Exploring sociodemographic predictors of depressive symptoms among French undergraduates.

Authors: Ben Pitter-Fagan, Courtney Reid, Kayesha Coley, Maya Chinegwundoh, Maya Ogonah, Myron Odingo, Sally Babiker, Shukrat Salisu-Olatunji, Wayne Matengo

Group Number: 2

1. Introduction

Mental health disorders are highly prevalent globally with variability by age, gender and location. Approximately 29.2% of respondents across 174 surveys in 26 high income countries (HICs) and 37 low and middle income countries (LMICs) experienced at least one common mental disorder (CMD) across their lifetime (Steel, et al 2014). The World Health organisation (WHO) Global Health estimates that 4.4% and 3.6% of the global population suffer from anxiety and depressive disorders respectively (WHO, 2017).

Empirical evidence suggests that the rate of psychopathology is higher among university students, compared to the general population in HICs such as the UK (Ibrahim et al., 2013) and Australia (Stallman, 2010). Research by the WHO Work Mental Health International College Student (WMH-ICS) project has found that approximately one third of first-year students screen positive for at least one common DSM-IV mood, anxiety or substance use disorder (Auerbach et al., 2018). This high prevalence of CMD is significant not only for the distress it can cause during a time of great transition, but it is associated with worsening social adjustment to university life (Alonso et al., 2018), substantial impairment in academic achievement (Bruffaerts et al., 2018), and adverse health outcomes including death by suicide (McLafferty et al., 2017). Thus, timely and effective treatment is needed to address CMD and its adverse outcomes in university students.

It is of note that the number of students requiring treatment for CMD far exceeds the mental health resources available, resulting in a substantial unmet need for mental health services amongst university students. Early identification of those at risk for mental health disorders is therefore essential as primary preventive interventions have been highlighted as part of the public health response by the WHO Euro region to reduce the burden of mental health disorders (WHO, 2018). Socio-demographic predictors could be useful for this purpose, hence this data analysis hopes to identify socio-demographic correlates to mental health disorders that could help in screening in the population.

Prior research by the WMH-ICS project, which sought to assess CMD and sociodemographic correlates using a series of surveys in 19 colleges across 8 countries (Australia, Belgium, Germany, Mexico, Northern-Ireland, South-Africa, Spain and the United States), suggests that female gender, older age (i.e., 19 and 20+), lack of religious affiliation, and non-heterosexual orientation are sociodemographic correlates of CMD in university students (Auerbach et al., 2018). Furthermore, evidence assessing the prevalence of socio-demographic correlates of CMD among students in South Africa found that female gender, non-heterosexual orientation, and disability were associated with higher lifetime and 12-month risk of CMD (Bantjes et al., 2019).

The challenge arises therefore in the identification of students requiring treatment for existing CMD and of those requiring preventive interventions due to susceptibility to CMDs. Reliable analysis of the prevalence and correlates of CMD in university students is vital, so that effective preventive interventions can be targeted towards those most at risk. Despite evidence assessing CMD and sociodemographic correlates in some HICs and middle-income countries (MIC), to date little research has assessed CMD and sociodemographic predictors in a French student sample – even though France records the second highest number suicides (9200), in absolute terms, in the European Union (EU; Eurostat, 2018). Even when adjusting for population size and structure, France records elevated levels of suicide (14 per 100,000), compared to an EU average of 11 per 100,000. (Eurostat, 2018). The minimal evidence assessing the sociodemographic correlates of CMD among French university students presents a gap in the literature, which this project seeks to address.

2. Aim

This is an exploratory study with the aim of describing the potential factors associated with depressive symptoms among university students in France.

3. Rationale

The study findings could potentially inform the aligning of student support services to address the potential factors associated with depressive symptoms. The work also serves as a feasibility study aimed at identifying statistically significant associations with depressive symptoms; and thereby informing future and more robust predictive modelling projects.

4. Methods

The dataset was downloaded from Dryad (https://datadryad.org/stash/) and presents information on 4,184 undergraduates who underwent compulsory medical assessments at the University Medical service in Nice, France between September 2012 and June 2013. Demographic, lifestyle-related and biophysical information was collected, including living conditions, dietary patterns, physical activity, financial status and anthropometric measurements (Tran, et al 2018).

4.1 Data Preprocessing

Once the data was obtained, we sought to pre-process the raw data in two stages. These stages were conducted using R and Python programming frameworks.

I. Data Quality assessment

The key objective of this stage was to examine the data in a bid to establish the accuracy, completeness and reliability of the data. This was made possible using the following questions:

I. What data is available?

II. Is the data complete?

III. Is the data accurate?

IV. Does the data have a wide range of variability?

The dataset had a total of 4,184 observations, each having 60 features. A list of the features present in the data is given below:

```
Rows: 4,184
Columns: 62
$ `Age (4 levels)`
$ Gender
$ `French nationality`
$ `Field of study`
$ `Year of university`
$ `Learning disabilities`
$ `Difficulty memorizing lessons`
$ `Professional objective`
$ `Informed about opportunities`
$ `Satisfied with living conditions`
$ `Living with a partner/child`
$ `Parental home`
$ `Having only one parent`
$ `At least one parent unemployed`
$ Siblings
$ `Long commute`
$ `Mode of transportation`
$ `Financial difficulties`
$ Grant
$ `Additional income`
$ `Public health insurance`
$ `Private health insurance`
$ C.M.U.
$ `Irregular rhythm of meals`
$ `Unbalanced meals`
$ `Eating junk food`
$ `On a diet`
$ `Irregular rhythm or unbalanced meals`
$ `Physical activity(3 levels)`
$ `Physical activity(2 levels)`
$ `Weight (kg)`
$ `Height (cm)`
$ `Overweight and obesity`
$ `Systolic blood pressure (mmHg)`
$ `Diastolic blood pressure (mmHg)`
$ `Prehypertension or hypertension`
$ `Heart rate (bpm)`
$ `Abnormal heart rate`
$ `Distant visual acuity of right eye (score /10)`
$ `Distant visual acuity of left eye (score /10)`
$ `Close visual acuity of right eye (score /10)`
$ `Close visual acuity of left eye (score /10)`
$ `Decreased in distant visual acuity`
$ `Decreased in close visual acuity`
$ `Urinalysis (glycosuria)`
$ `Urinalysis (proteinuria)`
$ `Urinalysis (hematuria)`
$ `Urinalysis leukocyturia)`
$ `Urinalysis (positive nitrite test)`
$ `Abnormal urinalysis`
$ `Vaccination up to date`
$ `Control examination needed`
$ `Anxiety symptoms`
$ `Panic attack symptoms`
$ `Depressive symptoms`
$ `Cigarette smoker (5 levels)`
$ `Cigarette smoker (3 levels)`
$ `Drinker (3 levels)`
$ `Drinker (2 levels)`
$ `Binge drinking`
$ `Marijuana use`
$ `Other recreational drugs`
```

Figure 1 - Snippet from R showing the features available in the dataset

A further two features, ‘BMI’ and ‘BMI_Category’, were derived using variables already present in the data. ‘BMI’ was derived based on weight (‘Weight (kg)’) and height (‘Height (cm)’) converted to m, using the formula BMI = weight/height^2. ‘BMI_Category’ was defined based on the value of BMI and corresponding BMI category (Centers for Disease Control and Prevention, 2021).

The features available were consistent with the research objectives and therefore there was no need to look for supplementary data. Furthermore, these features contained various information that alluded to the distinction of a single data point from another. The data, however, had several missing values.

```
                        Age (4 levels)                                  Gender                            French nationality
                                     0                                       0                                             0
                        Field of study                      Year of university                         Learning disabilities
                                     0                                       0                                             0
         Difficulty memorizing lessons                  Professional objective                  Informed about opportunities
                                     0                                       0                                             0
      Satisfied with living conditions            Living with a partner/child                                Parental home
                                   514                                       0                                           714
                Having only one parent          At least one parent unemployed                                     Siblings
                                    60                                       0                                             0
                          Long commute                  Mode of transportation                        Financial difficulties
                                   936                                    1120                                             0
                                 Grant                       Additional income                       Public health insurance
                                     0                                       0                                             0
              Private health insurance                                  C.M.U.                     Irregular rhythm of meals
                                     2                                       0                                             0
                       Unbalanced meals                        Eating junk food                                    On a diet
                                     0                                       0                                             0
  Irregular rhythm or unbalanced meals              Physical activity(3 levels)                  Physical activity(2 levels)
                                     0                                       0                                             0
                           Weight (kg)                             Height (cm)                        Overweight and obesity
                                   190                                     187                                           190
        Systolic blood pressure (mmHg)          Diastolic blood pressure (mmHg)              Prehypertension or hypertension
                                  1152                                    1073                                          1534
                      Heart rate (bpm)                     Abnormal heart rate Distant visual acuity of right eye (score /10)
                                   386                                     386                                             0
Distant visual acuity of left eye (score /10) Close visual acuity of right eye (score /10) Close visual acuity of left eye (score /10)
                                     0                                       0                                             0
     Decreased in distant visual acuity        Decreased in close visual acuity                     Urinalysis (glycosuria)
                                     0                                       0                                             0
              Urinalysis (proteinuria)                  Urinalysis (hematuria)                     Urinalysis leukocyturia)
                                     0                                       0                                             0
   Urinalysis (positive nitrite test)                     Abnormal urinalysis                       Vaccination up to date
                                     0                                       0                                          1071
           Control examination needed                        Anxiety symptoms                        Panic attack symptoms
                                     0                                       0                                             0
                   Depressive symptoms             Cigarette smoker (5 levels)                  Cigarette smoker (3 levels)
                                     0                                     620                                           620
                    Drinker (3 levels)                      Drinker (2 levels)                               Binge drinking
                                     8                                       8                                             0
                         Marijuana use                Other recreational drugs                                          BMI
                                   568                                    1273                                           190
                          BMI_Category
                                   209
```

Figure 2 - Snippet from R showing the missing values under each column

Out of the 4,184 observations, 3,257 observations had at least a single feature missing. We therefore, decided to drop these rows and this resulted in having 927 observations. The accuracy checks we conducted involved checking for duplicate observations, checking for outliers and verifying whether the information is consistent with general knowledge (reality).

II. Data Transformation

The next stage involved encoding the different features into machine-readable format. Our data had a large number of categorical variables, such as Gender, Age-group and others. We converted the data into machine readable format seamlessly using R-packages such as dplyr and tidyverse.

```{r}
categories <- c("Gender",
                "Age (4 levels)",
                "French nationality",
                "Field of study",
                "Year of university",
                "Learning disabilities",
                "Difficulty memorizing lessons",
                "Professional objective",
                "Informed about opportunities",
                "Satisfied with living conditions",
                "Living with a partner/child",
                "Parental home",
                "Having only one parent",
                "At least one parent unemployed",
                "Siblings",
                "Long commute",
                "Mode of transportation",
                "Financial difficulties",
                "Grant",
                "Additional income",
                "Public health insurance",
                "Private health insurance",
                "C.M.U.",
                "Irregular rhythm of meals",
                "Unbalanced meals",
                "Eating junk food",
                "On a diet",
                "Irregular rhythm or unbalanced meals",
                "Physical activity(3 levels)",
                "Physical activity(2 levels)",
                "Other recreational drugs",
                "Marijuana use",
                "Binge drinking",
                "Drinker (2 levels)",
                "Drinker (3 levels)",
                "Cigarette smoker (3 levels)",
                "Cigarette smoker (5 levels)",
                "Depressive symptoms",
                "Panic attack symptoms",
                "Anxiety symptoms"
                )
clean_data <-
  raw_data %>%
  mutate_at(categories, factor)
```

```
    Age (4 levels)    Gender
18          :366    female:553
19          :240    male  :374
20 and more:272
less 18     : 49

French nationality
no : 54
yes:873

                          Field of study
humanities                      :255
law and political sciences     :139
medicine and allied programs:165
other programs                  :193
sciences                        :108
sports science                  : 67
Year of university Learning disabilities
first :854          no :923
second: 39          yes:  4
third : 34

Difficulty memorizing lessons
no :911
yes: 16

Professional objective
no :171
yes:756

Informed about opportunities
no : 39
yes:888

Satisfied with living conditions
no : 46
yes:881
```

Figure 3 - Snippet from R showing the transformation process

4.2 Exploratory Data Analysis

The Pandas profiling module of Python was used to generate a HTML profile report for the dataset. Visualizations were created to provide preliminary insights into the relationships that existed within the data. (The Pandas profiling report is attached to this report as a separate document).

4.3 Correlation Analysis

Chi-square tests were used to assess the existence of relationships between the attribute 'depressive symptoms' and 62 independent attributes (p-value of <0.05).

4.4 Feature Identification and selection

Backward/ stepwise logistic regression analysis was conducted using the glm and stepAIC functions to aid feature selection by identifying statistically significant attributes incorporated into the model. The Akaike Information Criterion (AIC) was used to assess the quality of the model(s).

4.5 Statistical Analysis

Bootstrapping – The boot.stepAIC function was used to run the logistic model development multiple times (n=8), giving a proportion of the number of times that each attribute was included and the number of times the attributes were found to be statistically significant.

VIF – Multicollinearity was assessed using the variance inflation factor.

4.6 Model Validation

The tidymodels library was used to split the dataset into training and testing cohorts. Two logistic regression models were then developed and validated; (i) baseline model including all the features, (ii) model composed of selected features. Bootstrapping techniques were applied to run the validation process 25 times for each model and receiver operating characteristic (ROC) curves were created to compare model performance.

5. Modelling and Results

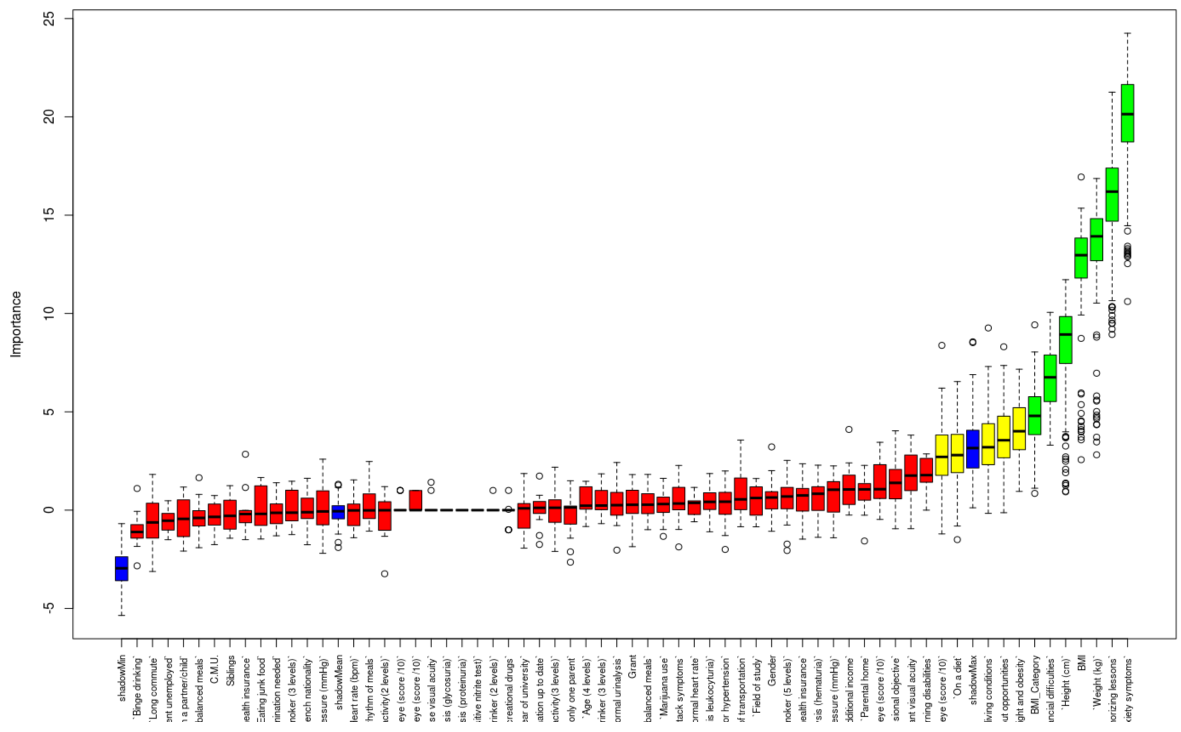


Figure 4 - Boruta algorithm output variable showing order of importance in variables

The Boruta analysis was conducted to establish any presence of correlation between the response variable and the explanatory variables. The analysis indicates that variables such as: BMI, anxiety symptoms and financial difficulties are some of the variables that were significantly related to the depressive symptoms.

This allowed for further exploration of the associations using a logistic regression model. The model with all the independent variables achieved an AIC value of 703.29 while the backward logistic regression model with 18 attributes achieved an AIC value of 636.5.

The bootstrapping procedure showed that four attributes were selected 100% of the time with every run of the model development process. The covariates selected were anxiety symptoms, difficulty memorizing lessons, grant, and professional objectives.

The VIF showed that 6 covariates were significantly associated with each other, and these were dropped from the model. The resulting model attained an AIC of 638.1.

Finally, covariates that did not have a statistically significant effect in the model were dropped leaving the following as the final model achieved from the exploratory analysis.

```
Call:
glm(formula = Depressive.symptoms ~ Difficulty.memorizing.lessons +
    Professional.objective + Financial.difficulties + Grant +
    Public.health.insurance + Anxiety.symptoms, family = "binomial",
    data = student_df1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9489  -0.5398  -0.4206  -0.3251   2.4357

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -2.1132     0.2608  -8.102 5.39e-16 ***
Difficulty.memorizing.lessonsyes  1.9794     0.5552   3.565 0.000364 ***
Professional.objectiveyes      -0.8003     0.2416  -3.312 0.000925 ***
Financial.difficultiesyes       2.6069     0.9935   2.624 0.008689 **
Grantyes                        0.5279     0.2120   2.490 0.012766 *
Public.health.insuranceyes      0.5330     0.2187   2.437 0.014828 *
Anxiety.symptomsyes             1.8708     0.2989   6.258 3.89e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 710.60  on 926  degrees of freedom
Residual deviance: 631.03  on 920  degrees of freedom
AIC: 645.03

Number of Fisher Scoring iterations: 5
```

Figure 5 - Logistic regression output

In the model validation process, the selected model outperformed the baseline model; mean ROC values were 0.59 for the baseline model (A) and 0.69 for the selected model (B). The validation process was conducted 25 times for each model, the ROC curves in figure 5 below show these results.

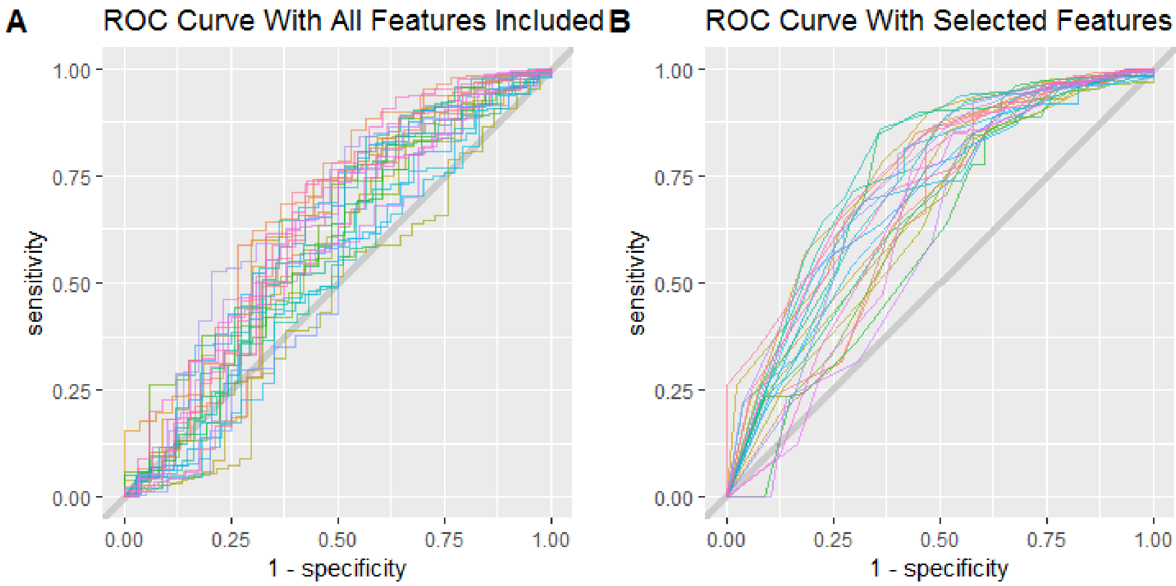


Figure 6 - ROC curves for baseline model and selected model

Discussion

The logistic regression model together with the filtering of the features led to the development of a model composed of 6 attributes that were deemed to be significantly associated with 'depressive symptoms'. This can be interpreted as follows; difficulty in memorising/ impaired memory, financial difficulties and anxiety were associated with higher odds of having depressive symptoms. This was similar to a study among Irish undergraduates that found that financial stress and poor family relationships increased the odds of experiencing depressive symptoms (Horgan, et al. 2018). Likewise another study among Malaysian undergraduates, found that higher level of study, living off-campus, poor financial status, pre-existing post-traumatic stress disorder and sleep disturbances were associated with a higher risk of developing depression (Ashraful Islam et al., 2018). In addition, a study among French undergraduates supported the evidence that academic and daily stress was associated with developing depression and further found that students with low resilience were more at risk (Kkou-Kpolou, et al., 2021). In the current analysis, having professional objectives/focus was associated with reduced odds of having depressive symptoms.

Having a study grant and public health insurance had a slight contribution to the odds of having depressive symptoms, but may warrant further interpretation as another study conversely reported that lack of access to scholarships and bursaries , especially for international students, was associated with developing anxiety and depression (Bradley, 2000).

The huge burden placed on the university's health services to provide support and treatment for CMDs has been acknowledged. The use of novel and time appropriate approaches like online counselling services could reduce the gap in mental health service provision (Mofatteh, 2020) by providing an accessible and affordable service for undergraduates (Horgan et al., 2013).

Finally, model validation showed that the selected model outperformed the baseline model composed of all the features in the dataset. This suggested that more predictive analysis work may be possible to help identify students who are at risk of developing depressive symptoms.

Challenges and Limitations

Since the research and modelling was based on secondary data, the appropriateness was always in question. There was a lot of missing data that reduced the range of responses used in the analysis after data cleaning. The reduction in sample size would normally reduce the power of a study, however this was of no serious consequence as this is an exploratory analysis.

Conclusion and Recommendations

We conclude that there may be scope for more rigorous predictive modelling for detecting depressive symptoms amongst students. This would help to tailor support services for students by focusing on those with risk factors that are significantly associated with depressive symptoms.

References

Alonso, J., Mortier, P., Auerbach, R. P., Bruffaerts, R., Vilagut, G., Cuijpers, P., ... & WHO WMH‐ICS Collaborators. (2018). Severe role impairment associated with mental disorders: results of the WHO world mental health surveys international college student project. Depression and anxiety, 35(9), 802-814.

Ashraful Islam, M., Yun Low, W., Ting Tong, W., Wan Yuen, C. C., & Abdullah, A. (2018). Factors Associated with Depression among University Students in Malaysia: A Cross-sectional Study. *KnE Life Sciences*, *4*(4), 415–427. https://doi.org/10.18502/kls.v4i4.2302

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., ... & Kessler, R. C. (2018). WHO world mental health surveys international college student project: prevalence and distribution of mental disorders. Journal of abnormal psychology, 127(7), 623-638.

Bantjes, J., Lochner, C., Saal, W., Roos, J., Taljaard, L., Page, D., ... & Stein, D. J. (2019). Prevalence and sociodemographic correlates of common mental disorders among first-year university students in post-apartheid South Africa: implications for a public mental health approach to student wellness. *BMC public health*, *19*(1), 1-12.

Bradley, G. Responding effectively to the mental health needs of international students. *Higher Education* 39, 417–433 (2000). https://doi.org/10.1023/A:1003938714191

Bruffaerts, R., Mortier, P., Kiekens, G., Auerbach, R. P., Cuijpers, P., Demyttenaere, K., ... & Kessler, R. C. (2018). Mental health problems in college freshmen: Prevalence and academic functioning. Journal of affective disorders, 225, 97-103.

Centers for Disease Control and Prevention. (2021). Defining Adult Overweight & Obesity. https://www.cdc.gov/obesity/adult/defining.html

Eurostat (2018) https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20180716-1

Horgan, A., Kelly, P., Goodwin, J., & Behan, L. (2018) Depressive Symptoms and Suicidal Ideation among Irish Undergraduate College Students, Issues in Mental Health Nursing, 39:7, 575-584, DOI: 10.1080/01612840.2017.1422199

Horgan, A., McCarthy, G., and Sweeney, J. (2013) An Evaluation of an Online Peer Support Forum for University Students With Depressive Symptoms. Archives of Psychiatric Nursing, Volume 27, Issue 2, Pages 84-89. ISSN 0883-9417. https://doi.org/10.1016/j.apnu.2012.12.005.

Ibrahim, A. K., Kelly, S. J., & Glazebrook, C. (2013). Socioeconomic status and the risk of depression among UK higher education students. Social psychiatry and psychiatric epidemiology, 48(9), 1491-1501.

Kokou-Kpolou, C.K., Jumageldinov, A., Park, S. *et al.* Prevalence of Depressive Symptoms and Associated Psychosocial Risk Factors among French University Students: the Moderating and Mediating Effects of Resilience. *Psychiatr Q* 92, 443–457 (2021). https://doi.org/10.1007/s11126-020-09812-8

McLafferty, M., Lapsley, C. R., Ennis, E., Armour, C., Murphy, S., Bunting, B. P., ... & O'Neill, S. M. (2017). Mental health, behavioural problems and treatment seeking among students commencing university in Northern Ireland. PloS one, 12(12), e0188785.

Mofatteh M. (2020). Risk factors associated with stress, anxiety, and depression among university undergraduate students. *AIMS public health*, *8*(1), 36–65. https://doi.org/10.3934/publichealth.2021004

Stallman, H. M. (2010). Psychological distress in university students: A comparison with general population data. Australian Psychologist, 45(4), 249-257.

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J.W., Patel, V., Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013, *International Journal of Epidemiology*, Volume 43, Issue 2, Pages 476–493, https://doi.org/10.1093/ije/dyu038

Tran, A., Tran, L., & Geghre, N, et al. (2018), Data from: Health assessment of French university students and risk factors associated with mental health disorders, Dryad, Dataset. Available at https://datadryad.org/stash/dataset/doi:10.5061/dryad.54qt7

W. H. O. (2017). Depression. Other common mental disorders: global health estimates. Geneva: World Health Organization, 1-24.

WHO Regional office for Europe (2018). Child and adolescent mental health. Available at https://www.euro.who.int/en/health-topics/noncommunicable-diseases/mental-health/areas-of-work/child-and-adolescent-mental-health

Appendices

1. Data Utility Framework

Dataset 1 (please specify)	Dryad dataset: Health assessment of French university students and risk factors associated with mental health disorders https://datadryad.org/stash/dataset/doi:10.5061/dryad.54qt7

Category	Dimension	Definition	Bronze	Silver	Gold	Platinum	Please select the description that most closely matches the dataset	Notes *(free text)*
Data Documentation	Documentation Completeness	This element will be calculated separately						
	Availability of additional documentation and support	Available dataset documentation in addition to the data dictionary	Past journal articles demonstrate that knowledge of the data exists	Comprehensive ReadMe describing extracting and use of data, Dataset FAQS available, Visual data model provided	Dataset publication was supported with a journal article explaining the dataset in detail, or dataset training materials	As Gold, plus support personnel available to answer questions	Gold	An open-access publication outlining details on recruitment strategy, study period and data collection methods. There are also journal articles which have cited findings from the paper, and further additional analyses have been performed using the dataset e.g. predictive modelling.
	Data Model	Availability of clear, documented data model	Known and accepted data model but some key field un-coded or free text	Key fields codified using a local standard	Key fields codified using a national or international standard	Data Model conforms to a national standard and key fields codified using a national / international standard	Bronze	There are no free text fields in the dataset. Column names do not follow standard naming conventions.
	Data Dictionary	Provided documented data dictionary and terminologies	Data definitions available	Definitions compiled into local data dictionary which is available online	Dictionary relates to national definitions	Dictionary is based on international standards and includes mapping	Bronze	Data definitions can be found in accompanying published journal article, but there is currently no data dictionary.
	Provenance	Clear description of source and history of the dataset, providing a "transparent data pipeline"	Source of the dataset is documented	Source of the dataset and any transformations, rules and exclusions documented	All original data items listed, all transformations, rules and exclusion listed and impact of these	Ability to view earlier versions, including versions before any transformations have been applied data (in line with deidentification and IG approval) and review the impact of	Bronze	A single 'final' version of the dataset is publically available.
Technical Quality	Data Quality Management Process	The level of maturity of the data quality management processes	A documented data management plan covering collection, auditing, and management is available for the dataset	Evidence that the data management plan has been implemented is available		Externally verified compliance with the data management plan, e.g. by ISO, CQC, ICO or other body	Not yet Bronze	An outline of the data collection methods is presented in the published journal article, however this lacks specific details on planning, auditing and management.
	Data Management Association (DAMA) Quality Dimensions	These elements will be calculated separately						
Coverage	Pathway coverage	Representation of multi-disciplinary healthcare data	Contains data from a single speciality or area	Contains data from multiple specialties or services within a single tier of care	Contains multimodal data or data that is linked across two tiers (e.g. primary and secondary care)	Contains data across more than two tiers	Bronze	Data was collected from undergraduate students from a range of faculties at the University of Nice Sophia-Antipolis (UNSA) in France.
	Length of follow up	Average timeframe in which a patient appears in a dataset (follow up period)	Between 1 - 6 months	Between 6 - 12 months	Between 1 - 10 years	More than 10 years	Not yet Bronze	No follow-up data present in dataset.

Access & Provision	Allowable uses	Allowable dataset usages as per the licencing agreement		Non-consented, aggregate data for specific academic uses (following IG approval)	Aggregate data, for academic and specific commercial uses (following IG approval)	Fully consented for commercial uses (following IG approval)	Platinum	The dataset is licenced under a CC0 1.0 Universal Public Domain Dedication license, which permits use of the data academically, and for commercial purposes, even without permission.
	Time Lag	Lag between the data being collected and added to the dataset	Approximately 1 year	Approximately 1 month	Approximately 1 week	Effectively real-time data	Other (please specify)	There is no information on the time it took to add the data to the data set.
	Timeliness	Average data access request timeframe	Less than 6 months	Less than 3 months	Less than 1 month	Less than 2 weeks	Platinum	Readily available from https://datadryad.org/stash/dataset/doi:10.5061/dryad.54qt7
Value & Interest	Linkages	Ability to link with other datasets	Identifiers to demonstrate ability to link to other datasets	Available linkages outlined and/or List of datasets previously successfully linked provided	List of restrictions on the type of linkages detailed. List of previously successful dataset linkages performed, with navigable links to linked datasets via at DOI/URL	Existing linkage with reusable or downstream approvals	Other (please specify)	No identifiers present for linkage to other datasets.
	Data Enrichments	Data sources enriched with annotations, image labels, phenomes, derivations, NLP derived data labels	The data include additional derived fields, or enriched data.	The data include additional derived fields, or enriched data used by other available data sources.	The derived fields or enriched data were generated from, or used by, a peer reviewed algorithm.	The data includes derived fields or enriched data from a national report.	Not yet Bronze	There are no additional derived fields in the dataset.

2. Tables showing statistically significant associations (made using Stata)

```
. tabulate Financialdifficulties Depressivesymptoms, column row nokey chi2 lrchi2 V

+-------------------+
| Key               |
|-------------------|
|     frequency     |
|  row percentage   |
| column percentage |
+-------------------+

 Financial |
difficulty |   Depressive symptoms
           |        no        yes |     Total
-----------+----------------------+----------
        no |       806        115 |       921
           |     87.51      12.49 |    100.00
           |     99.75      96.64 |     99.35
-----------+----------------------+----------
       yes |         2          4 |         6
           |     33.33      66.67 |    100.00
           |      0.25       3.36 |      0.65
-----------+----------------------+----------
     Total |       808        119 |       927
           |     87.16      12.84 |    100.00
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =  15.6392   Pr = 0.000
 likelihood-ratio chi2(1) =   9.4361   Pr = 0.002
               Cramér's V =   0.1299
```

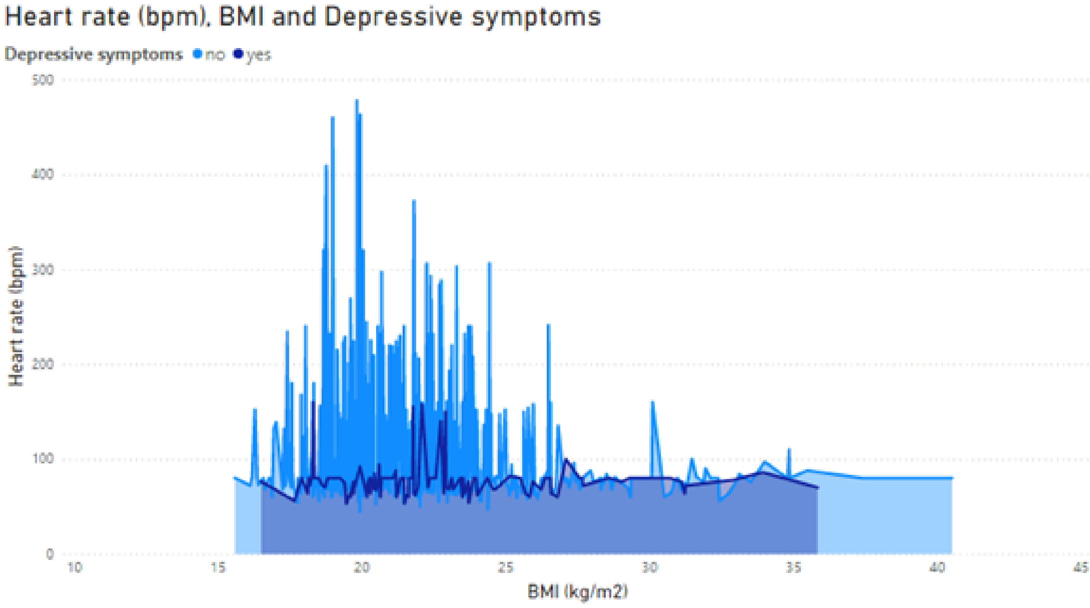
```
. tabulate Difficultymemorizinglessons Depressivesymptoms,  column row key chi2 lrchi2 V

+-------------------+
| Key               |
|-------------------|
|     frequency     |
|  row percentage   |
| column percentage |
+-------------------+

 Difficulty |
memorizing |  Depressive symptoms
   lessons |        no        yes |     Total
-----------+----------------------+----------
        no |       801        110 |       911
           |     87.93      12.07 |    100.00
           |     99.13      92.44 |     98.27
-----------+----------------------+----------
       yes |         7          9 |        16
           |     43.75      56.25 |    100.00
           |      0.87       7.56 |      1.73
-----------+----------------------+----------
     Total |       808        119 |       927
           |     87.16      12.84 |    100.00
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =  27.4233   Pr = 0.000
 likelihood-ratio chi2(1) =  17.4259   Pr = 0.000
               Cramér's V =   0.1720
```

```
. tabulate Anxietysymptoms Depressivesymptoms,  column row key chi2 lrchi2 V

+-------------------+
| Key               |
|-------------------|
|     frequency     |
|  row percentage   |
| column percentage |
+-------------------+

   Anxiety |  Depressive symptoms
  symptoms |        no        yes |     Total
-----------+----------------------+----------
        no |       774         92 |       866
           |     89.38      10.62 |    100.00
           |     95.79      77.31 |     93.42
-----------+----------------------+----------
       yes |        34         27 |        61
           |     55.74      44.26 |    100.00
           |      4.21      22.69 |      6.58
-----------+----------------------+----------
     Total |       808        119 |       927
           |     87.16      12.84 |    100.00
           |    100.00     100.00 |    100.00

          Pearson chi2(1) =  57.6300   Pr = 0.000
 likelihood-ratio chi2(1) =  40.4330   Pr = 0.000
               Cramér's V =   0.2493
```

```
. tabulate Publichealthinsurance Depressivesymptoms,  column row key chi2 lrchi2 V

+-------------------+
| Key               |
|-------------------|
|     frequency     |
|  row percentage   |
| column percentage |
+-------------------+

    Public |
    health |  Depressive symptoms
 insurance |        no        yes |     Total
-----------+----------------------+----------
        no |       366         42 |       408 
           |     89.71      10.29 |    100.00 
           |     45.30      35.29 |     44.01 
-----------+----------------------+----------
       yes |       442         77 |       519 
           |     85.16      14.84 |    100.00 
           |     54.70      64.71 |     55.99 
-----------+----------------------+----------
     Total |       808        119 |       927 
           |     87.16      12.84 |    100.00 
           |    100.00     100.00 |    100.00 

          Pearson chi2(1) =   4.2118   Pr = 0.040
 likelihood-ratio chi2(1) =   4.2838   Pr = 0.038
               Cramér's V =   0.0674
```

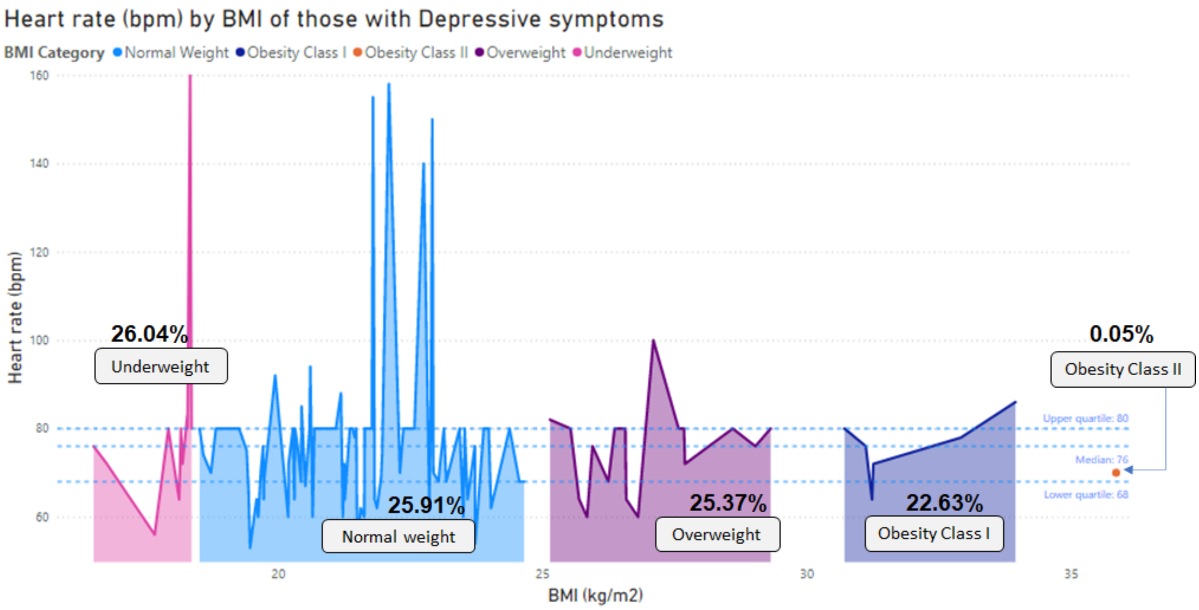
```
. tabulate Grant Depressivesymptoms,  column row key chi2 lrchi2 V

+-------------------+
| Key               |
|-------------------|
|     frequency     |
|  row percentage   |
| column percentage |
+-------------------+

           |  Depressive symptoms
     Grant |        no        yes |     Total
-----------+----------------------+----------
        no |       551         66 |       617 
           |     89.30      10.70 |    100.00 
           |     68.19      55.46 |     66.56 
-----------+----------------------+----------
       yes |       257         53 |       310 
           |     82.90      17.10 |    100.00 
           |     31.81      44.54 |     33.44 
-----------+----------------------+----------
     Total |       808        119 |       927 
           |     87.16      12.84 |    100.00 
           |    100.00     100.00 |    100.00 

          Pearson chi2(1) =   7.5528   Pr = 0.006
 likelihood-ratio chi2(1) =   7.2773   Pr = 0.007
               Cramér's V =   0.0903
```

3. Visualisations comparing the depressive symptoms (made using PowerBi)



Appendix 3a

BMI for those with depressive symptoms, ranges from 16.51 kg/m^2 to 35.86kg/m^2. According to the NHS, a healthy BMI range for adults is 18.5-24.9kg/m^2. Below 18.5kg/m^2 is the underweight category. We've focused on this group that falls under having depressive symptoms, in order to produce another graph, which looks at the BMI categorisations of those with depressive symptoms.



Appendix 3b

Appendix 3b, is filtered with only BMIs of 16.51-35.86 kg/m^2 and those with depressive symptoms only. It can be seen here that there is an even split in proportion between those that are underweight, normal weight, overweight and in obesity class I. The proportion falling in obesity class II is minuscule. This split shows that there is no direct link between the weight distribution and mental wellbeing of this particular set of French students. The heart rates of those in all BMI categories overall, edges above the upper quartile figure of 80 bpm. There is one massive outlier in the underweight category at 160 bpm and several outliers in the normal weight category, close to 160 bpm. NHS states on their website that the normal resting heart rate for an adult is between 60 and 100 bpm. Apart from the outliers mentioned, most of these figures are below 100 bpm. From this information alone, it is suggestive that the depressive issues that these students face, may have nothing to do with anxiety or panic attacks, but this is not conclusive.