Analyses for Psychedelic Mushroom Dataset

Jeffrey Xing and Liam Conaboy

2022-12-03

Notes for Pivoting Dataset

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.

Dataset Abstract

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

Data Cleaning

Load necessary libraries

```r
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()
```

Load data

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

Select necessary columns

```r
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,
    PSY1_POSITIVE_USE,
    PM_USE_ONLY_YN,
    PM_VS_PSY_YN,
    DATA_WEIGHT
    )
```

Calculate Past-year PM use

```r
## pool all past-year PM columns together
PM_cleaned = PM_cleaned %>%
```

```r
  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)
```

Refactor Predictor Variables

```r
## 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')
```

Hypothesis 1:

$PM = MentalHealthStatus + Error$

Multivariate Logistical Regression

```r
## modified helper from
## https://rdrr.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)
}
```

Add helper

```
empty_model = glm(PM_12M ~ NULL, data = PM_cleaned, family = binomial)
empty_model_results = logistic.regression.or.ci(empty_model)
empty_model_results
```

Empty Model

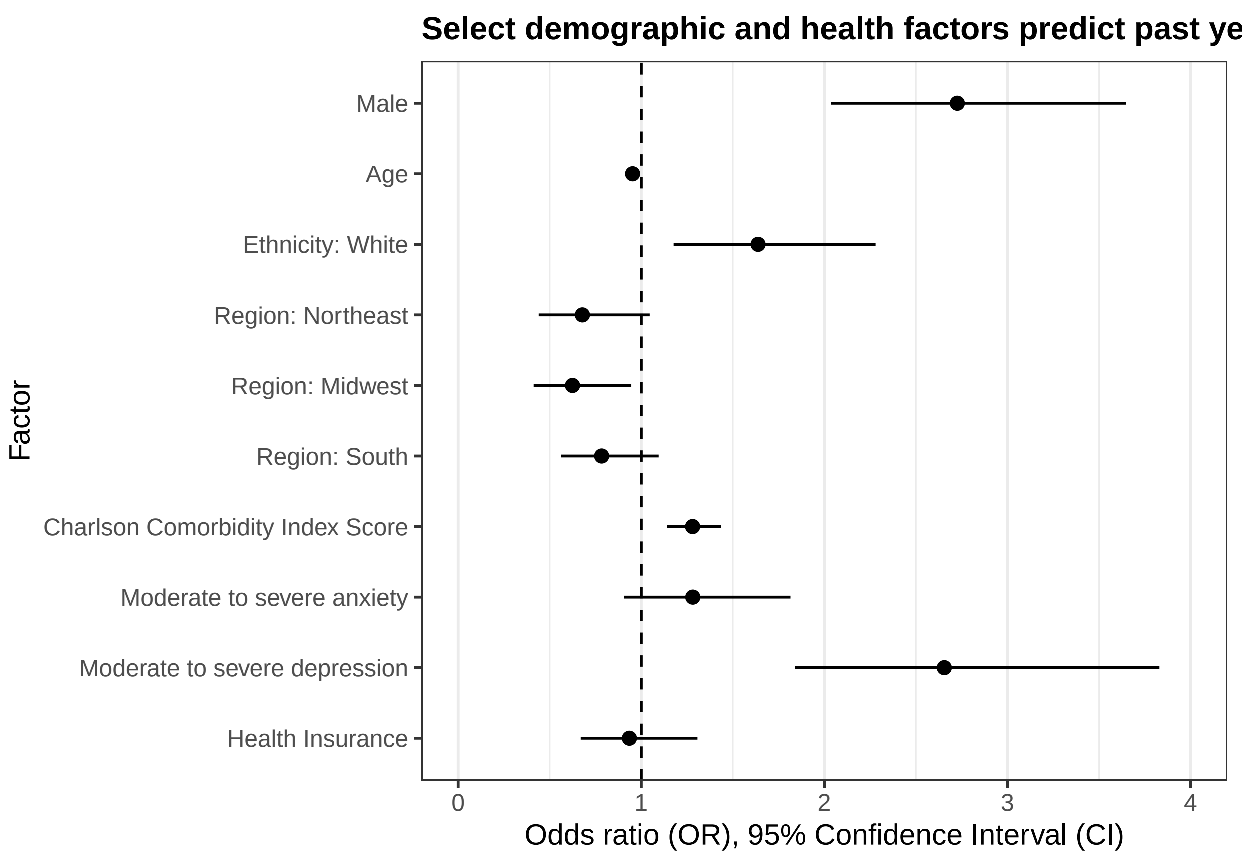
```
## $regression.table
##
## Call:
## glm(formula = PM_12M ~ NULL, family = binomial, data = PM_cleaned)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.2548  -0.2548  -0.2548  -0.2548   2.6245
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.41152    0.06731  -50.68   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2019.1  on 7138  degrees of freedom
## Residual deviance: 2019.1  on 7138  degrees of freedom
## AIC: 2021.1
##
## Number of Fisher Scoring iterations: 6
##
##
## $intercept.ci
## [1] -3.543449 -3.279599
##
## $slopes.ci
##      [,1] [,2]
##
## $OR
## named numeric(0)
##
## $OR.ci
##      [,1] [,2]
```

```r
full_model = glm(PM_12M ~ SEX + AGE + ETHNICITY + REGION + CCI_SCORE + GAD7_GE10 + PHQ9_GE10 + INSURANC
full_model_results = logistic.regression.or.ci(full_model)
full_model_results
```

Full Model

```
## $regression.table
##
## Call:
## glm(formula = PM_12M ~ SEX + AGE + ETHNICITY + REGION + CCI_SCORE +
##     GAD7_GE10 + PHQ9_GE10 + INSURANCE, family = binomial, data = PM_cleaned)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9536  -0.2812  -0.1814  -0.1113   3.3346
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.793904   0.337698  -8.273  < 2e-16 ***
## SEXMale           1.002818   0.148650   6.746 1.52e-11 ***
## AGE              -0.048536   0.006197  -7.832 4.79e-15 ***
## ETHNICITYWhite    0.493329   0.168726   2.924  0.00346 **
## REGIONNortheast -0.388157   0.221003  -1.756  0.07903 .
## REGIONMidwest   -0.471080   0.211757  -2.225  0.02611 *
## REGIONSouth     -0.243976   0.170793  -1.428  0.15315
## CCI_SCORE        0.247087   0.058807   4.202 2.65e-05 ***
## GAD7_GE10        0.247902   0.177679   1.395  0.16295
## PHQ9_GE10        0.976374   0.186931   5.223 1.76e-07 ***
## INSURANCE       -0.067125   0.170770  -0.393  0.69427
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2019.1  on 7138  degrees of freedom
## Residual deviance: 1766.5  on 7128  degrees of freedom
## AIC: 1788.5
##
## Number of Fisher Scoring iterations: 7
##
##
## $intercept.ci
## [1] -3.455780 -2.132027
##
## $slopes.ci
##              [,1]        [,2]
##  [1,]  0.71146923  1.29416691
##  [2,] -0.06068113 -0.03638991
##  [3,]  0.16263260  0.82402449
##  [4,] -0.82131482  0.04500019
##  [5,] -0.88611659 -0.05604350
##  [6,] -0.57872348  0.09077187
```

```
##  [7,]  0.13182829  0.36234554
##  [8,] -0.10034179  0.59614668
##  [9,]  0.60999473  1.34275255
## [10,] -0.40182830  0.26757920
##
## $OR
##         SEXMale             AGE  ETHNICITYWhite REGIONNortheast   REGIONMidwest
##       2.7259529       0.9526235       1.6377585       0.6783056       0.6243276
##     REGIONSouth       CCI_SCORE       GAD7_GE10       PHQ9_GE10       INSURANCE
##       0.7835066       1.2802904       1.2813349       2.6548115       0.9350787
##
## $OR.ci
##            [,1]      [,2]
##  [1,] 2.0369818 3.6479556
##  [2,] 0.9411233 0.9642642
##  [3,] 1.1766043 2.2796559
##  [4,] 0.4398529 1.0460281
##  [5,] 0.4122536 0.9454980
##  [6,] 0.5606135 1.0950192
##  [7,] 1.1409124 1.4366953
##  [8,] 0.9045282 1.8151111
##  [9,] 1.8404217 3.8295701
## [10,] 0.6690956 1.3067971
```



Visualize

Hypothesis 2:

Positive PM Perception = PM Use + Error

Distinguish PM-Only users and Non-psychedelic users

```
PM_cleaned_PMvsNonPsy = PM_cleaned %>%
  ## leave only pm use only and non-psyc
  filter(PM_USE_ONLY_YN == 1 | PM_VS_PSY_YN == -99)

## make PM_USE_ONLY_YN factor
PM_cleaned_PMvsNonPsy$PM_USE_ONLY_YN = PM_cleaned_PMvsNonPsy$PM_USE_ONLY_YN %>%
  recode_factor(`0` = 'NONE', `1` = 'PM ONLY')
```

One-way ANOVA

```
PMUser_model = lm(PSY1_POSITIVE_USE ~ PM_USE_ONLY_YN, data = PM_cleaned_PMvsNonPsy, weights = DATA_WEIG
```

```
summary(PMUser_model)
```

Summary

```
##
## Call:
## lm(formula = PSY1_POSITIVE_USE ~ PM_USE_ONLY_YN, data = PM_cleaned_PMvsNonPsy,
##     weights = DATA_WEIGHT)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -969.90 -118.06  -70.02  234.81  604.22
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.56349    0.01471  242.27   <2e-16 ***
## PM_USE_ONLY_YNPM ONLY -1.50572    0.11168  -13.48   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 224.7 on 6733 degrees of freedom
## Multiple R-squared:  0.02629,    Adjusted R-squared:  0.02615
## F-statistic: 181.8 on 1 and 6733 DF,  p-value: < 2.2e-16
```

```
anova(PMUser_model)
```

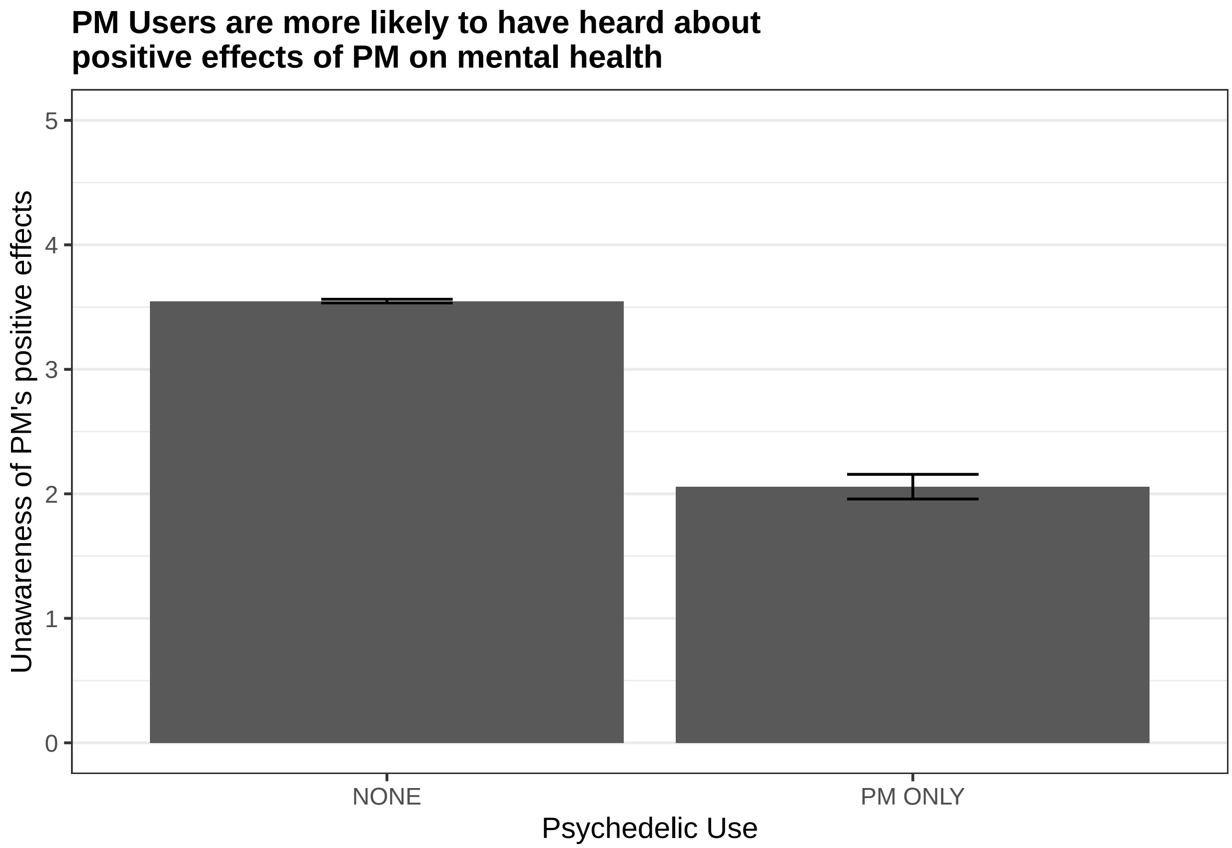
ANOVA

```
## Analysis of Variance Table
##
## Response: PSY1_POSITIVE_USE
##                   Df    Sum Sq Mean Sq F value    Pr(>F)
## PM_USE_ONLY_YN    1   9181279 9181279  181.79 < 2.2e-16 ***
## Residuals      6733 340043444   50504
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plot

```
PM_cleaned_PMvsNonPsy %>%
  ggplot(mapping = aes(x= PM_USE_ONLY_YN, y = PSY1_POSITIVE_USE, weight = DATA_WEIGHT)) +
  stat_summary(geom = 'bar') +
  stat_summary(geom = 'errorbar', width = 0.25) +
  theme_bw() +
  theme(
    plot.title = element_text(size=12,face = "bold"),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank()
  )+
  labs(
    title = 'PM Users are more likely to have heard about \npositive effects of PM on mental health',
    x = 'Psychedelic Use',
    y = "Unawareness of PM's positive effects",
    color = 'OR'
  ) +
  ylim(0, 5)
```

```
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
```




```
ggsave(filename = 'f2rep.pdf', width = 4, height = 4, dpi = 300)
```

```
## No summary function supplied, defaulting to `mean_se()`
## No summary function supplied, defaulting to `mean_se()`
```

```
knowledge_model = glm(PM_12M ~ PSY1_POSITIVE_USE, data = PM_cleaned, family = binomial)
knowledge_model_results = logistic.regression.or.ci(knowledge_model)
knowledge_model_results
```

Logistic Regression

```
## $regression.table
##
## Call:
## glm(formula = PM_12M ~ PSY1_POSITIVE_USE, family = binomial,
##     data = PM_cleaned)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6370  -0.2020  -0.2020  -0.0614   3.5423
##
## Coefficients:
```

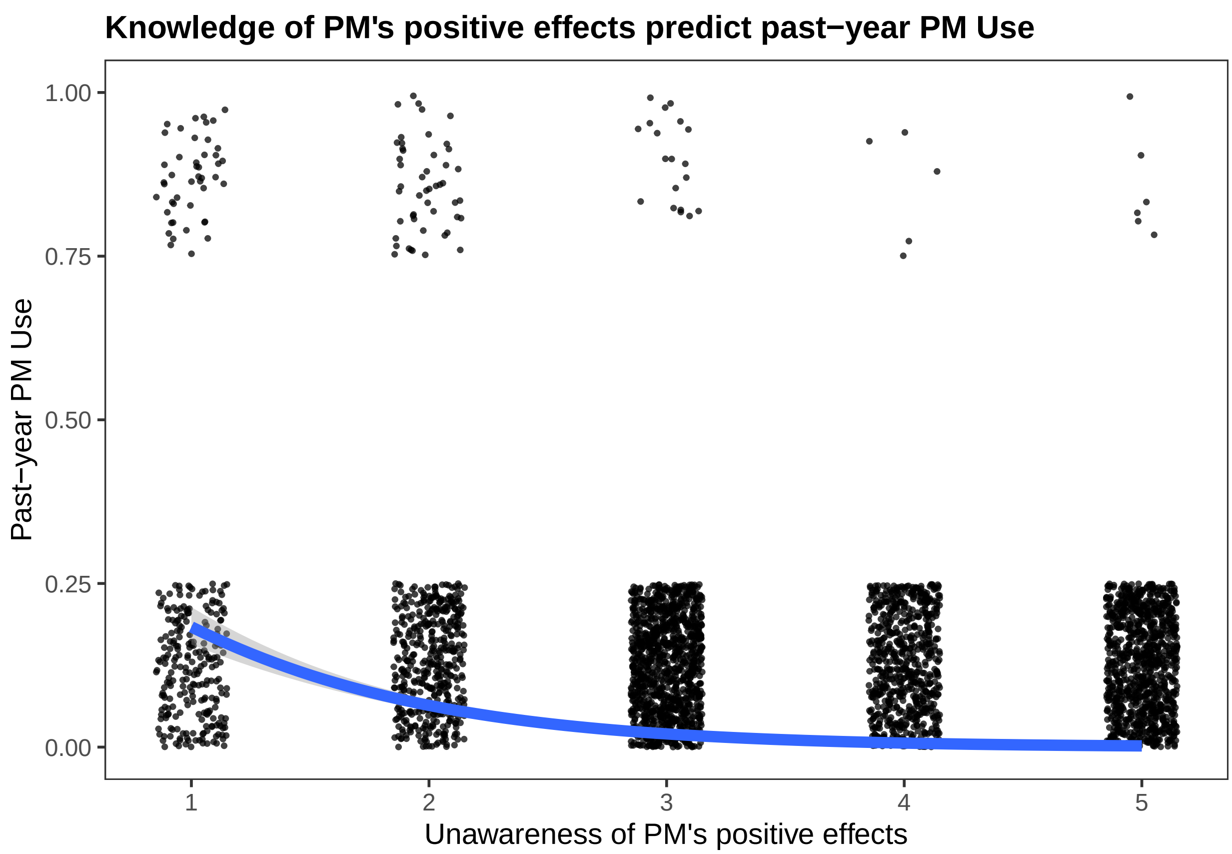
```
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.29690    0.15522  -1.913   0.0558 .
## PSY1_POSITIVE_USE -1.19506    0.07168 -16.672   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2019.1  on 7138  degrees of freedom
## Residual deviance: 1632.1  on 7137  degrees of freedom
## AIC: 1636.1
##
## Number of Fisher Scoring iterations: 7
##
##
## $intercept.ci
## [1] -0.60112988  0.00733167
##
## $slopes.ci
## [1] -1.335546 -1.054567
##
## $OR
## PSY1_POSITIVE_USE
##         0.3026868
##
## $OR.ci
## [1] 0.2630145 0.3483431
```

Plot Logistic Regression

```
PM_cleaned %>%
  ggplot(mapping = aes(x = PSY1_POSITIVE_USE, y = PM_12M)) +
  geom_jitter(size = 0.5, height = 0.25, width = 0.15, alpha = 0.75) +
  stat_smooth(method="glm", method.args = list(family=binomial), size = 2) +
  labs(
    title = "Knowledge of PM's positive effects predict past-year PM Use",
    x = "Unawareness of PM's positive effects",
    y = 'Past-year PM Use'
  ) +
  theme_bw() +
  theme(
    plot.title = element_text(size=12,face = "bold"),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank(),
    panel.grid.major.y = element_blank(),
    panel.grid.minor.y = element_blank()
  ) +
  ylim(0, 1)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 3521 rows containing missing values (geom_point).
```




```
ggsave(filename = 'f3rep.pdf', width = 6, height = 4, dpi = 300)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 3525 rows containing missing values (geom_point).
```

Hypothesis 3

Interaction models

```
health_interaction_model = glm(PM_12M ~ SEX + AGE + ETHNICITY + REGION + CCI_SCORE + (GAD7_GE10 * PHQ9_G
health_interaction_model_results = logistic.regression.or.ci(health_interaction_model)
health_interaction_model_results
```

```
## $regression.table
##
## Call:
## glm(formula = PM_12M ~ SEX + AGE + ETHNICITY + REGION + CCI_SCORE +
##     (GAD7_GE10 * PHQ9_GE10) + INSURANCE, family = binomial, data = PM_cleaned)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9590  -0.2807  -0.1817  -0.1121   3.3276
##
```

```
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -2.760409   0.339853  -8.122 4.57e-16 ***
## SEXMale             1.002961   0.148697   6.745 1.53e-11 ***
## AGE                -0.048733   0.006199  -7.861 3.82e-15 ***
## ETHNICITYWhite      0.492934   0.168736   2.921  0.00349 **
## REGIONNortheast    -0.389071   0.221040  -1.760  0.07838 .
## REGIONMidwest      -0.468985   0.211785  -2.214  0.02680 *
## REGIONSouth        -0.244175   0.170807  -1.430  0.15285
## CCI_SCORE           0.246615   0.058801   4.194 2.74e-05 ***
## GAD7_GE10          -0.041356   0.432526  -0.096  0.92383
## PHQ9_GE10           0.897464   0.215967   4.156 3.24e-05 ***
## INSURANCE          -0.067412   0.170830  -0.395  0.69313
## GAD7_GE10:PHQ9_GE10  0.359119   0.477278   0.752  0.45179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2019.1  on 7138  degrees of freedom
## Residual deviance: 1765.9  on 7127  degrees of freedom
## AIC: 1789.9
##
## Number of Fisher Scoring iterations: 7
##
##
## $intercept.ci
## [1] -3.426509 -2.094309
##
## $slopes.ci
##              [,1]        [,2]
##  [1,]  0.71152085  1.29440059
##  [2,] -0.06088395 -0.03658245
##  [3,]  0.16221671  0.82365163
##  [4,] -0.82230154  0.04415908
##  [5,] -0.88407568 -0.05389335
##  [6,] -0.57895130  0.09060177
##  [7,]  0.13136622  0.36186324
##  [8,] -0.88909086  0.80637935
##  [9,]  0.47417603  1.32075283
## [10,] -0.40223292  0.26740934
## [11,] -0.57632873  1.29456580
##
## $OR
##             SEXMale                 AGE      ETHNICITYWhite     REGIONNortheast
##           2.7263418           0.9524352           1.6371127           0.6776860
##       REGIONMidwest         REGIONSouth           CCI_SCORE           GAD7_GE10
##           0.6256373           0.7833507           1.2796860           0.9594877
##           PHQ9_GE10           INSURANCE GAD7_GE10:PHQ9_GE10
##           2.4533745           0.9348102           1.4320665
##
## $OR.ci
##            [,1]      [,2]
##  [1,] 2.0370870 3.6488082
```

```
##  [2,] 0.9409324 0.9640786
##  [3,] 1.1761151 2.2788060
##  [4,] 0.4394191 1.0451486
##  [5,] 0.4130958 0.9475332
##  [6,] 0.5604858 1.0948329
##  [7,] 1.1403853 1.4360025
##  [8,] 0.4110293 2.2397838
##  [9,] 1.6066898 3.7462406
## [10,] 0.6688249 1.3065752
## [11,] 0.5619577 3.6494110
```

```
demographic_interaction_model = glm(PM_12M ~ (SEX * AGE) + ETHNICITY + REGION + CCI_SCORE + GAD7_GE10 +
demographic_interaction_model_results = logistic.regression.or.ci(demographic_interaction_model)
demographic_interaction_model_results
```

```
## $regression.table
##
## Call:
## glm(formula = PM_12M ~ (SEX * AGE) + ETHNICITY + REGION + CCI_SCORE +
##     GAD7_GE10 + PHQ9_GE10 + INSURANCE, family = binomial, data = PM_cleaned)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9288  -0.2803  -0.1845  -0.1013   3.4878
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -2.03139    0.43453  -4.675 2.94e-06 ***
## SEXMale         -0.18127    0.44873  -0.404  0.68624
## AGE             -0.07120    0.01092  -6.518 7.13e-11 ***
## ETHNICITYWhite   0.49491    0.16820   2.942  0.00326 **
## REGIONNortheast -0.38836    0.22080  -1.759  0.07860 .
## REGIONMidwest   -0.47039    0.21151  -2.224  0.02615 *
## REGIONSouth     -0.24730    0.17040  -1.451  0.14670
## CCI_SCORE        0.24253    0.05895   4.114 3.88e-05 ***
## GAD7_GE10        0.24961    0.17758   1.406  0.15983
## PHQ9_GE10        0.97993    0.18711   5.237 1.63e-07 ***
## INSURANCE       -0.07609    0.17046  -0.446  0.65535
## SEXMale:AGE      0.03411    0.01255   2.719  0.00655 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2019.1  on 7138  degrees of freedom
## Residual deviance: 1758.6  on 7127  degrees of freedom
## AIC: 1782.6
##
## Number of Fisher Scoring iterations: 8
##
##
## $intercept.ci
## [1] -2.883062 -1.179727
##
```

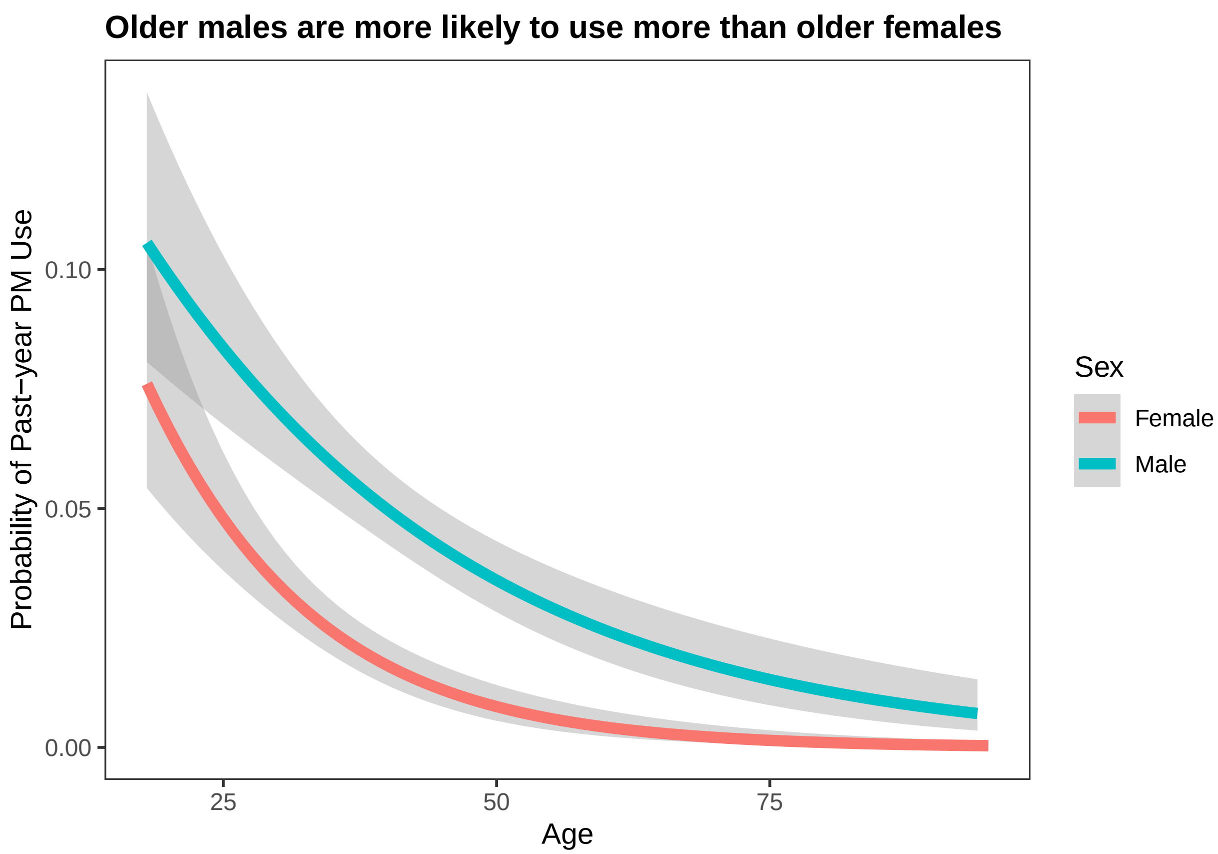
```
## $slopes.ci
##                [,1]        [,2]
##  [1,] -1.060775609  0.69822817
##  [2,] -0.092611921 -0.04979079
##  [3,]  0.165249280  0.82457783
##  [4,] -0.821114815  0.04439892
##  [5,] -0.884939029 -0.05584611
##  [6,] -0.581284096  0.08668100
##  [7,]  0.126994319  0.35807236
##  [8,] -0.098432765  0.59765677
##  [9,]  0.613201630  1.34666400
## [10,] -0.410188419  0.25801681
## [11,]  0.009523053  0.05870249
##
## $OR
##         SEXMale            AGE  ETHNICITYWhite REGIONNortheast   REGIONMidwest
##       0.8342070      0.9312744       1.6403564       0.6781696       0.6247570
##     REGIONSouth      CCI_SCORE       GAD7_GE10       PHQ9_GE10       INSURANCE
##       0.7809052      1.2744737       1.2835273       2.6642772       0.9267367
##     SEXMale:AGE
##       1.0347013
##
## $OR.ci
##            [,1]      [,2]
##  [1,] 0.3461872 2.0101878
##  [2,] 0.9115472 0.9514285
##  [3,] 1.1796872 2.2809176
##  [4,] 0.4399409 1.0453993
##  [5,] 0.4127393 0.9456847
##  [6,] 0.5591799 1.0905487
##  [7,] 1.1354106 1.4305691
##  [8,] 0.9062566 1.8178542
##  [9,] 1.8463332 3.8445786
## [10,] 0.6635252 1.2943606
## [11,] 1.0095685 1.0604597
```

Plot Demographic Interaction

```
PM_cleaned %>%
  ggplot(mapping = aes(x = AGE, y = PM_12M, color = SEX)) +
  #geom_jitter(size = 0.5, height = 0.25, width = 0.15, alpha = 0.75) +
  stat_smooth(method="glm", method.args = list(family=binomial), size = 2) +
  labs(
    title = "Older males are more likely to use more than older females",
    x = "Age",
    y = 'Probability of Past-year PM Use',
    color = "Sex"
  ) +
  theme_bw() +
  theme(
    plot.title = element_text(size=12,face = "bold"),
    panel.grid.major.x = element_blank(),
```

```
    panel.grid.minor.x = element_blank(),
    panel.grid.major.y = element_blank(),
    panel.grid.minor.y = element_blank()
  )
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggsave(filename = 'f4nov.pdf', width = 6, height = 4, dpi = 300)
```

```
## `geom_smooth()` using formula 'y ~ x'
```