

Testing Figure 1

2022-11-18

Working on our initial report, it became evident that the raw data from Cavanna et al. (2021), “Lifetime use of psychedelics is associated with better mental health indicators during the COVID-19 pandemic” would not be appropriate to use for all aspects of the Final Project for PSYC 201A. The authors of said paper applied some exclusion criteria to their data which we simply would not be able to replicate, and this became apparent as we first attempted to replicate their exclusion criteria and then to replicate some of their statistical findings. Faced with a difficult decision, we opted to replicate the results and perform novel analysis on a similar paper: “Psychedelic mushrooms in the USA: Knowledge, patterns of use, and association with health outcomes,” by Matzopoulos, Morlock, Morlock, Lerer & Lerer (2021). In this open dataset, we continued to find publication decisions on the part of authors which complicate complete replication, but feel they have provided sufficient data to allow for partial replication of some published results and novel analysis. We reviewed about ten papers in our search for a dataset for this project and learned a valuable lesson. We believe the challenges we’ve faced in finding an appropriate open dataset potentially reflect facets of an ongoing replication crisis in the fields of shared interest to us and highlight the importance of rigorous and open science as we move forward in our careers.

Matzopoulos, Morlock, Morlock, Lerer & Lerer (2021) examined the use of psychedelic mushrooms (PM) in association with measures of mental health status and outcomes and quality of life. They examined participants’ motivations for consumption of PM including desires for general mental health and well-being, as medication for medically diagnosed conditions, and as self-medication for undiagnosed conditions. The authors observed that users of PM reported more depression and anxiety as measured by the GAD-7 and PHQ-9, as well as several other factors predictive of PM use such as gender and not having health insurance. Examining a sample of 7,139 participants weighted to current estimates of the US population, the authors report that a significant number of US adults are already self-medicating with PM. Positive press coverage and “hype” indicate that this proportion of the population is likely to increase, and this necessitates further research into the association between PM use and poor mental health outcomes.

Link to paper and dataset:

Published paper permalink: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8761614/>

Dataset available via Zenodo: <https://zenodo.org/record/5791226#.Y3lGFuzMKrO>

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
```

```
## v tidyr 1.2.1 v stringr 1.4.1
## v readr 2.1.2 v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
```

Replication of Figure 1: Figure 1 represents a multivariate logistic regression model using predictive factors to predict use of PM. Predictive factors of PM use include being male, having a higher score on the Charlson Comorbidity Index, or living in the Western Census Region of the US. Self-report of PM use was less likely among participants who had health insurance, were relatively older, or who lived outside of the Western US Census region (i.e., in the Northeast, Midwest, or South). These relationships are reported as odds ratios (ORs) and 95% Confidence Intervals (95% CI; lower, upper).

Load data

```
PM = read.csv('data/external/AHRI_DATASET_PM_MANUSCRIPT_DATA.csv')
```

Clean data

```
# only necessary columns for figure 1's multivariate logistical regression
PM_cleaned = PM %>%
  select(
    CASEID_7139,
    SEX,
    AGE,
    ETHNICITY,
    HLS_YN,
    REGION,
    CCI_SCORE,
    GAD7_GE10,
    PHQ9_GE10,
    INSURANCE,
    PM1_GEN_HEALTH,
    PM1_DIAG_CONDITION,
    PM1_UNDIAG_CONCERN
  ) %>%
  mutate(
    PM_12M = PM1_GEN_HEALTH + PM1_UNDIAG_CONCERN + PM1_DIAG_CONDITION
  )

# calculate boolean for PM_12M
PM_cleaned$PM_12M = PM_cleaned$PM_12M %>%
  recode(`-297` = 0, `0` = 0, `1` = 1, `2` = 1, `3` = 1)
```

```
## Make additional columns for individual risk factors
PM_cleaned$BLACK = (PM_cleaned$ETHNICITY == 1)
```

```

PM_cleaned$WHITE = (PM_cleaned$ETHNICITY == 2)
PM_cleaned$OTHERETHNICITY = (PM_cleaned$ETHNICITY == 3)

PM_cleaned$NORTHWEST = (PM_cleaned$REGION == 1)
PM_cleaned$MIDWEST = (PM_cleaned$REGION == 2)
PM_cleaned$SOUTH = (PM_cleaned$REGION == 3)
PM_cleaned$WEST = (PM_cleaned$REGION == 4)

## refactor columns
PM_cleaned$SEX = PM_cleaned$SEX %>%
  recode_factor(., `0` = 'Female', `1` = 'Male')

PM_cleaned$ETHNICITY = PM_cleaned$ETHNICITY %>%
  recode_factor(., `1` = 'Other', `2` = 'White', `3` = 'Other')

PM_cleaned$HLS_YN = PM_cleaned$HLS_YN %>%
  recode_factor(., `0` = 'None-Hispanic', `1` = 'Hispanic')

PM_cleaned$REGION = PM_cleaned$REGION %>%
  recode_factor(., `4` = 'West', `1` = 'Northeast', `2` = 'Midwest', `3` = 'South')

```

Multivariate logistical regression

```

## modified helper from https://rdr.io/github/eringrand/RUncommon/src/R/logistic_regression_or_ci.R
logistic_regression_or_ci <- function(regress.out, level = 0.95) {
  usual.output <- summary(regress.out)
  z.quantile <- stats::qnorm(1 - (1 - level) / 2)
  number.vars <- length(regress.out$coefficients)
  OR <- exp(regress.out$coefficients[-1])
  temp.store.result <- matrix(rep(NA, number.vars * 2), nrow = number.vars)
  for (i in 1:number.vars) {
    temp.store.result[i, ] <- summary(regress.out)$coefficients[i] +
      c(-1, 1) * z.quantile * summary(regress.out)$coefficients[i + number.vars]
  }
  intercept.ci <- temp.store.result[1, ]
  slopes.ci <- temp.store.result[-1, ]
  OR.ci <- exp(slopes.ci)

  output <- list(
    regression.table = usual.output, intercept.ci = intercept.ci,
    slopes.ci = slopes.ci, OR = OR, OR.ci = OR.ci
  )
  return(output)
}

```

```

full_model = glm(PM_12M ~ SEX + AGE + ETHNICITY + REGION + CCI_SCORE + GAD7_GE10 + PHQ9_GE10 + INSURANCE_COVERED)
full_model_results = logistic_regression_or_ci(full_model)

```

```

sdf = full_model_results$regression.table$coefficients

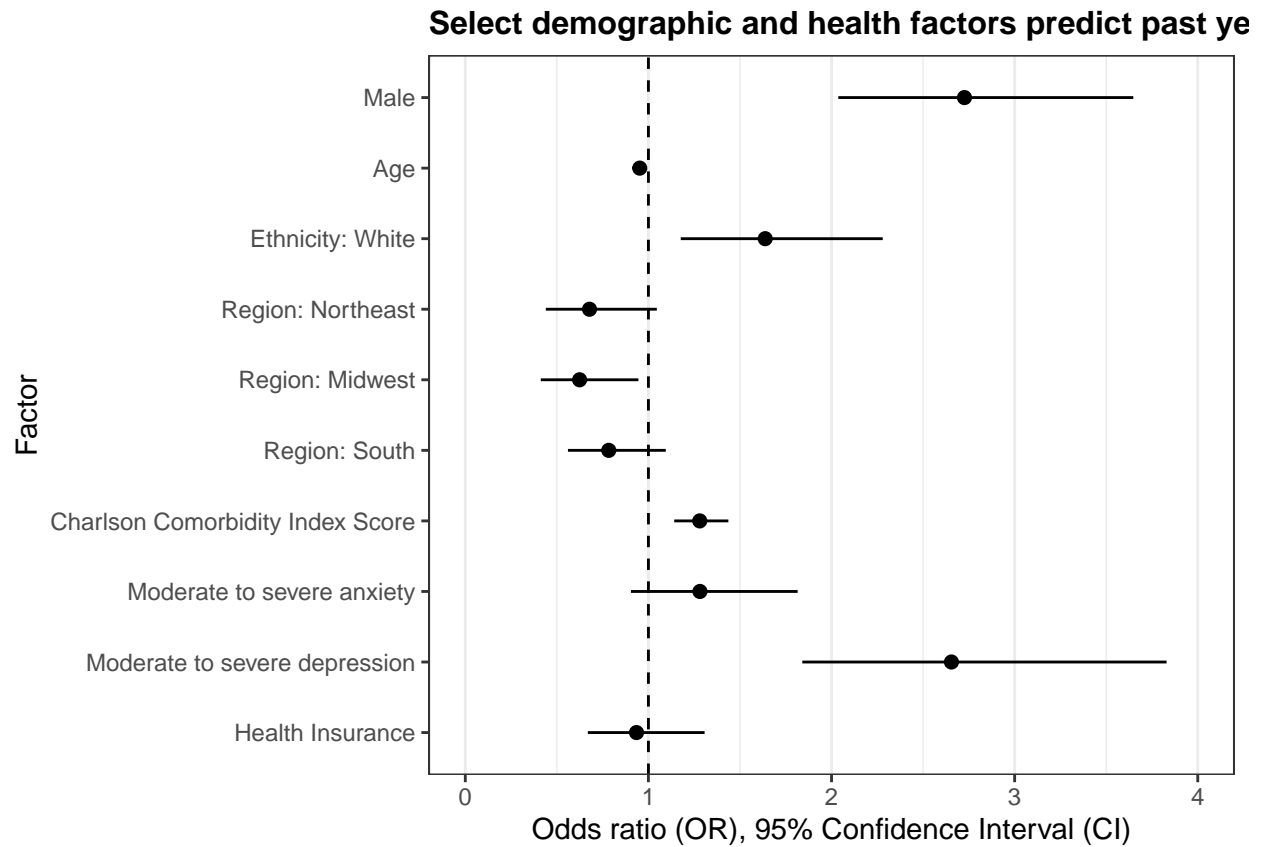
```

```
figure1df = data.frame(full_model_results$OR)
figure1df = cbind(Factor = rownames(figure1df), figure1df)
rownames(figure1df) = 1:nrow(figure1df)
```

```
figure1df$or.cimin = full_model_results$OR.ci[,1]
figure1df$or.cimax = full_model_results$OR.ci[,2]
```

```
figure1df$Factor = figure1df %>%
  pull(Factor) %>%
  recode_factor(
    `SEXMale` = "Male",
    `AGE` = "Age",
    `ETHNICITYWhite` = "Ethnicity: White",
    `REGIONNortheast` = "Region: Northeast",
    `REGIONMidwest` = "Region: Midwest",
    `REGIONSouth` = "Region: South",
    `CCI_SCORE` = "Charlson Comorbidity Index Score",
    `GAD7_GE10` = "Moderate to severe anxiety",
    `PHQ9_GE10` = "Moderate to severe depression",
    `INSURANCE` = "Health Insurance"
  )
```

```
## try plotting
ggplot(data = figure1df, aes(x = full_model_results.OR, y = Factor, xmin = or.cimin, xmax = or.cimax)) +
  geom_linerange() +
  geom_point(size = 2) +
  theme_bw() +
  scale_y_discrete(limits=rev) +
  geom_vline(aes(xintercept = 1), linetype = 'dashed') +
  labs(
    title = 'Select demographic and health factors predict past year psychedelic mushroom use',
    x = 'Odds ratio (OR), 95% Confidence Interval (CI)',
    color = 'OR'
  ) +
  xlim(0, 4) +
  theme(
    plot.title = element_text(size=12,face = "bold"),
    panel.grid.major.y = element_blank(),
    panel.grid.minor.y = element_blank()
  ) +
  guides(fill=guide_legend(title="New Legend Title"))
```



```
ggsave(filename = 'f1rep.pdf', width = 9, height = 4)
```

The plot for our replication of Matzopoulos, Morlock, Morlock, Lerer & Lerer (2021)'s multivariate logistical regression resembles the original

For novel analysis, we elect to