Project 2 Codebook

Yingxi Kong

Table 1: Participant Characteristics by Treatment Arm

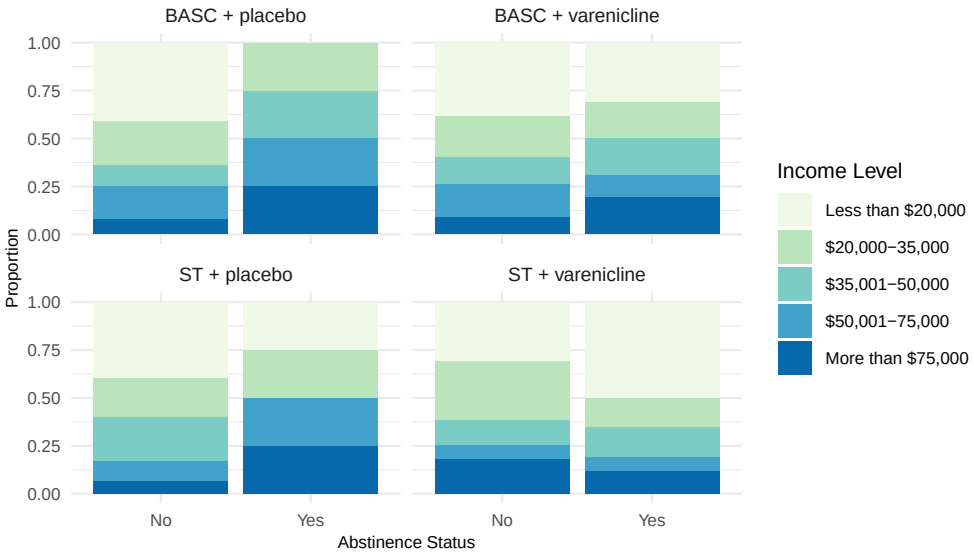
Characteristic	Behavioral and Pharmacological Treatment Assignment				
	BASC + placebo, N = 68	**BASC + varenicline, N = 83**	**ST + placebo, N = 68**	**ST + varenicline, N = 81**	**Overall, N = 300**
Smoking abstinence	4 (5.9%)	26 (31%)	8 (12%)	26 (32%)	64 (21%)
Age	51 (14)	50 (13)	50 (11)	49 (13)	50 (13)
Sex					
Male	30 (44%)	39 (47%)	29 (43%)	37 (46%)	135 (45%)
Female	38 (56%)	44 (53%)	39 (57%)	44 (54%)	165 (55%)
Income					
Less than $20,000	26 (38%)	30 (36%)	26 (38%)	30 (37%)	112 (37%)
$20,000-35,000	16 (24%)	17 (20%)	14 (21%)	21 (26%)	68 (23%)
$35,001-50,000	8 (12%)	13 (16%)	14 (21%)	11 (14%)	46 (15%)
$50,001-75,000	12 (18%)	13 (16%)	8 (12%)	6 (7.4%)	39 (13%)
More than $75,000	6 (8.8%)	10 (12%)	6 (8.8%)	13 (16%)	35 (12%)
Education					
Grade School	1 (1.5%)	0 (0%)	0 (0%)	0 (0%)	1 (0.3%)
Some high school	3 (4.4%)	7 (8.4%)	2 (2.9%)	4 (4.9%)	16 (5.3%)
High school graduate or GED	23 (34%)	15 (18%)	11 (16%)	27 (33%)	76 (25%)
Some college/technical school	22 (32%)	32 (39%)	38 (56%)	24 (30%)	116 (39%)
College graduate	19 (28%)	29 (35%)	17 (25%)	26 (32%)	91 (30%)
FTCD score	5.31 (2.02)	5.07 (2.34)	5.42 (2.09)	5.17 (2.08)	5.23 (2.14)
Smoking within 5 mins of waking up	32 (47%)	33 (40%)	35 (51%)	38 (47%)	138 (46%)
BDI score	19 (12)	18 (11)	18 (11)	20 (12)	19 (11)
Cigarettes smoked per day	16 (9)	16 (9)	15 (7)	14 (7)	15 (8)
Cigarette reward value	7.4 (3.8)	7.3 (3.9)	6.9 (3.6)	7.0 (3.4)	7.2 (3.6)
Pleasurable events (substitute reinforcers)	23 (20)	23 (19)	21 (20)	23 (19)	23 (20)
Pleasurable events (complementary reinforcers)	28 (22)	22 (17)	27 (20)	25 (19)	25 (19)
Anhedonia	2.36 (3.41)	2.25 (3.12)	2.51 (3.35)	2.11 (3.00)	2.30 (3.20)
Other lifetime DSM-5 diagnosis	35 (51%)	30 (36%)	28 (41%)	40 (49%)	133 (44%)
Taking antidepressant	28 (41%)	24 (29%)	15 (22%)	15 (19%)	82 (27%)
Current vs. past MDD					

Table 1: Participant Characteristics by Treatment Arm *(continued)*

Characteristic	Behavioral and Pharmacological Treatment Assignment				
	BASC + placebo, N = 68	**BASC + varenicline, N = 83**	**ST + placebo, N = 68**	**ST + varenicline, N = 81**	**Overall, N = 300**
Past MDD	36 (53%)	43 (52%)	37 (54%)	37 (46%)	153 (51%)
Current MDD	32 (47%)	40 (48%)	31 (46%)	44 (54%)	147 (49%)
Nicotine metabolism ratio	0.35 (0.17)	0.38 (0.24)	0.37 (0.26)	0.35 (0.20)	0.36 (0.22)
Exclusive mentholated cigarette user	40 (59%)	49 (59%)	43 (63%)	47 (58%)	179 (60%)
Readiness to quit smoking	6.77 (1.33)	6.67 (1.16)	6.96 (1.30)	6.74 (1.09)	6.78 (1.21)
Race					
Black	37 (54%)	37 (45%)	40 (59%)	43 (53%)	157 (52%)
Hispanic	4 (5.9%)	3 (3.6%)	4 (5.9%)	5 (6.2%)	16 (5.3%)
Non-Hispanic White	24 (35%)	34 (41%)	22 (32%)	25 (31%)	105 (35%)
Other	3 (4.4%)	9 (11%)	2 (2.9%)	8 (9.9%)	22 (7.3%)

[1] Mean (SD) for continuous; n (%) for categorical

Figure 1: Income Levels by Abstinence Status and Treatment Group



Appendix

```r
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)

# load necessary packages
library(tidyverse)
library(mice)
library(gt)
library(gtsummary)
library(kableExtra)
library(RColorBrewer)
library(scico)
# set working directory
setwd("C:/Users/yingx/OneDrive/Desktop/Fall 2024/PHP 2550/Data/")

# read in data
data <- read.csv("project2.csv")

data[, c("abst", "Var", "BA", "sex_ps", "NHW",
         "Black", "Hisp", "inc", "edu", "ftcd.5.mins",
         "otherdiag", "antidepmed", "mde_curr",
         "Only.Menthol")] <- lapply(data[, c("abst", "Var", "BA", "sex_ps", "NHW",
                                             "Black", "Hisp", "inc", "edu",
                                             "ftcd.5.mins", "otherdiag", "antidepmed",
                                             "mde_curr", "Only.Menthol")], as.factor)
# multiple imputation with m = 5
imputed_data <- mice(data, m = 5, method = 'pmm', maxit = 50, seed = 2550, printFlag = FALSE)

# extract the five imputed datasets
completed_datasets <- list()
for (i in 1:5) {
  completed_datasets[[i]] <- complete(imputed_data, i)
}

# calculate average/mode of each missing variable
averaged_data <- completed_datasets[[1]]

for (var in names(averaged_data)) {
  if (any(is.na(data[[var]]))) {
    if (is.numeric(averaged_data[[var]])) {
      averaged_data[[var]] <- rowMeans(sapply(completed_datasets, function(x) x[[var]]))
    } else {
      averaged_data[[var]] <- apply(sapply(completed_datasets, function(x) x[[var]]), 1, function(vals)
        vals <- as.factor(vals)
        unique_vals <- unique(vals)
        unique_vals[which.max(tabulate(match(vals, unique_vals)))]
      })
    }
  }
}
# Recode factor levels in the dataset
averaged_data_factor <- averaged_data %>%
  mutate(abst = fct_recode(as.factor(abst), "Yes" = "1", "No" = "0"),
```

```
        inc = fct_recode(as.factor(inc),
                         "Less than $20,000" = "1",
                         "$20,000-35,000" = "2",
                         "$35,001-50,000" = "3",
                         "$50,001-75,000" = "4",
                         "More than $75,000" = "5"),
        sex_ps = fct_recode(as.factor(sex_ps), "Male" = "1", "Female" = "2"),
        edu = fct_recode(as.factor(edu),
                         "Grade School" = "1",
                         "Some high school" = "2",
                         "High school graduate or GED" = "3",
                         "Some college/technical school" = "4",
                         "College graduate" = "5"),
        ftcd.5.mins = fct_recode(as.factor(ftcd.5.mins), "Yes" = "1", "No" = "0"),
        otherdiag = fct_recode(as.factor(otherdiag), "Yes" = "1", "No" = "0"),
        antidepmed = fct_recode(as.factor(antidepmed), "Yes" = "1", "No" = "0"),
        mde_curr = fct_recode(as.factor(mde_curr), "Current MDD" = "1", "Past MDD" = "0"),
        Only.Menthol = fct_recode(as.factor(Only.Menthol), "Yes" = "1", "No" = "0"),
        race = case_when(Black == 1 ~ "Black",
                         Hisp == 1 ~ "Hispanic",
                         NHW == 1 ~ "Non-Hispanic White",
                         TRUE ~ "Other"),
        trt = case_when(Var == 1 & BA == 1 ~ "BASC + varenicline",
                        Var == 0 & BA == 1 ~ "BASC + placebo",
                        Var == 1 & BA == 0 ~ "ST + varenicline",
                        Var == 0 & BA == 0 ~ "ST + placebo",
                        TRUE ~ NA_character_))

averaged_data_factor <- averaged_data_factor %>%
  mutate(inc = fct_relevel(inc, "Less than $20,000", "$20,000-35,000",
                           "$35,001-50,000", "$50,001-75,000", "More than $75,000"),
         edu = fct_relevel(edu, "Grade School", "Some high school", "High school graduate or GED",
                           "Some college/technical school", "College graduate"))

# Now create the summary table
summary_table <- averaged_data_factor %>%
  select(-c("id", "Var", "BA", "Black", "Hisp", "NHW")) %>%
  tbl_summary(by = trt, label = list(abst ~ "Smoking abstinence",
                                     race ~ "Race",
                                     age_ps ~ "Age",
                                     sex_ps ~ "Sex",
                                     inc ~ "Income",
                                     edu ~ "Education",
                                     ftcd_score ~ "FTCD score",
                                     ftcd.5.mins ~ "Smoking within 5 mins of waking up",
                                     bdi_score_w00 ~ "BDI score",
                                     cpd_ps ~ "Cigarettes smoked per day",
                                     crv_total_pq1 ~ "Cigarette reward value",
                                     hedonsum_n_pq1 ~ "Pleasurable events (substitute reinforcers)",
                                     hedonsum_y_pq1 ~ "Pleasurable events (complementary reinforcers)",
                                     shaps_score_pq1 ~ "Anhedonia",
                                     otherdiag ~ "Other lifetime DSM-5 diagnosis",
                                     antidepmed ~ "Taking antidepressant",
```

```
                                    mde_curr ~ "Current vs. past MDD",
                                    NMR ~ "Nicotine metabolism ratio",
                                    Only.Menthol ~ "Exclusive mentholated cigarette user",
                                    readiness ~ "Readiness to quit smoking"),
              statistic = all_continuous() ~ "{mean} ({sd})",
              missing = "ifany",
              missing_text = "Missing") %>%
  add_overall(last = TRUE) %>%
  modify_spanning_header(update = all_stat_cols() ~ "**Behavioral and Pharmacological Treatment Assignme
  modify_footnote(update = all_stat_cols() ~ "Mean (SD) for continuous; n (%) for categorical") %>%
  bold_labels()

summary_table %>%
  as_kable_extra(booktabs = TRUE, caption = "Participant Characteristics by Treatment Arm",
                 longtable = TRUE, linesep = "") %>%
  kableExtra::kable_styling(font_size = 9,
                            latex_options = c("repeat_header", "HOLD_position", "scale_down"))%>%
  column_spec(1, width = "3.5cm") %>%
  column_spec(2, width = "2.5cm") %>%
  column_spec(3, width = "2.5cm") %>%
  column_spec(4, width = "2.5cm") %>%
  column_spec(5, width = "2.5cm") %>%
  column_spec(6, width = "2.5cm") %>%
  row_spec(0, bold = TRUE, font_size = 9)
ggplot(averaged_data_factor, aes(x = abst, fill = inc)) +
  geom_bar(position = "fill") +
  facet_wrap(~ trt) +
  labs(title = "Figure 1: Income Levels by Abstinence Status and Treatment Group",
       x = "Abstinence Status",
       y = "Proportion",
       fill = "Income Level") +
  theme_minimal() +
  scale_fill_brewer(palette = "GnBu") +
  theme(axis.title = element_text(size = 8),
        title = element_text(size = 10),
        axis.text = element_text(size = 8),
        legend.text = element_text(size = 8))
```