes. | X | -c |
| Treatment\_diabetes | Diagnosed or treated with diabetes | 0= no, 1= yes. | X | -c |
| Treatment\_heart | Diagnosed or treated with heart disease | 0= no, 1= yes. | X | -c |
| disability | Living with condition: intellectual disability, learning disability, Autism, Asperger’s or neurodiverse spectrum, difficulty in seeing, hearing, walking, using hands or fingers, doing physical activities, chronic pain | 0= no, 1= yes | X | X |
| Disability\_\_\_1 | Living with intellectual disability | 0= no, 1= yes | X | -c |
| Disability\_\_\_2 | Living with learning disability | 0= no, 1= yes | X | -c |
| Disability\_\_\_3 | Living with Autism, Asperger’s or neurodiverse spectrum | 0= no, 1= yes | X | -c |
| Disability\_\_\_4 | Living with difficulty seeing (blind, low vision or visual impairment) | 0= no, 1= yes | X | -c |
| Disability\_\_\_5 | Living with hearing difficulty | 0= no, 1= yes | X | -c |
| Disability\_\_\_6 | Living with difficulty walking, using stairs, using hands or fingers, or doing other physical activities | 0= no, 1= yes | X | -c |
| Disability\_\_\_7 | Living with chronic pain | 0= no, 1= yes | X | -c |
| Disability\_\_\_9 | Living with other disability | 0= no, 1= yes | X | -c |
| **Mental health** | | | |  |
| mental\_health | Mental health ratings | 1 – 5 (poor to excellent) | X | X |
| stresslife | Stress levels in most days | 1 – 5 (not stressful to extremely stressful) | X | X |
| con\_eating | diagnosed mental health condition: anorexia or bulimia | 0= no, 1= yes | X | -d |
| con\_anxiety | diagnosed mental health condition: anxiety disorder | 0= no, 1= yes | X | -d |
| con\_ADD | diagnosed mental health condition: ADD | 0= no, 1= yes | X | -d |
| con\_ADHD | diagnosed mental health condition: ADHD | 0= no, 1= yes | X | -d |
| con\_bipolar | diagnosed mental health condition: bipolar disorder | 0= no, 1= yes | X | -d |
| con\_depression | diagnosed mental health condition: depression disorder | 0= no, 1= yes | X | -d |
| con\_OCD | diagnosed mental health condition: OCD | 0= no, 1= yes | X | -d |
| con\_panic | diagnosed mental health condition: panic disorder | 0= no, 1= yes | X | -d |
| con\_PTSD | diagnosed mental health condition: PTSD | 0= no, 1= yes | X | -d |
| con\_others | diagnosed mental health condition: others (including dysthymia, mania, phobia, psychosis, and schizophrenia) | 0= no, 1= yes | X | -d |
| suicidal | Past 12 month suicidal thought | 1 = never, 2 = less than once a month, 3 = 1 to 3 times per month, 4 = 1 to 6 times per week, 5 = every day | X | X |
| Seek\_help | Wanted to talk or seek help with a professional | 0 = No, 1 = Yes | X | X |
| Receive\_help | Delay the mental health care need | 0 = No, 1 = Yes, 2 = did not seek help | X | X |
| **Smoking related behaviours** | | | |  |
| curr\_smoke | Current smoking frequency | 0= non-smoker, 1= daily or almost daily, 2= less than daily but at least weekly, 3 = less than weekly but at least monthly, 4 = less than monthly | X | X |
| use\_cigar | Past 30 day use of cigar or cigarillos | 0= no, 1= yes | X | X |
| use\_wp | Past 30 day use of waterpipe (e.g., hookah, sheesha) | 0= no, 1= yes | X | X |
| use\_smokeless | Past 30 day use of smokeless tobacco (e.g., snus.chew) | 0= no, 1= yes | X | X |
| covid | COVID-19 impact on smoking | 0= Don’t smoke, 1= Increased, 2= Decreased or quit, 3 = No change | X | X |
| quit\_attempts | Past 12 month quit attempt | 0= no, 1 = yes, 2= non-smoker | X | X |
| tailored | Wants service tailored to LGBTQI2S+ people | 0= no, 1= yes, 2= don’t know,  3= non-smoker, | X | X |
| quit\_support | Confidence of finding support for quitting smoking | 0= not confident, 1= confident, 2= neutral, 3=non-smoker | X | X |
| curr\_vape | Current vaping frequency | 0= never vaper, 1= current vaper, 2= ever vaper | X | X |
| **Substance use** |  |  |  |  |
| alcohol\_amount | Past 12 month frequency of 5 or more drinks in one occasion | 0 = none, 1 = yes, 2= non-drinker | X | X |
| cannabis | Past 12 month cannabis use frequency | 0 = none, 1 = regular use, 2 = occasional use | X | X |
| drug\_12m | Past 12 month use of any drug (other than alcohol or cannabis) | 0 = no, 1= yes | X | X |
| poppers | Past 12 month use of poppers/amyl | 0 = no, 1= yes | X | -c |
| special\_K | Past 12 month use of ketamine/Special K | 0 = no, 1= yes | X | -c |
| MDMA | Past 12 month use of ecstasy/MDMA | 0 = no, 1= yes | X | -c |
| crystal\_meth | Past 12 month use of crystal meth/Tina | 0 = no, 1= yes | X | -c |
| crack | Past 12 month use of crack, free base | 0 = no, 1= yes | X | -c |
| cocaine | Past 12 month use of cocaine | 0 = no, 1= yes | X | -c |
| heroin | Past 12 month use of heroin(smack) | 0 = no, 1= yes | X | -c |
| pres\_opioids | Past 12 month use of prescription opioids (e.g., Percocet, Dialudid, OxyContin) | 0 = no, 1= yes | X | -c |
| fentanyl | Past 12 month use of Fentanyl | 0 = no, 1= yes | X | -c |
| GHB | Past 12 month use of GHB/ “G” | 0 = no, 1= yes | X | -c |
| tranquilizers | Past 12 month use of tranquilizers or benzos (e.g., Valium, Xanax) | 0 = no, 1= yes | X | -c |
| psychedelics | Past 12 month use of psychedelics (e.g., LSD, mescaline, acid, mushrooms) | 0 = no, 1= yes | X | -c |
| drug\_others | Past 12 month use of other drugs | 0 = no, 1= yes | X | -c |
| substances\_covid | COVID-19 impact on consumption of alcohol, cannabis or other drugs | 0= Don’t smoke, 1= Increased, 2= Decreased or quit, 3 = No change | X | X |
| **Composite scores: made following imputation, all scores were standardized into standard Normal distribution** | | | | |
| cen\_identity | Overall identity centrality score (e.g.: agreement with statements such as “My sexual orientation/gender identity is a central part of my identity”) | Range 4-20 (strongly disagree to strongly agree); standardized to standard normal with mean 0 and SD 1 | X | X |
| outness | Average openness score regarding sexual orientation with family, friends and other people | Range 0-7 (“definitely doesn’t know” to “definitely knows, talk about”); standardized to standard normal with mean 0 and SD 1 | X | X |
| connect\_com | Connection to the community (e.g.: “You feel you're a part of your community”) | Range 7-35 (“strongly disagree” to “strongly agree”; standardized to standard normal with mean 0 and SD 1 | X | X |
| per\_stigma | Overall perceived stigma score (e.g. Heard that being LGBTQI2S+ is not normal? ) | Range 5-25 (“never” to “many times”); standardized to standard normal with mean 0 and SD 1 | X | X |
| en\_stigma | Overall enacted stigma score (e.g. Been hit or beaten up for being LGBTQI2S+? ) | Range 8-40 (“never” to “many times”); standardized to standard normal with mean 0 and SD 1 | X | X |
| phobia | Overall internalized homophobia score (I am comfortable with people knowing that I identify as LGBTQI2S+) | Range 13-53 (“Strongly disagree” to “strongly agree”; standardized to standard normal with mean 0 and SD 1 | X | X |
| ace | Overall adverse childhood experiences (ACEs) score (e.g. “Did you live with anyone who was depressed, mentally ill or suicidal?” Or “How often did anyone at least five years older than you or an adult ever force you to have sex?”) | Range 0-12 (for binary items: 1 = Yes; 0 = No; frequency items: 1 = more than once or once; 0 = never) standardized to standard normal with mean 0 and SD 1 | X | X |
| Cesd\_score | Sum of CESD score at baseline | Continuous score (7 to 28); standardized to standard normal with mean 0 and SD 1 | X | X |
| **Smoking and mental health stigma** | | | | |
| rates | Opinion- smoking rates are higher among LGBTQI2S+ young adults | 0= disagree, 1= agree, 2=neutral | X | X |
| accept | Opinion – Smoking is more acceptable among LGBTQI2S+ youngadults | 0= disagree, 1= agree, 2=neutral | X | X |
| stress | Opinion- LGBTQ2S+ young adults deal with more stress | 0= disagree, 1= agree, 2=neutral | X | X |
| stigma | Opinion- LGBTQI2S+ young adults deal with more stigma (e.g., discrimination) | 0= disagree, 1= agree, 2=neutral | X | X |
| pressure | Opinion- social pressure to smoke are more among LGBTQI2S+ young adults | 0= disagree, 1= agree, 2=neutral | X | X |
| mhealth | Opinion- LGBTQI2S+ young adults have more mental health concerns | 0= disagree, 1= agree, 2=neutral | X | X |
| culture | Opinion- smoking is part of LGBTQI2S+ culture | 0= disagree, 1= agree, 2=neutral | X | X |
| **Perception of smoking risk** |  |  |  |  |
| risk1 | Risk of smoking cigarette daily | 0= no or low risk, 1= moderate risk, 2=high or very high risk | X | X |
| risk2 | Risk of smoking cigarette some day but not daily | 0= no or low risk, 1= moderate risk, 2=high or very high risk | X | X |
| risk3 | Risk of using e-cigarette daily | 0= no or low risk, 1= moderate risk, 2=high or very high risk | X | X |
| risk4 | Risk of using e-cigarette some day but not daily | 0= no or low risk, 1= moderate risk, 2=high or very high risk | X | X |

1. Analysis 1: changes in scores of depressive symptoms; Analysis 2: incidence of new mental health diagnosis from a physician.
2. Two spirit and questioning are combined as “other” orientation due to small cells in Analysis 2
3. All comorbidities, drug use in the past 12 months, and disabilities were combined into three indicator variables in Analysis 2, not broken down by each condition or drug as in Analysis 1 to prevent cases of zero conditions.
4. The analytic sample in Analysis 2 does not have existing mental health conditions by definition since they were never diagnosed by a healthcare professional, so this variable cannot be included.

Appendix 2. Correlations between predictor variables that were larger than 0.6

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Appendix 3. Performances of Random Forest and LASSO models at each of the 5 imputed datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Imputed Data # | | | | |
|  | 1 | 2 | 3 | 4 | 5 |
| **Depressive Symptoms scores** |  |  |  |  |  |
| RMSE (RF) | 0.274 | 0.292 | 0.327 | 0.303 | 0.306 |
| RMSE (LASSO) | 0.297 | 0.352 | 0.289 | 0.318 | 0.343 |
| **New Mental Health Diagnosis** |  |  |  |  |  |
| Brier’s score (RF) | 0.164 | 0.135 | 0.162 | 0.129 | 0.135 |
| Brier’s score (LASSO) | 0.141 | 0.131 | 0.144 | 0.156 | 0.133 |
| AUCROC (RF) | 0.544 (0.444, 0.644) | 0.583 (0.473, 0.694) | 0.544 (0.444, 0.644) | 0.598 (0.485, 0.711) | 0.533 (0.449, 0.616) |
| AUCROC (LASSO) | 0.791 (0.669, 0.913) | 0.744 (0.589, 0.899) | 0.794 (0.651, 0.937) | 0.697 (0.543, 0.851) | 0.712 (0.531, 0.893) |

Appendix 4. Predictors selected by the LASSO models for each imputed data

**Depressive symptom scores:**

Data 1

[1] "suicidal" "curr\_orient23" "gender3" "trans1"

[5] "poc1" "education2" "ind\_income2" "ind\_income5"

[9] "fitness11" "stresslife" "con\_depression1" "treat\_comorbid1"

[13] "receive\_help" "curr\_smoke3" "covid1" "time\_vape1"

[17] "disability\_\_\_11" "disability\_\_\_21" "disability\_\_\_9" "pres\_opioids1"

[21] "fentanyl1" "substances\_covid2" "stress" "stigma"

[25] "pressure" "ethnicity2" "ethnicity4" "ethnicity6"

[29] "connect\_com" "ace" "phobia" "per\_stigma"

[33] "en\_stigma" "cesd\_score"

Data 2

[1] "suicidal" "province9" "curr\_orient23" "gender3"

[5] "trans1" "intersex1" "poc1" "education2"

[9] "house\_income3" "house\_income5" "house\_income6" "ind\_income2"

[13] "ind\_income5" "gen\_health" "con\_depression1" "treat\_comorbid1"

[17] "cigs\_smoked5" "covid2" "covid3" "plan\_quit2"

[21] "tailored3" "quit\_support1" "cannabis1" "cannabis2"

[25] "pres\_opioids1" "fentanyl1" "MDMA1" "substances\_covid2"

[29] "accept" "stress" "mhealth" "culture"

[33] "ethnicity4" "ethnicity7" "connect\_com" "ace"

[37] "per\_stigma" "en\_stigma" "cesd\_score"

Data 3

[1] "suicidal" "curr\_orient23" "gender3" "trans1"

[5] "intersex1" "poc1" "education2" "ind\_income5"

[9] "gen\_health" "fitness11" "stresslife" "diagnosis1"

[13] "con\_depression1" "cigs\_smoked3" "cigs\_smoked4" "covid1"

[17] "quit\_support1" "time\_vape1" "cannabis1" "cannabis2"

[21] "pres\_opioids1" "substances\_covid2" "stress" "stigma"

[25] "pressure" "ethnicity2" "ethnicity4" "connect\_com"

[29] "ace" "per\_stigma" "en\_stigma" "cesd\_score"

Data 4

[1] "suicidal" "trans1" "poc1" "education2"

[5] "ind\_income5" "stresslife" "con\_depression1" "treat\_comorbid1"

[9] "curr\_smoke2" "curr\_smoke3" "covid2" "quit\_support1"

[13] "disability\_\_\_11" "cannabis1" "cannabis2" "crystal\_meth1"

[17] "pres\_opioids1" "substances\_covid1" "substances\_covid2" "stress"

[21] "stigma" "pressure" "ethnicity4" "ethnicity6"

[25] "ace" "per\_stigma" "en\_stigma" "cesd\_score"

Data 5

[1] "suicidal" "age" "province9" "curr\_orient24"

[5] "gender2" "trans1" "poc1" "education2"

[9] "ind\_income5" "where\_live1" "con\_depression1" "treat\_comorbid1"

[13] "curr\_smoke3" "covid1" "risk1" "cannabis1"

[17] "cannabis2" "poppers1" "cocaine1" "pres\_opioids1"

[21] "substances\_covid2" "stress" "stigma" "pressure"

[25] "ethnicity4" "ethnicity5" "ethnicity6" "connect\_com"

[29] "ace" "phobia" "per\_stigma" "en\_stigma"

[33] "cesd\_score"

**New Mental Health Diagnosis:**

Data 1

[1] "suicidal" "curr\_orient23" "curr\_orient26" "education2" "gen\_health" "mental\_health"

[7] "stresslife" "disability1" "use\_wp1" "risk3" "alcohol1" "drug\_12m1"

[13] "seek\_help1" "ethnicity3"

Data 2

[1] "age" "education2" "mental\_health" "stresslife" "disability1" "covid2"

[7] "seek\_help1" "cesd\_score"

Data 3

[1] "suicidal" "mental\_health" "stresslife" "seek\_help1"

Data 4

[1] "education2" "gen\_health" "mental\_health" "stresslife"

[5] "disability1" "cigs\_smoked4" "use\_wp1" "alcohol1"

[9] "alcohol\_amount1" "cannabis1" "substances\_covid4" "seek\_help1"

[13] "ethnicity3" "ethnicity5"

Data 5

[1] "age" "curr\_orient24" "education2" "house\_income6" "mental\_health" "stresslife"

[7] "disability1" "alcohol1" "seek\_help1"

Appendix 5. Top Ten Interaction Variables from the Random Forest Model

|  |  |
| --- | --- |
| CESD Scores |  |
| **Interaction** | **Variables** |

|  |  |
| --- | --- |
| 0.013223 | cesd\_score\*ethnicity |
| 0.008775 | cesd\_score\*age |
| 0.007385 | cesd\_score\*curr\_orient2 |
| 0.007026 | cesd\_score\*house\_income |
| 0.005655 | suicidal\*ethnicity |
| 0.004643 | cesd\_score\*education |
| 0.003385 | cesd\_score\*where\_live |
| 0.003033 | suicidal\*house\_income |
| 0.002728 | cesd\_score\*gender |
| 0.002653 | per\_stigma\*age |

|  |  |
| --- | --- |
| New Mental Health Diagnosis |  |
| **Interaction** | **Variables** |

|  |  |
| --- | --- |
| 0.076979 | stresslife\*age |
| 0.064452 | stresslife\*ethnicity |
| 0.057381 | en\_stigma\*age |
| 0.051778 | per\_stigma\*age |
| 0.050317 | en\_stigma\*ethnicity |
| 0.048451 | en\_stigma\*gender |
| 0.047913 | cesd\_score\*ethnicity |
| 0.046983 | en\_stigma\*education |
| 0.04662 | stresslife\*gender |
| 0.046527 | phobia\*gender |