

Telomere Exploration

Hayden Mountcastle and Andrew Ratanatharathorn

Feb. 25, 2025

Contents

1	Contents	1
2	Data	1
2.1	Finding Duplicates	3
2.2	Rosner Batch Correction 3-29-24	6
3	Stratified analysis	50
3.1	By batch	50
3.2	Models (all sets) with K10 + antipsychotics	56
3.3	Stratified by sex	63

1 Contents

Purpose:

This is an exploratory analysis of 971 participants, 487 cases with Schizophrenia and 484 controls, with telomere and phenotype data from NGAP Psychosis. The original sample has 1000 participants, evenly split between cases and controls, but 29 participants were not included in the analysis due to insufficient DNA or failed samples, leaving a total of n=971 participants.

Here, we explore associations between Telomere length and Schizophrenia, age, and sex.

We also explore other associations on available phenotypic data.

Methods:

We viewed association between phenotype variables with sufficient prevalence with log telomere length (*ttl*). We used linear regression to view associations between said variables and *ttl* at first without adjusting for age, and then adjusting for age. Additionally, we ran an ANOVA test with models with and without interaction terms to determine whether there was any functional difference between these models.

Results

In the unadjusted model, case status (*is_case*), sex (*msex*), and educational attainment (*educ_ord*) were associated with *ttl* (p<0.05).

Of particular note, congruent with previous literature, men on average had shorter telomere length both adjusting and not adjusting for age (T-test, p = 0.0018).

In the adjusted linear regression model, controlling for age and sex, schizophrenia (*is_case*=1) is associated with LTL ($p=0.016$)

UPDATE 4-30-24

On 4-10-24 we received data that brought to attention certain duplicated phenotypes in our data. It appears that some of these are located in this telomere dataset. We are now removing them and continuing our analysis.

2 Data

Libraries

```
library(tidyverse)
library(knitr)
library(batchma)
library(ggplot2)
library(stargazer)
library(tableone)
library(arsenal)
library(ggpubr)
library(rstatix)
library(broom)
library(ggrepel)
library(knitr)
library(kableExtra)
library(gtsummary)
library(table1)
library(flextable)
library(janitor)
library(table1)
library(ggeasy)
library(rcompanion)
library(nnet)
library(gt)

#devtools::install_github("thomasp85/patchwork")
library(patchwork)
```

Data specifics

- I was sent the data specified in *tel_data* from Steven Senese ssenese@hsph.harvard.edu on 09-26-2023. This data includes the telomere lengths. This is a modified version of the *Koenen Sorted Data*, that included 5 sets of data based on the plates run for telomere analysis. I copied those into one document, *tel_data*, but specified in column Set which plate it came from.
- I was sent the data specified in *manifest* from Patrice Soule spsoule@hsph.harvard.edu on 10-16-2023.
- The data specified in *freeze* is from the latest data freeze file located in DropBox NeuroGAP Psychosis called *NeuroGAP-P_Release8_Kenya_as-of-2023-08-29*

Loading and merging data

```

#Adjust to fit your computer

path_analytic <- "/Users/ham593/Dropbox (Harvard University)/NeuroGAP-Psychosis/Telomeres/Data/Analytic
path_raw <- "/Users/ham593/Dropbox (Harvard University)/NeuroGAP-Psychosis/Telomeres/Data/Raw Data"

#Load in data-----
#Analytical Data
setwd(path_analytic)
tel_data <- read.csv('Telomeredata_new.csv')
freeze <- read.csv('Copy of NeuroGAP_DataFreeze8.csv')

#Raw Data
setwd(path_raw)
manifest <- read.csv('Full manifest Broad PDO-31674 Plate Map.csv')

colnames(manifest)[colnames(manifest) == "Collaborator.Participant.ID"] <- "subjid"
colnames(freeze)[colnames(freeze) == "subj_id"] <- "subjid"

#Remove all Reference/QC rows and Insufficient DNA
tel_data <- tel_data[grepl("^SM-", tel_data$Sample.ID), ]
tel_data <- tel_data[!grepl("Failed", tel_data$Tel.CT.1),]
tel_data <- tel_data[!grepl("Insufficient DNA", tel_data$Tel.CT.1),]

#MERGE
tel_all <- merge(manifest, tel_data, by="Sample.ID")
tel_all <- merge(tel_all, freeze, by="subjid")

#We found that Age was duplicated, the below will remove on of the Age variables.
#Remove duplicated columns
tel_all <- tel_all %>%
  subset(select=which(!duplicated(names(.))))

setwd("/Users/ham593/Dropbox (Harvard University)/NeuroGAP-Psychosis/Data Working Group (DAWG)/Data/Data
ngap_7 <- read.csv("NeuroGAP-P_Release7_Final.csv")

setwd(path_analytic)

```

2.1 Finding Duplicates

Loading in duplicate data

```

setwd("/Users/ham593/Dropbox (Harvard University)/NeuroGAP-Psychosis/Data Working Group (DAWG)/Phenotyp
duplicate <- read.csv("Phenotype Swap Data_4-10-24.csv")

```

Finding duplicates

```

dups1 <- duplicate$ID1_COLLABORATOR_PARTICIPANT_ID
dups2 <- duplicate$ID2_COLLABORATOR_PARTICIPANT_ID

```

```

tel_all_id <- tel_all$subjid

dup_tel_all <- data.frame()

# Iterate over each unique element in dups1
for (i in 1:length(dups1)) {

  #if the item is in tel_all_id append it and it's corresponding duplicate
  if (dups1[i] %in% tel_all_id) {
    #print(dups1[i])
    # Create a new row with the corresponding elements from dups1 and dups2
    new_row <- data.frame(dups1 = dups1[i], dups2 = dups2[i])
    # Append the new row to dup_tel_all
    dup_tel_all <- rbind(dup_tel_all, new_row)
  }
}

#Check how many of the duplicates from column 2 are in the tel_all data
duplicated_final <- intersect(dup_tel_all$dups2, tel_all_id)

#Remove only these participants from tel_all

#Logic: Only these participants are duplicated, no need to remove those that appear in both list, but d
#For duplicate in tel_all, only need to remove 1 participant of the pair.

tel_all <- tel_all %>% filter(!(subjid %in% duplicated_final))

#Reconfirm that there are no remaining duplicates.
tel_all_id_new <- tel_all$subjid
count <- 0

for (i in 1:length(dups1)) {

  if (dups1[i] %in% tel_all_id_new &
      dups2[i] %in% tel_all_id_new) {
    count <- count+1
    print(paste("Dups1:",dups1[i], "and Dups2:",dups2[i] ))
  }
}

if (count == 0 ) {
  print("No problem, all duplicates found")
} else {print("Problem: Check IDS")}

## [1] "No problem, all duplicates found"

```

```

tel_all <- tel_all %>% mutate(is_case_clean = case_when(
  is_case == 1 &

  (tobacco_3cat %in% c("Irregular Users",
    "Regular Users") |

  cidi_q1 == 1 |
  cidi_q2 == 1 |
  cidi_q3 == 1 |
  cidi_q4 == 1 |
  cidi_q5 == 1 |
  cidi_q6 == 1 |
  cidi_q7 == 1 |
  cidi_q8 == 1 |
  cidi_q9 == 1 |
  cidi_q10 == 1 |
  cidi_q11 == 1 |
  cidi_q12 == 1 |
  cidi_q13 == 1 |
  cidi_q14 == 1 |
  cidi_q15 == 1 |
  cidi_q16 == 1 |
  cidi_q17 == 1) ~ "Unhealthy Case",
  is_case == 0 ~ "Control",
  TRUE ~ "Healthy Case"
)) %>% mutate(is_case_clean = factor(is_case_clean,
  levels = c("Control", "Healthy Case",
    "Unhealthy Case")))

```

```

tel_all_truescd <- tel_all %>% mutate(is_case_clean = case_when(
  (is_case == 1 &

  tobacco_3cat %in% c("Irregular Users",
    "Regular Users") |
  alcohol_3cat %in% c("Irregular Users",
    "Regular Users") |
  cannabis_3cat %in% c("Irregular Users",
    "Regular Users") |
  khat_3cat %in% c("Irregular Users",
    "Regular Users") |

  cidi_q1 == 1 |
  cidi_q2 == 1 |
  cidi_q3 == 1 |
  cidi_q4 == 1 |
  cidi_q5 == 1 |
  cidi_q6 == 1 |
  cidi_q7 == 1 |
  cidi_q8 == 1 |
  cidi_q9 == 1 |
  cidi_q10 == 1 |
  cidi_q11 == 1 |
  cidi_q12 == 1 |
  cidi_q13 == 1 |

```

```

cidi_q14 == 1 |
cidi_q15 == 1 |
cidi_q16 == 1 |
cidi_q17 == 1) ~ "Unhealthy Case",
is_case == 0 ~ "Control",
TRUE ~ "Healthy Case"
)) %>% mutate(is_case_clean = factor(is_case_clean,
  levels = c("Control", "Healthy Case",
    "Unhealthy Case")))

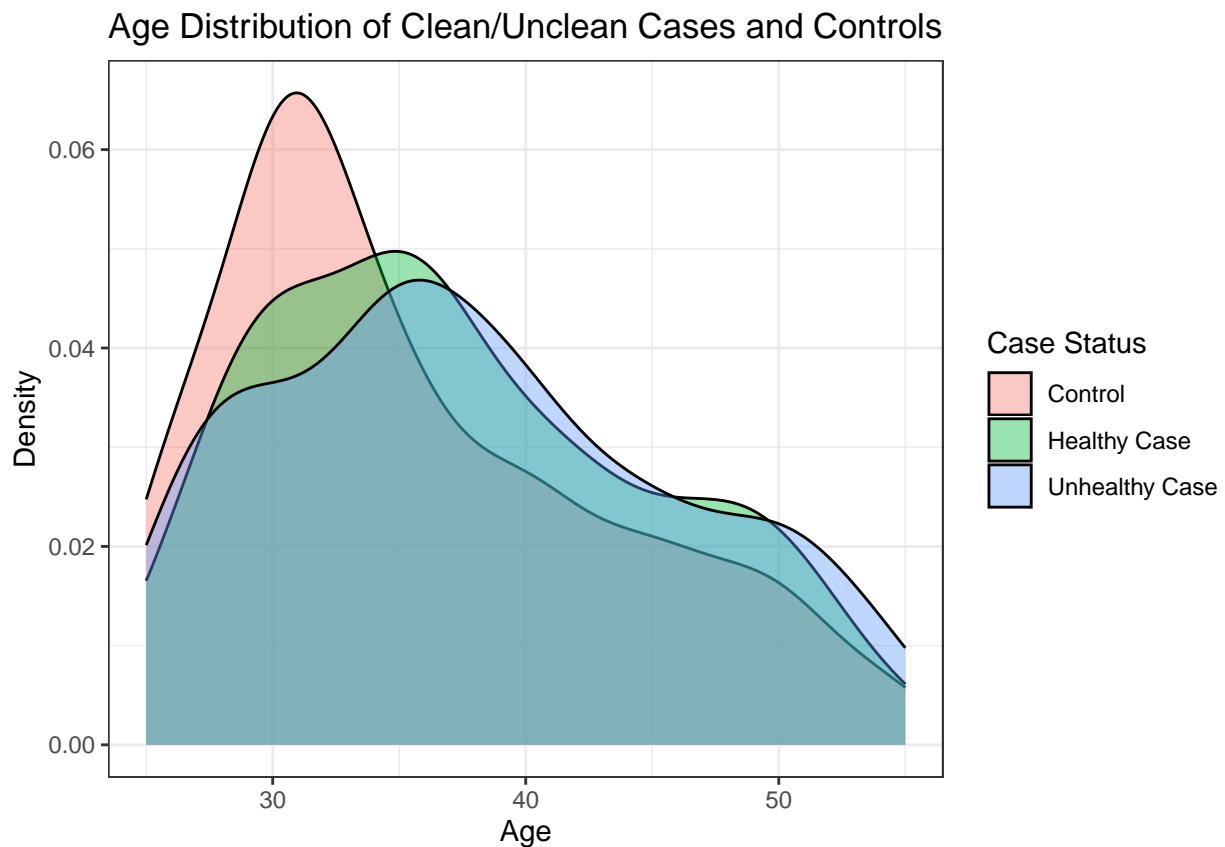
```

Age distribution of clean/unclean cases and controls.

```

ggplot(tel_all, aes(x=age_at_iview, group = is_case_clean, fill=is_case_clean)) +
  geom_density(alpha=0.4) +
  xlab("Age") + ylab("Density") +
  ggtitle("Age Distribution of Clean/Unclean Cases and Controls") +
  ggeasy::easy_center_title() +
  labs(fill='Case Status') +
  #scale_y_continuous(labels = scales::percent_format(scale = 1)) +
  theme_bw()

```



2.2 Rosner Batch Correction 3-29-24

Next steps: 1. Filter 26 missing after batch correction

2. Demo table, order-> controls, filtered cases (clean cases), unfiltered cases
3. Scatterplot: x-axis = age, y-axis = batch corrected LTL
4. Boxplots of LTL by case status (controls, filtered, unfiltered)
5. Correlations between age and batch corrected LTL by case status so r for controls = ? r for filtered cases = ?
6. Rename variables and values (e.g., High blood pressure 0 = “Normal blood pressure”, 1 = “High blood pressure”)
7. Run 6 models, individual tables for each model + summary table
8. Stratified analyses

A vs B vs C vs D

If A and C look similar: Run

ACE vs B vs D

Models: 1. $\text{Lm}(\text{Rosner LTL} \sim \text{age} + \text{case status (clean/unclean)})$ 2. $\text{Lm}(\text{Rosner LTL} \sim \text{age} + \text{case status (clean/unclean)} + \text{sex})$ 3. $\text{lm}(\text{Rosner LTL} \sim \text{Age} + \text{case} + \text{sex} + \text{education} + \text{BMI})$ 4. $\text{lm}(\text{Rosner LTL} \sim \text{Age} + \text{case} + \text{sex} + \text{education} + \text{BMI} + \text{All Others})$ 5. Model 4 + sex * case status interaction 6. Model 4 + education * case status interaction

2.2.1 1. Rosner Batch Correction

```
#Rosner batch correction with clean/unclean cases

ros_df <- tel_all %>% select(is_case_clean, subjid, ltl, Set, Age)

ros <- adjust_batch(
  data = ros_df,
  markers = ltl,
  batch = Set,
  method = simple
)
```

How many people do we lose excluding CIDI variables (Excluding HIV and cancer)

```
tel_all %>% filter(cidi_q1 %in% c(777) |
  cidi_q2 %in% c(777) |
  cidi_q3 %in% c(777) |
  cidi_q4 %in% c(777) |
  cidi_q5 %in% c(777) |
  cidi_q6 %in% c(777) |
  cidi_q7 %in% c(777) |
  cidi_q8 %in% c(777) |
  cidi_q9 %in% c(777) |
  cidi_q10 %in% c(777) |
  cidi_q11 %in% c(777) |
  cidi_q12 %in% c(777) |
  cidi_q13 %in% c(777) |
  cidi_q14 %in% c(777) |
  cidi_q16 %in% c(777)
) %>% group_by(is_case) %>% summarise(`Count with missing` = n())
```

```
## # A tibble: 1 x 2
##   is_case `Count with missing`
##   <fct>      <int>
## 1 1          26
```

2.2.2 2. Filter

Filter these 26 these people from analysis.

```
tel_all_nomissing <- tel_all %>% filter(!(cidi_q1 %in% c(777) |
  cidi_q2 %in% c(777) |
  cidi_q3 %in% c(777) |
  cidi_q4 %in% c(777) |
  cidi_q5 %in% c(777) |
  cidi_q6 %in% c(777) |
  cidi_q7 %in% c(777) |
  cidi_q8 %in% c(777) |
  cidi_q9 %in% c(777) |
  cidi_q10 %in% c(777) |
  cidi_q11 %in% c(777) |
  cidi_q12 %in% c(777) |
  cidi_q13 %in% c(777) |
  cidi_q14 %in% c(777) |
  #Skip cidi_q15, HIV
  cidi_q16 %in% c(777)
  #Skip cidi_q17, Cancer
))

#Update 01-22-2025
tel_all_nomissing <- tel_all_nomissing %>%
  mutate(
    across(c(cidi_q1:cidi_q14), ~ factor(., levels = c("0", "1"))),
    cidi_q16 = factor(cidi_q16, levels = c("0", "1"))
```



```

)

###TESTING

tel_all_nomissing_truescd <- tel_all %>% filter(!(cidi_q1 %in% c(777) |
  cidi_q2 %in% c(777) |
  cidi_q3 %in% c(777) |
  cidi_q4 %in% c(777) |
  cidi_q5 %in% c(777) |
  cidi_q6 %in% c(777) |
  cidi_q7 %in% c(777) |
  cidi_q8 %in% c(777) |
  cidi_q9 %in% c(777) |
  cidi_q10 %in% c(777) |
  cidi_q11 %in% c(777) |
  cidi_q12 %in% c(777) |
  cidi_q13 %in% c(777) |
  cidi_q14 %in% c(777) |
  cidi_q15 %in% c(777) |
  cidi_q16 %in% c(777) |
  cidi_q17 %in% c(777)
))

#Update 01-22-2025
tel_all_nomissing_truescd <- tel_all_nomissing_truescd %>%
  mutate(
    across(c(cidi_q1:cidi_q17), ~ factor(., levels = c("0", "1")))
  )

```

2.2.3 3. Figures

Scatterplot – age and batch corrected ltl

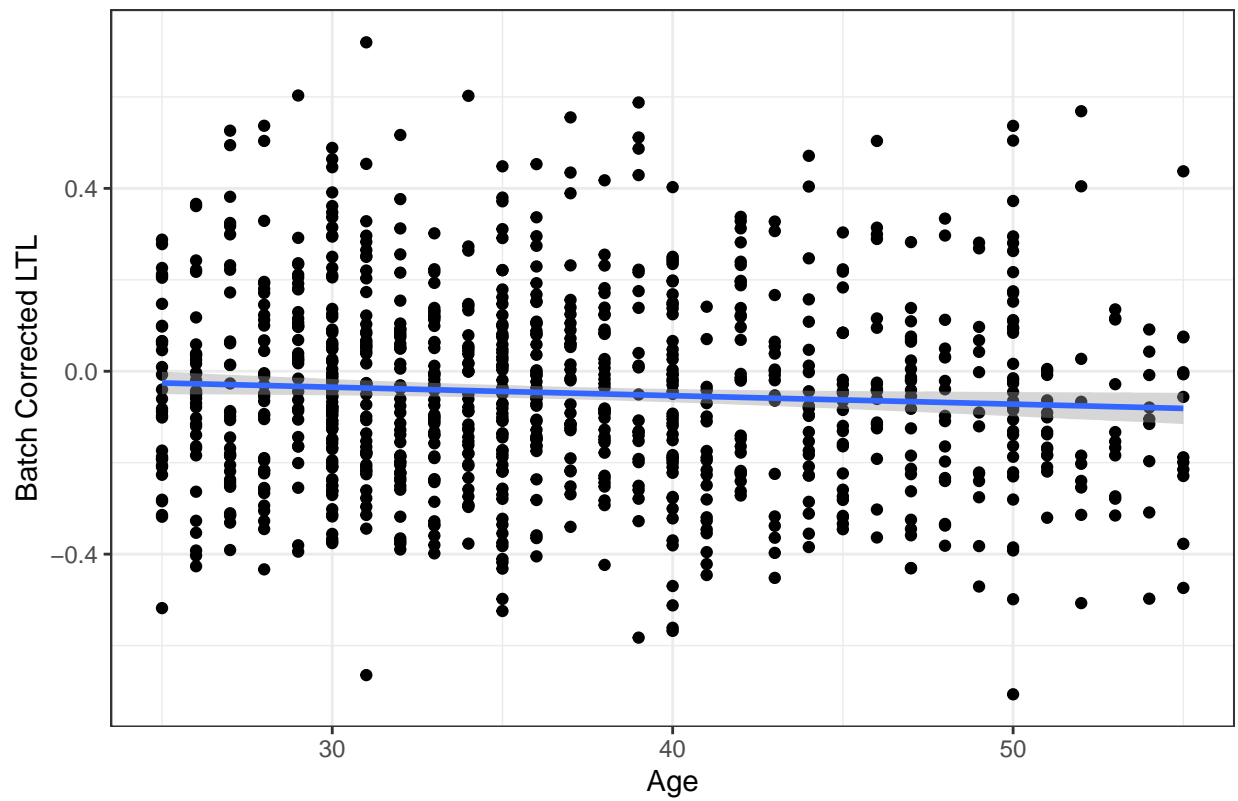
```

ros %>%
  ggplot(aes(x = Age, y = ltl_adj2)) +
  geom_point() +
  stat_smooth(method = "lm") + theme_bw() +
  xlab("Age") + ylab("Batch Corrected LTL") +
  ggtitle("Age vs. Batch Corrected LTL Scatterplot")

```

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

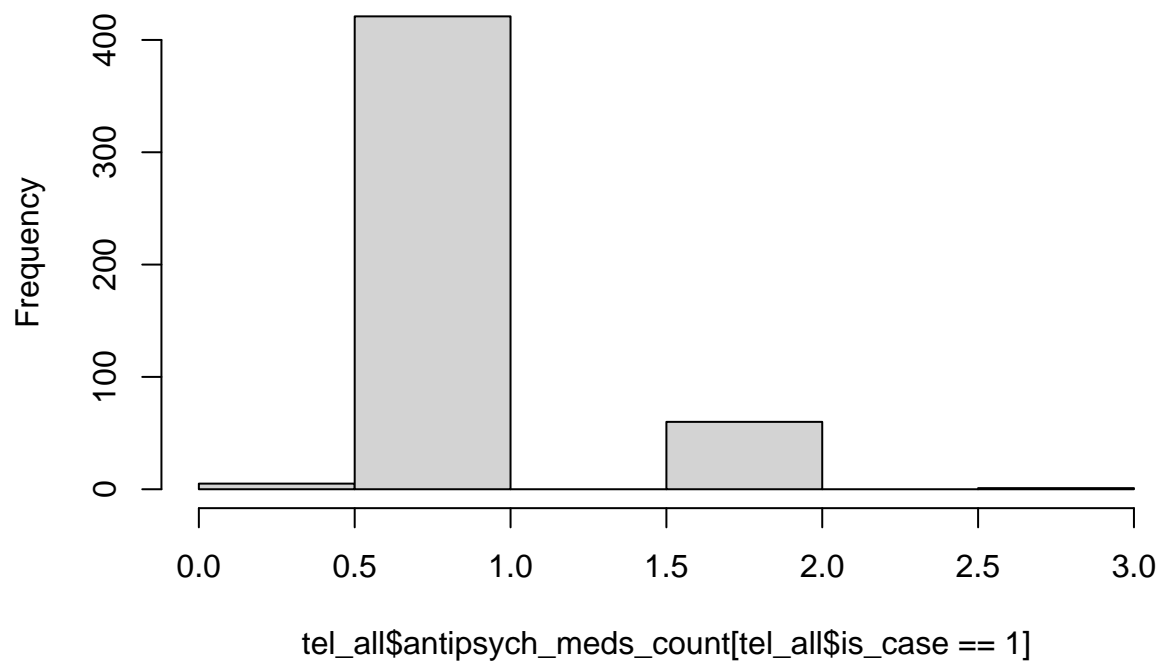
Age vs. Batch Corrected LTL Scatterplot



Histogram of antipsychotic use

```
hist(tel_all$antipsych_meds_count[tel_all$is_case==1])
```

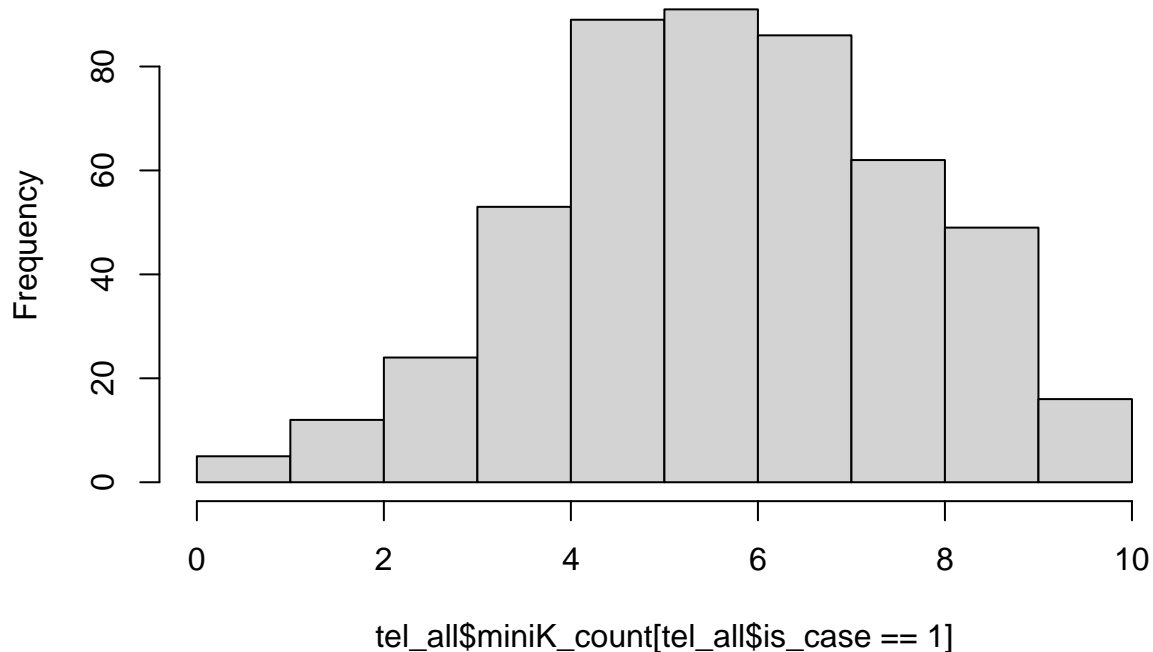
Histogram of tel_all\$antipsych_meds_count[tel_all\$is_case == 1]



Histogram of psychosis severity

```
hist(tel_all$miniK_count[tel_all$is_case == 1])
```

Histogram of tel_all\$miniK_count[tel_all\$is_case == 1]



Boxplots of LTL and filtered/unfiltered cases

```
compare_means(ltl ~ is_case_clean, data = tel_all_nomissing, method = "t.test")
```

```
## # A tibble: 3 x 8
##   .y. group1      group2          p p.adj p.format p.signif method
##   <chr> <chr>      <chr>      <dbl> <dbl> <chr>      <chr>      <chr>
## 1 ltl   Control    Healthy Case  0.000449 0.0013 0.00045 ***      T-test
## 2 ltl   Control    Unhealthy Case 0.364    0.36  0.36446 ns       T-test
## 3 ltl   Healthy Case Unhealthy Case 0.0119   0.024 0.01194 *       T-test
```

```
my_comparisons <- list( c("Control", "Healthy Case"),
                        c("Healthy Case", "Unhealthy Case"),
                        c("Control", "Unhealthy Case") )
```

```
tel_all_nomissing_2 <- tel_all_nomissing %>%
  mutate(case_2 = case_when(
    is_case_clean == "Control" ~ "Control",
    is_case_clean == "Healthy Case" ~ "Cw/oSCD",
    is_case_clean == "Unhealthy Case" ~ "CwSCD",
  ))
```

```
compare_means(ltl ~ case_2, data = tel_all_nomissing_2, method = "t.test")
```

```
## # A tibble: 3 x 8
```

```
##   .y.  group1 group2      p  p.adj p.format p.signif method
##   <chr> <chr>  <chr>    <dbl> <dbl> <chr>    <chr>    <chr>
## 1 ltl   CwSCD   Control 0.364   0.36  0.36446  ns      T-test
## 2 ltl   CwSCD   Cw/oSCD 0.0119  0.024 0.01194  *      T-test
## 3 ltl   Control Cw/oSCD 0.000449 0.0013 0.00045  ***    T-test
```

```
my_comparisons <- list( c("Control", "Cw/oSCD"),
                        c("Cw/oSCD", "CwSCD"),
                        c("Control", "CwSCD"))

tel_all_nomissing_2 %>%
  ggplot(aes(x=case_2, y=ltl)) +
  geom_boxplot() +
  ylab("Log Telomere Length") + xlab("") +
  stat_compare_means(comparisons = my_comparisons,
                    label = "p.signif") + # Add pairwise comparisons p-value
  stat_compare_means(label.y = 1.4, method = "anova") + theme_classic2() +
  labs(caption = "Figure 1: LTL between case-control groups") +
  theme(plot.caption = element_text(size = 11, hjust = 0, face = "italic"))
```

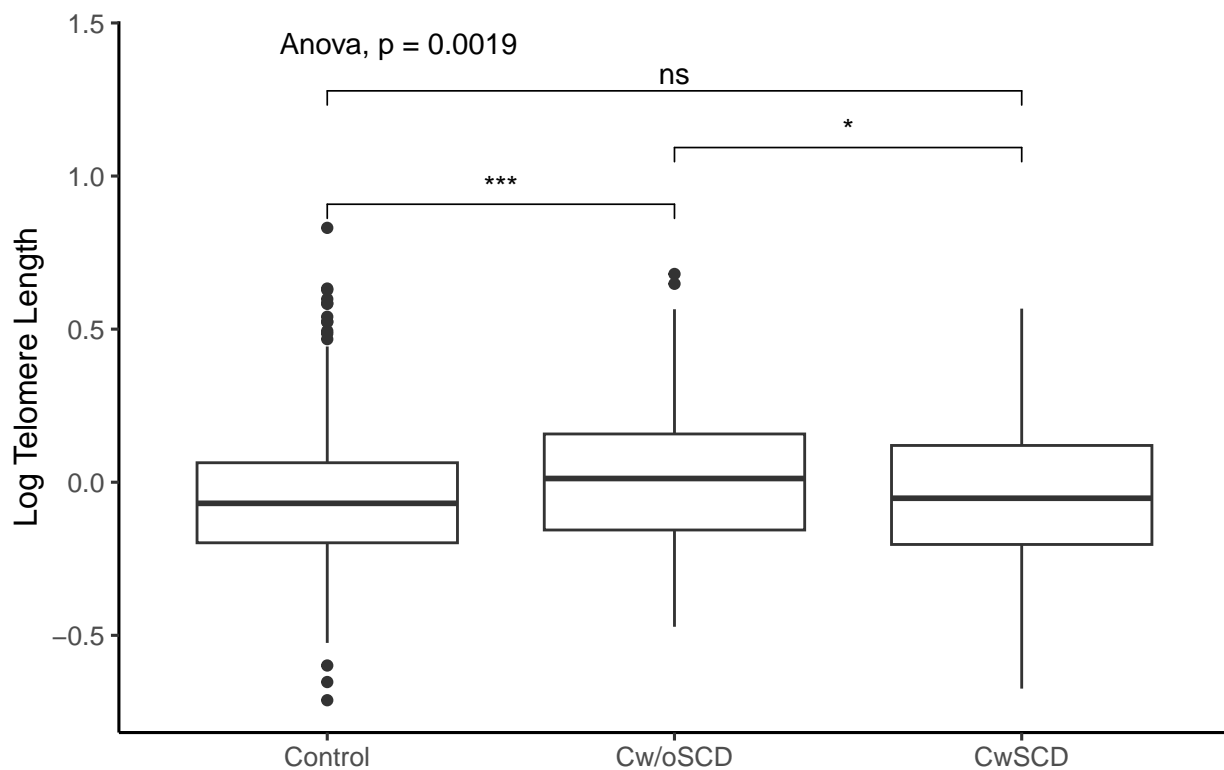


Figure 1: LTL between case-control groups

Compare means sex

```
compare_means(ltl ~ msex, data = tel_all_nomissing, method = "t.test")
```

```
## # A tibble: 1 x 8
```

```
##   .y.   group1 group2      p  p.adj p.format p.signif method
##   <chr> <chr>  <chr>    <dbl> <dbl> <chr>    <chr>    <chr>
## 1 ltl    Male   Female 0.00262 0.0026 0.0026    **      T-test
```

```
tel_all_nomissing %>% group_by(msex) %>% summarise(mean_ltl = mean(ltl),
                                                    sd_ltl = sd(ltl))
```

```
## # A tibble: 2 x 3
##   msex   mean_ltl sd_ltl
##   <fct>     <dbl> <dbl>
## 1 Male    -0.0518  0.216
## 2 Female -0.00336  0.242
```

2.2.4 4. Correlations between age and batch corrected LTL by case status so r for controls = ? r for filtered cases

```
cor_df <- ros %>% group_by(is_case_clean) %>%
  summarise(R_age_adj.ltl = round(cor(Age, ltl_adj2),3))

colnames(cor_df) <- c("Case Status", "r")
kable(cor_df, format="html", caption="Correlation of Age and Batch Corrected LTL") %>%
  kable_classic(full_width = F, html_font = "Cambria")
```

Correlation of Age and Batch Corrected LTL

Case Status

r

Control

-0.072

Healthy Case

-0.140

Unhealthy Case

-0.048

HIV

```
tel_all_nomissing %>% group_by(cidi_q15, hiv_positive) %>% summarize(count = n())
```

```
## 'summarise()' has grouped output by 'cidi_q15'. You can override using the
## '.groups' argument.
```

```
## # A tibble: 7 x 3
## # Groups:   cidi_q15 [3]
##   cidi_q15 hiv_positive count
##   <fct>         <dbl> <int>
## 1 0             0     561
## 2 0             1       1
```

```
## 3 0          777  339
## 4 1           0    1
## 5 1           1    8
## 6 777         0    9
## 7 777        777   20
```

2.2.5 5. Table with new data excluding CIDI missing

```
rndr <- function(x, name, ...) {
  if (!is.numeric(x)) return(render.categorical.default(x))
  what <- switch(name,
    ltl = "Mean (SD)",
    age_at_iview = "Mean (SD)",
    Age = "Mean (SD)",
  )
  parse.abbrev.render.code(c("", what))(x)
}

pvalueANOVA <- function(x, ...) {
  # Construct vectors of data y, and groups (strata) g
  y <- unlist(x)
  g <- factor(rep(1:length(x), times=sapply(x, length)))

  if (is.numeric(y)) {
    # For numeric variables, perform a standard 2-sample t-test
    ano <- aov(y ~ g)
    p <- summary(ano)[[1]][[5]][1]

  } else {
    # For categorical variables, perform a chi-squared test of independence
    p <- chisq.test(table(y, g))$p.value
  }

  # Format the p-value, using an HTML entity for the less-than sign.
  # The initial empty string places the output on the line below the variable label.
  c("", sub("<", "&lt;", format.pval(p, digits=3, eps=0.001)))
}

label(tel_all_nomissing$bmi_bin) <- "BMI"

tbl <- table1(~ ltl + age_at_iview + msex + educ_ord +
  bmi_bin + alcohol_3cat + khat_3cat + cannabis_3cat +
  tobacco_3cat +
  cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 | is_case_clean,
  data=tel_all_nomissing, overall=F,
  caption = "Demographics by Case Status",
  render=rndr)

tbl
```

Table 1: Demographics by Case Status

	Control (N=478)	Healthy Case (N=183)	Unhealthy Case (N=278)
Log Telomere Length			
Mean (SD)	-0.0531 (0.226)	0.0163 (0.225)	-0.0377 (0.224)
Age			
Mean (SD)	35.7 (7.70)	37.5 (7.52)	37.8 (8.10)
Sex			
Male	312 (65.3%)	103 (56.3%)	199 (71.6%)
Female	166 (34.7%)	80 (43.7%)	79 (28.4%)
Education			
Primary or less	135 (28.2%)	64 (35.0%)	90 (32.4%)
Secondary	136 (28.5%)	78 (42.6%)	120 (43.2%)
College	207 (43.3%)	41 (22.4%)	68 (24.5%)
BMI			
Normal Weight (18.6 - 24.9)	303 (63.4%)	113 (61.7%)	183 (65.8%)
Underweight (≤ 18.5)	28 (5.9%)	8 (4.4%)	19 (6.8%)
Overweight (25.0 - 29.9)	114 (23.8%)	42 (23.0%)	50 (18.0%)
Obese (≥ 30)	33 (6.9%)	20 (10.9%)	26 (9.4%)
Alcohol Use			
Never Users	198 (41.4%)	136 (74.3%)	89 (32.0%)
Irregular Users	185 (38.7%)	40 (21.9%)	157 (56.5%)
Regular Users	95 (19.9%)	7 (3.8%)	32 (11.5%)
Khat Use			
Never Users	398 (83.3%)	147 (80.3%)	100 (36.0%)
Irregular Users	52 (10.9%)	26 (14.2%)	108 (38.8%)
Regular Users	28 (5.9%)	10 (5.5%)	70 (25.2%)
Cannabis Use			
Never Users	472 (98.7%)	183 (100%)	244 (87.8%)
Irregular Users	6 (1.3%)	0 (0%)	30 (10.8%)
Regular Users	0 (0%)	0 (0%)	4 (1.4%)
Tobacco Use			
Never Users	478 (100%)	183 (100%)	105 (37.8%)
Irregular Users	0 (0%)	0 (0%)	107 (38.5%)
Regular Users	0 (0%)	0 (0%)	66 (23.7%)
cidi_q1			
0	478 (100%)	183 (100%)	241 (86.7%)
1	0 (0%)	0 (0%)	37 (13.3%)
cidi_q2			
0	478 (100%)	183 (100%)	225 (80.9%)
1	0 (0%)	0 (0%)	53 (19.1%)
cidi_q3			
0	478 (100%)	183 (100%)	202 (72.7%)
1	0 (0%)	0 (0%)	76 (27.3%)
cidi_q4			
0	478 (100%)	183 (100%)	264 (95.0%)
1	0 (0%)	0 (0%)	14 (5.0%)


```
tbl <- table1(
  ~ ltl + age_at_iview + msex + educ_ord +
    bmi_bin + cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 +
    cidi_q6 + cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
    cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
    cidi_q17 | is_case_clean,
  data = tel_all_nomissing, overall = F,
  caption = "Demographics by Case Status",
  render = rndr, extra.col = list(`ANOVA/Chi-Sq` = pvalueANOVA)
)
```

```
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
```

```
tbl
```

```
tbl <- table1(
  ~ ltl + age_at_iview + is_case_clean + educ_ord +
    bmi_bin + cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 +
    cidi_q6 + cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
    cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
    cidi_q17 | msex ,
  data = tel_all_nomissing, overall = F,
  caption = "Demographics by Case Status",
  render = rndr, extra.col = list(`ANOVA/Chi-Sq` = pvalueANOVA)
)
```

```
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
```

Table 2: Demographics by Case Status

	Control (N=478)	Healthy Case (N=183)	Unhealthy Case (N=278)	ANOVA/Chi-Sq
Log Telomere Length				0.0019
Mean (SD)	-0.0531 (0.226)	0.0163 (0.225)	-0.0377 (0.224)	
Age				<0.001
Mean (SD)	35.7 (7.70)	37.5 (7.52)	37.8 (8.10)	
Sex				0.00332
Male	312 (65.3%)	103 (56.3%)	199 (71.6%)	
Female	166 (34.7%)	80 (43.7%)	79 (28.4%)	
Education				<0.001
Primary or less	135 (28.2%)	64 (35.0%)	90 (32.4%)	
Secondary	136 (28.5%)	78 (42.6%)	120 (43.2%)	
College	207 (43.3%)	41 (22.4%)	68 (24.5%)	
BMI				0.296
Normal Weight (18.6 - 24.9)	303 (63.4%)	113 (61.7%)	183 (65.8%)	
Underweight (≤ 18.5)	28 (5.9%)	8 (4.4%)	19 (6.8%)	
Overweight (25.0 - 29.9)	114 (23.8%)	42 (23.0%)	50 (18.0%)	
Obese (≥ 30)	33 (6.9%)	20 (10.9%)	26 (9.4%)	
cidi_q1				<0.001
0	478 (100%)	183 (100%)	241 (86.7%)	
1	0 (0%)	0 (0%)	37 (13.3%)	
cidi_q2				<0.001
0	478 (100%)	183 (100%)	225 (80.9%)	
1	0 (0%)	0 (0%)	53 (19.1%)	
cidi_q3				<0.001
0	478 (100%)	183 (100%)	202 (72.7%)	
1	0 (0%)	0 (0%)	76 (27.3%)	
cidi_q4				<0.001
0	478 (100%)	183 (100%)	264 (95.0%)	
1	0 (0%)	0 (0%)	14 (5.0%)	
cidi_q5				<0.001
0	478 (100%)	183 (100%)	255 (91.7%)	
1	0 (0%)	0 (0%)	23 (8.3%)	
cidi_q6				0.304
0	478 (100%)	183 (100%)	277 (99.6%)	
1	0 (0%)	0 (0%)	1 (0.4%)	
cidi_q7				<0.001
0	478 (100%)	183 (100%)	268 (96.4%)	
1	0 (0%)	0 (0%)	10 (3.6%)	
cidi_q8				<0.001
0	478 (100%)	183 (100%)	272 (97.8%)	
1	0 (0%)	0 (0%)	6 (2.2%)	
cidi_q9				<0.001
0	478 (100%)	183 (100%)	257 (92.4%)	
1	0 (0%)	0 (0%)	21 (7.6%)	
cidi_q10				<0.001
0	478 (100%)	183 (100%)	271 (97.5%)	
1	0 (0%)	0 (0%)	7 (2.5%)	
cidi_q11		18		
0	478 (100%)	183 (100%)	253 (91.0%)	<0.001
1	0 (0%)	0 (0%)	25 (9.0%)	
cidi_q12				

```
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
```

```
tbl
```

```
tbl <- table1(
  ~ ltl + age_at_iview + msex + educ_ord +
    bmi_bin + cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 +
    cidi_q6 + cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
    cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
    cidi_q17 | is_case_clean,
  data = tel_all_nomissing_truescd, overall = F,
  caption = "Demographics by Case Status",
  render = rndr, extra.col = list(`ANOVA/Chi-Sq` = pvalueANOVA)
)
```

```
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
```

Table 3: Demographics by Case Status

	Male	Female	ANOVA/Chi-Sq
	(N=614)	(N=325)	
Log Telomere Length			
Mean (SD)	-0.0518 (0.216)	-0.00336 (0.242)	0.0018
Age			
Mean (SD)	36.3 (7.62)	37.4 (8.21)	0.043
is_case_clean			
Control	312 (50.8%)	166 (51.1%)	0.00332
Healthy Case	103 (16.8%)	80 (24.6%)	
Unhealthy Case	199 (32.4%)	79 (24.3%)	
Education			
Primary or less	175 (28.5%)	114 (35.1%)	<0.001
Secondary	201 (32.7%)	133 (40.9%)	
College	238 (38.8%)	78 (24.0%)	
BMI			
Normal Weight (18.6 - 24.9)	437 (71.2%)	162 (49.8%)	<0.001
Underweight (≤ 18.5)	40 (6.5%)	15 (4.6%)	
Overweight (25.0 - 29.9)	117 (19.1%)	89 (27.4%)	
Obese (≥ 30)	20 (3.3%)	59 (18.2%)	
cidi_q1			
0	597 (97.2%)	305 (93.8%)	0.0183
1	17 (2.8%)	20 (6.2%)	
cidi_q2			
0	583 (95.0%)	303 (93.2%)	0.348
1	31 (5.0%)	22 (6.8%)	
cidi_q3			
0	575 (93.6%)	288 (88.6%)	0.0103
1	39 (6.4%)	37 (11.4%)	
cidi_q4			
0	606 (98.7%)	319 (98.2%)	0.711
1	8 (1.3%)	6 (1.8%)	
cidi_q5			
0	601 (97.9%)	315 (96.9%)	0.495
1	13 (2.1%)	10 (3.1%)	
cidi_q6			
0	613 (99.8%)	325 (100%)	1
1	1 (0.2%)	0 (0%)	
cidi_q7			
0	608 (99.0%)	321 (98.8%)	0.979
1	6 (1.0%)	4 (1.2%)	
cidi_q8			
0	611 (99.5%)	322 (99.1%)	0.716
1	3 (0.5%)	3 (0.9%)	
cidi_q9			
0	605 (98.5%)	313 (96.3%)	0.0496
1	9 (1.5%)	12 (3.7%)	
cidi_q10			
0	610 (99.3%)	322 (99.1%)	0.951
1	4 (0.7%) ²⁰	3 (0.9%)	
cidi_q11			
0	603 (98.2%)	311 (95.7%)	0.0389
1	11 (1.8%)	14 (4.3%)	

```
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
## Warning in stats::chisq.test(x, y, ...): Chi-squared approximation may be
## incorrect
```

```
tbl
```

```
tel_all_nomissing_2 <- tel_all_nomissing_2 %>%
  mutate(bmi_bin_ordered = factor(bmi_bin,
                                  levels = c("Underweight (<= 18.5)",
                                               "Normal Weight (18.6 - 24.9)",
                                               "Overweight (25.0 - 29.9)",
                                               "Obese (>= 30)")))

label(tel_all_nomissing_2$bmi_bin_ordered) <- "BMI"

tbl <- table1(~ ltl + age_at_iview + msex + educ_ord +
              bmi_bin_ordered | case_2,
              data=tel_all_nomissing_2, overall=F,
              caption = "Demographics by Case Status",
              render=rndr)

pdf("Demographics")
tbl
```

```
dev.off()
```

```
## pdf
## 2
```

```
tbl <- table1(~ ltl + age_at_iview + msex + educ_ord +
              bmi_bin + alcohol_3cat + khat_3cat + cannabis_3cat +
              tobacco_3cat
              | is_case_f,
              data=tel_all_nomissing, overall=F,
              caption = "Demographics by Case Status",
              render=rndr)

tbl
```

```
tbl <- table1(~ cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
              cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 + cidi_q12 +
              cidi_q13 + cidi_q14 + cidi_q15 +
              cidi_q16 + cidi_q17 | is_case_clean,
              data=tel_all_nomissing, overall=F,
              caption = "Demographics by Case Status",
              render=rndr)

tbl
```

Table 4: Demographics by Case Status

	Control (N=478)	Healthy Case (N=173)	Unhealthy Case (N=252)	ANOVA/Chi-Sq
Log Telomere Length				
Mean (SD)	-0.0531 (0.226)	0.0168 (0.223)	-0.0371 (0.227)	0.00233
Age				
Mean (SD)	35.7 (7.70)	37.4 (7.56)	37.9 (8.04)	<0.001
Sex				
Male	312 (65.3%)	98 (56.6%)	179 (71.0%)	0.00929
Female	166 (34.7%)	75 (43.4%)	73 (29.0%)	
Education				
Primary or less	135 (28.2%)	60 (34.7%)	84 (33.3%)	<0.001
Secondary	136 (28.5%)	75 (43.4%)	108 (42.9%)	
College	207 (43.3%)	38 (22.0%)	60 (23.8%)	
BMI				
Normal Weight (18.6 - 24.9)	303 (63.4%)	106 (61.3%)	168 (66.7%)	0.133
Underweight (≤ 18.5)	28 (5.9%)	7 (4.0%)	17 (6.7%)	
Overweight (25.0 - 29.9)	114 (23.8%)	40 (23.1%)	42 (16.7%)	
Obese (≥ 30)	33 (6.9%)	20 (11.6%)	25 (9.9%)	
cidi_q1				
0	478 (100%)	173 (100%)	218 (86.5%)	<0.001
1	0 (0%)	0 (0%)	34 (13.5%)	
cidi_q2				
0	478 (100%)	173 (100%)	205 (81.3%)	<0.001
1	0 (0%)	0 (0%)	47 (18.7%)	
cidi_q3				
0	478 (100%)	173 (100%)	181 (71.8%)	<0.001
1	0 (0%)	0 (0%)	71 (28.2%)	
cidi_q4				
0	478 (100%)	173 (100%)	240 (95.2%)	<0.001
1	0 (0%)	0 (0%)	12 (4.8%)	
cidi_q5				
0	478 (100%)	173 (100%)	230 (91.3%)	<0.001
1	0 (0%)	0 (0%)	22 (8.7%)	
cidi_q6				
0	478 (100%)	173 (100%)	251 (99.6%)	0.274
1	0 (0%)	0 (0%)	1 (0.4%)	
cidi_q7				
0	478 (100%)	173 (100%)	242 (96.0%)	<0.001
1	0 (0%)	0 (0%)	10 (4.0%)	
cidi_q8				
0	478 (100%)	173 (100%)	246 (97.6%)	<0.001
1	0 (0%)	0 (0%)	6 (2.4%)	
cidi_q9				
0	478 (100%)	173 (100%)	232 (92.1%)	<0.001
1	0 (0%)	0 (0%)	20 (7.9%)	
cidi_q10				
0	478 (100%)	173 (100%)	245 (97.2%)	<0.001
1	0 (0%)	0 (0%)	7 (2.8%)	
cidi_q11				
0	478 (100%)	173 (100%)	229 (90.9%)	<0.001
1	0 (0%)	0 (0%)	23 (9.1%)	
cidi_q12				

Table 5: Demographics by Case Status

	Control	Cw/oSCD	CwSCD
	(N=478)	(N=183)	(N=278)
Log Telomere Length			
Mean (SD)	-0.0531 (0.226)	0.0163 (0.225)	-0.0377 (0.224)
Age			
Mean (SD)	35.7 (7.70)	37.5 (7.52)	37.8 (8.10)
Sex			
Male	312 (65.3%)	103 (56.3%)	199 (71.6%)
Female	166 (34.7%)	80 (43.7%)	79 (28.4%)
Education			
Primary or less	135 (28.2%)	64 (35.0%)	90 (32.4%)
Secondary	136 (28.5%)	78 (42.6%)	120 (43.2%)
College	207 (43.3%)	41 (22.4%)	68 (24.5%)
BMI			
Underweight (≤ 18.5)	28 (5.9%)	8 (4.4%)	19 (6.8%)
Normal Weight (18.6 - 24.9)	303 (63.4%)	113 (61.7%)	183 (65.8%)
Overweight (25.0 - 29.9)	114 (23.8%)	42 (23.0%)	50 (18.0%)
Obese (≥ 30)	33 (6.9%)	20 (10.9%)	26 (9.4%)

Original table of cases/controls

```

rndr <- function(x, name, ...) {
  if (!is.numeric(x)) return(render.categorical.default(x))
  what <- switch(name,
    ltl = "Mean (SD)",
    age_at_iview = c("Mean (SD)", "Median [Min, Max]"),
  )
  parse.abbrev.render.code(c("", what))(x)
}

tbl <- table1(~ ltl + age_at_iview + msex + educ_ord + Set | is_case_f,
  data=tel_all,
  caption = "Demographics by Case Status",
  render=rndr
)

tbl

```

Table 6: Demographics by Case Status

	Case	Control
	(N=461)	(N=478)
Log Telomere Length		
Mean (SD)	-0.0163 (0.226)	-0.0531 (0.226)
Age		
Mean (SD)	37.7 (7.87)	35.7 (7.70)
Sex		
Male	302 (65.5%)	312 (65.3%)
Female	159 (34.5%)	166 (34.7%)
Education		
Primary or less	154 (33.4%)	135 (28.2%)
Secondary	198 (43.0%)	136 (28.5%)
College	109 (23.6%)	207 (43.3%)
BMI		
Normal Weight (18.6 - 24.9)	296 (64.2%)	303 (63.4%)
Underweight (≤ 18.5)	27 (5.9%)	28 (5.9%)
Overweight (25.0 - 29.9)	92 (20.0%)	114 (23.8%)
Obese (≥ 30)	46 (10.0%)	33 (6.9%)
Alcohol Use		
Never Users	225 (48.8%)	198 (41.4%)
Irregular Users	197 (42.7%)	185 (38.7%)
Regular Users	39 (8.5%)	95 (19.9%)
Khat Use		
Never Users	247 (53.6%)	398 (83.3%)
Irregular Users	134 (29.1%)	52 (10.9%)
Regular Users	80 (17.4%)	28 (5.9%)
Cannabis Use		
Never Users	427 (92.6%)	472 (98.7%)
Irregular Users	30 (6.5%)	6 (1.3%)
Regular Users	4 (0.9%)	0 (0%)
Tobacco Use		
Never Users	288 (62.5%)	478 (100%)
Irregular Users	107 (23.2%)	0 (0%)
Regular Users	66 (14.3%)	0 (0%)

Table 7: Demographics by Case Status

	Control	Healthy Case	Unhealthy Case
	(N=478)	(N=183)	(N=278)
cidi_q1			
0	478 (100%)	183 (100%)	241 (86.7%)
1	0 (0%)	0 (0%)	37 (13.3%)
cidi_q2			
0	478 (100%)	183 (100%)	225 (80.9%)
1	0 (0%)	0 (0%)	53 (19.1%)
cidi_q3			
0	478 (100%)	183 (100%)	202 (72.7%)
1	0 (0%)	0 (0%)	76 (27.3%)
cidi_q4			
0	478 (100%)	183 (100%)	264 (95.0%)
1	0 (0%)	0 (0%)	14 (5.0%)
cidi_q5			
0	478 (100%)	183 (100%)	255 (91.7%)
1	0 (0%)	0 (0%)	23 (8.3%)
cidi_q6			
0	478 (100%)	183 (100%)	277 (99.6%)
1	0 (0%)	0 (0%)	1 (0.4%)
cidi_q7			
0	478 (100%)	183 (100%)	268 (96.4%)
1	0 (0%)	0 (0%)	10 (3.6%)
cidi_q8			
0	478 (100%)	183 (100%)	272 (97.8%)
1	0 (0%)	0 (0%)	6 (2.2%)
cidi_q9			
0	478 (100%)	183 (100%)	257 (92.4%)
1	0 (0%)	0 (0%)	21 (7.6%)
cidi_q10			
0	478 (100%)	183 (100%)	271 (97.5%)
1	0 (0%)	0 (0%)	7 (2.5%)
cidi_q11			
0	478 (100%)	183 (100%)	253 (91.0%)
1	0 (0%)	0 (0%)	25 (9.0%)
cidi_q12			
0	478 (100%)	183 (100%)	272 (97.8%)
1	0 (0%)	0 (0%)	6 (2.2%)
cidi_q13			
0	478 (100%)	183 (100%)	269 (96.8%)
1	0 (0%)	0 (0%)	9 (3.2%)
cidi_q14			
0	478 (100%)	183 (100%)	245 (88.1%)
1	0 (0%)	0 (0%)	33 (11.9%)
HIV/AIDS			
0	478 (100%)	177 (96.7%)	246 (88.5%)
1	0 (0%)	0 (0%)	9 (3.2%)
777	0 (0%)	6 (3.3%)	23 (8.3%)
cidi_q16			
0	478 (100%)	183 (100%)	276 (99.3%)
1	0 (0%)	0 (0%)	2 (0.7%)
Cancer			
0	478 (100%)	179 (97.8%)	275 (98.9%)
777	0 (0%)	4 (2.2%)	3 (1.1%)

Table 8: Demographics by Case Status

	Case	Control	Overall
	(N=487)	(N=478)	(N=965)
Log Telomere Length			
Mean (SD)	-0.0209 (0.225)	-0.0531 (0.226)	-0.0368 (0.226)
Age			
Mean (SD)	37.8 (7.88)	35.7 (7.70)	36.8 (7.85)
Median [Min, Max]	36.0 [25.0, 55.0]	34.0 [25.0, 55.0]	35.0 [25.0, 55.0]
Sex			
Male	317 (65.1%)	312 (65.3%)	629 (65.2%)
Female	170 (34.9%)	166 (34.7%)	336 (34.8%)
Education			
Primary or less	161 (33.1%)	135 (28.2%)	296 (30.7%)
Secondary	212 (43.5%)	136 (28.5%)	348 (36.1%)
College	114 (23.4%)	207 (43.3%)	321 (33.3%)
Set			
A	118 (24.2%)	113 (23.6%)	231 (23.9%)
B	118 (24.2%)	117 (24.5%)	235 (24.4%)
C	124 (25.5%)	114 (23.8%)	238 (24.7%)
D	116 (23.8%)	120 (25.1%)	236 (24.5%)
E	11 (2.3%)	14 (2.9%)	25 (2.6%)

2.2.6 6 models with new data

Models: 1. Lm(Rosner LTL ~ age + case status (clean/unclean) 2. Lm(Rosner LTL ~ age + case status (clean/unclean) + sex 3. lm(Rosner LTL ~ Age + case + sex + education + BMI 4. lm(Rosner LTL ~ Age + case + sex + education + BMI + All Others 5. Model 4 + sex * case status interaction 6. Model 4 + education * case status interaction

Merge Ros to get other variables

```
ros_merged <- merge(ros, tel_all_nomissing)
```

Model 1

```
model1 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged)
#summary(model)

tbl_regression(model1) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.001	0.00, 0.00	0.028*
is_case_clean				

Control	—	—	—	
Healthy Case	0.07	0.019	0.03, 0.10	<0.001***
Unhealthy Case	0.02	0.016	-0.01, 0.05	0.2

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 2

```
model2 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged)
#summary(model)

tbl_regression(model2) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.001	0.00, 0.00	0.017*
is_case_clean				
Control	—	—	—	
Healthy Case	0.06	0.019	0.03, 0.10	<0.001***
Unhealthy Case	0.03	0.016	-0.01, 0.06	0.12
msex				
Male	—	—	—	
Female	0.04	0.015	0.01, 0.07	0.005**

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin, ros_merged)

tbl_regression(model3) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.001	0.00, 0.00	0.006**
is_case_clean				
Control	—	—	—	
Healthy Case	0.06	0.019	0.02, 0.09	0.002**
Unhealthy Case	0.02	0.016	-0.01, 0.05	0.3

msex					
Male	—	—	—		
Female	0.03	0.015	0.00, 0.06		0.030*
educ_ord					
Primary or less	—	—	—		
Secondary	-0.03	0.017	-0.06, 0.01		0.15
College	-0.06	0.018	-0.09, -0.02		0.002**
bmi_bin					
Normal Weight (18.6 - 24.9)	—	—	—		
Underweight (<= 18.5)	0.04	0.030	-0.02, 0.10		0.2
Overweight (25.0 - 29.9)	0.00	0.018	-0.04, 0.03		>0.9
Obese (>= 30)	0.02	0.027	-0.03, 0.07		0.4

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 4

```
model4 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
  alcohol_3cat + tobacco_3cat +
  cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
  cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
  cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
  cidi_q17, ros_merged)

tbl_regression(model4) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.001	0.00, 0.00	0.022*
is_case_clean				
Control	—	—	—	
Healthy Case	0.07	0.020	0.03, 0.11	<0.001***
Unhealthy Case	0.03	0.035	-0.04, 0.10	0.4
msex				
Male	—	—	—	
Female	0.04	0.017	0.01, 0.07	0.021*
educ_ord				
Primary or less	—	—	—	
Secondary	-0.03	0.018	-0.06, 0.01	0.15
College	-0.05	0.018	-0.08, -0.01	0.008**
bmi_bin				
Normal Weight (18.6 - 24.9)	—	—	—	
Underweight (<= 18.5)	0.03	0.030	-0.03, 0.09	0.3

Overweight (25.0 - 29.9)	0.00	0.018	-0.04, 0.03	0.8
Obese (≥ 30)	0.02	0.027	-0.03, 0.07	0.5
khat_3cat				
Never Users	—	—	—	
Irregular Users	-0.01	0.021	-0.05, 0.03	0.7
Regular Users	-0.01	0.027	-0.06, 0.05	0.8
alcohol_3cat				
Never Users	—	—	—	
Irregular Users	0.00	0.016	-0.03, 0.04	0.8
Regular Users	0.03	0.023	-0.01, 0.08	0.2
cannabis_3cat				
Never Users	—	—	—	
Irregular Users	-0.03	0.040	-0.11, 0.05	0.5
Regular Users	-0.08	0.110	-0.30, 0.14	0.5
tobacco_3cat				
Never Users	—	—	—	
Irregular Users	0.02	0.038	-0.06, 0.09	0.7
Regular Users	0.00	0.043	-0.09, 0.08	>0.9
cidi_q1				
cidi_q10	—	—	—	
cidi_q11	0.01	0.042	-0.08, 0.09	0.9
cidi_q2				
0	—	—	—	
1	-0.02	0.035	-0.09, 0.05	0.5
cidi_q3				
0	—	—	—	
1	0.02	0.032	-0.04, 0.08	0.5
cidi_q4				
0	—	—	—	
1	0.18	0.064	0.05, 0.30	0.006**
cidi_q5				
0	—	—	—	
1	-0.05	0.049	-0.15, 0.05	0.3
cidi_q6				
0	—	—	—	
1	-0.17	0.234	-0.63, 0.29	0.5
cidi_q7				
0	—	—	—	
1	0.17	0.076	0.02, 0.32	0.028*
cidi_q8				
0	—	—	—	
1	-0.03	0.094	-0.22, 0.15	0.7
cidi_q9				
0	—	—	—	
1	-0.01	0.054	-0.12, 0.09	0.8

cidi_q10					
0	—	—	—		
1	-0.22	0.087	-0.39, -0.05		0.011*
cidi_q11					
0	—	—	—		
1	0.05	0.047	-0.05, 0.14		0.3
cidi_q12					
0	—	—	—		
1	0.11	0.091	-0.07, 0.29		0.2
cidi_q13					
0	—	—	—		
1	-0.03	0.078	-0.18, 0.13		0.7
cidi_q14					
0	—	—	—		
1	-0.04	0.042	-0.13, 0.04		0.3
cidi_q15					
0	—	—	—		
1	-0.18	0.077	-0.33, -0.03		0.022*
777	-0.02	0.041	-0.11, 0.06		0.5
cidi_q16					
0	—	—	—		
1	-0.18	0.154	-0.48, 0.13		0.3
cidi_q17					
0	—	—	—		
777	-0.15	0.081	-0.31, 0.00		0.057

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 5

```
model5 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
  alcohol_3cat + tobacco_3cat +
  cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
  cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
  cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
  cidi_q17 + msex*is_case_clean, ros_merged)

tbl_regression(model5) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.001	0.00, 0.00	0.022*
is_case_clean				

Control	—	—	—	
Healthy Case	0.07	0.025	0.03, 0.12	0.003**
Unhealthy Case	0.01	0.044	-0.08, 0.10	0.8
msex				
Male	—	—	—	
Female	0.04	0.022	0.00, 0.08	0.074
educ_ord				
Primary or less	—	—	—	
Secondary	-0.03	0.018	-0.06, 0.01	0.2
College	-0.05	0.018	-0.08, -0.01	0.008**
bmi_bin				
Normal Weight (18.6 - 24.9)	—	—	—	
Underweight (≤ 18.5)	0.03	0.030	-0.03, 0.09	0.3
Overweight (25.0 - 29.9)	0.00	0.018	-0.04, 0.03	0.9
Obese (≥ 30)	0.02	0.027	-0.03, 0.07	0.5
khat_3cat				
Never Users	—	—	—	
Irregular Users	-0.01	0.021	-0.05, 0.03	0.6
Regular Users	-0.01	0.028	-0.06, 0.05	0.8
alcohol_3cat				
Never Users	—	—	—	
Irregular Users	0.01	0.016	-0.03, 0.04	0.7
Regular Users	0.03	0.023	-0.01, 0.08	0.2
cannabis_3cat				
Never Users	—	—	—	
Irregular Users	-0.03	0.040	-0.11, 0.05	0.5
Regular Users	-0.08	0.110	-0.30, 0.14	0.5
tobacco_3cat				
Never Users	—	—	—	
Irregular Users	0.03	0.044	-0.05, 0.12	0.5
Regular Users	0.02	0.049	-0.08, 0.11	0.8
cidi_q1				
cidi_q10	—	—	—	
cidi_q11	0.01	0.042	-0.08, 0.09	0.9
cidi_q2				
0	—	—	—	
1	-0.02	0.035	-0.09, 0.05	0.5
cidi_q3				
0	—	—	—	
1	0.02	0.032	-0.05, 0.08	0.6
cidi_q4				
0	—	—	—	
1	0.18	0.064	0.05, 0.31	0.005**
cidi_q5				
0	—	—	—	

1	-0.05	0.049	-0.15, 0.05	0.3
cidi_q6	—	—	—	
0				
1	-0.17	0.234	-0.63, 0.29	0.5
cidi_q7	—	—	—	
0				
1	0.17	0.076	0.02, 0.32	0.024*
cidi_q8	—	—	—	
0				
1	-0.04	0.094	-0.22, 0.15	0.7
cidi_q9	—	—	—	
0				
1	-0.01	0.055	-0.12, 0.09	0.8
cidi_q10	—	—	—	
0				
1	-0.22	0.087	-0.39, -0.05	0.012*
cidi_q11	—	—	—	
0				
1	0.05	0.047	-0.05, 0.14	0.3
cidi_q12	—	—	—	
0				
1	0.13	0.093	-0.06, 0.31	0.2
cidi_q13	—	—	—	
0				
1	-0.02	0.078	-0.18, 0.13	0.8
cidi_q14	—	—	—	
0				
1	-0.04	0.042	-0.12, 0.04	0.3
cidi_q15	—	—	—	
0				
1	-0.19	0.079	-0.34, -0.03	0.017*
777	-0.03	0.041	-0.11, 0.06	0.5
cidi_q16	—	—	—	
0				
1	-0.18	0.154	-0.48, 0.12	0.2
cidi_q17	—	—	—	
0				
777	-0.15	0.081	-0.31, 0.01	0.063
is_case_clean * msex				
Healthy Case * Female	-0.01	0.038	-0.09, 0.06	0.7
Unhealthy Case * Female	0.03	0.045	-0.06, 0.12	0.5

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 6


```

model6 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
  alcohol_3cat + tobacco_3cat +
  cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
  cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
  cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
  cidi_q17 + educ_ord*is_case_clean, ros_merged)

tbl_regression(model6) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)

```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.001	0.00, 0.00	0.020*
is_case_clean				
Control	—	—	—	
Healthy Case	0.05	0.033	-0.02, 0.11	0.2
Unhealthy Case	0.03	0.043	-0.05, 0.12	0.4
msex				
Male	—	—	—	
Female	0.04	0.017	0.01, 0.07	0.024*
educ_ord				
Primary or less	—	—	—	
Secondary	-0.03	0.026	-0.08, 0.02	0.3
College	-0.06	0.024	-0.10, -0.01	0.020*
bmi_bin				
Normal Weight (18.6 - 24.9)	—	—	—	
Underweight (<= 18.5)	0.03	0.030	-0.03, 0.09	0.3
Overweight (25.0 - 29.9)	0.00	0.018	-0.04, 0.03	>0.9
Obese (>= 30)	0.02	0.027	-0.03, 0.07	0.4
khat_3cat				
Never Users	—	—	—	
Irregular Users	-0.01	0.021	-0.05, 0.03	0.7
Regular Users	-0.01	0.028	-0.06, 0.05	0.8
alcohol_3cat				
Never Users	—	—	—	
Irregular Users	0.00	0.016	-0.03, 0.04	0.8
Regular Users	0.03	0.023	-0.01, 0.07	0.2
cannabis_3cat				
Never Users	—	—	—	
Irregular Users	-0.03	0.040	-0.11, 0.05	0.4
Regular Users	-0.07	0.111	-0.29, 0.15	0.5
tobacco_3cat				
Never Users	—	—	—	

Irregular Users	0.01	0.039	-0.06, 0.09	0.7
Regular Users	0.00	0.043	-0.09, 0.08	>0.9
cidi_q1	—	—	—	
cidi_q10	—	—	—	
cidi_q11	0.01	0.043	-0.08, 0.09	0.9
cidi_q2	—	—	—	
0	—	—	—	
1	-0.03	0.035	-0.09, 0.04	0.5
cidi_q3	—	—	—	
0	—	—	—	
1	0.02	0.032	-0.05, 0.08	0.6
cidi_q4	—	—	—	
0	—	—	—	
1	0.18	0.065	0.05, 0.30	0.007**
cidi_q5	—	—	—	
0	—	—	—	
1	-0.05	0.050	-0.15, 0.05	0.3
cidi_q6	—	—	—	
0	—	—	—	
1	-0.17	0.235	-0.63, 0.29	0.5
cidi_q7	—	—	—	
0	—	—	—	
1	0.17	0.076	0.02, 0.32	0.029*
cidi_q8	—	—	—	
0	—	—	—	
1	-0.03	0.094	-0.22, 0.15	0.7
cidi_q9	—	—	—	
0	—	—	—	
1	-0.01	0.055	-0.12, 0.10	0.8
cidi_q10	—	—	—	
0	—	—	—	
1	-0.23	0.088	-0.40, -0.06	0.010**
cidi_q11	—	—	—	
0	—	—	—	
1	0.05	0.047	-0.04, 0.14	0.3
cidi_q12	—	—	—	
0	—	—	—	
1	0.12	0.092	-0.06, 0.30	0.2
cidi_q13	—	—	—	
0	—	—	—	
1	-0.02	0.078	-0.17, 0.13	0.8
cidi_q14	—	—	—	
0	—	—	—	
1	-0.04	0.042	-0.13, 0.04	0.3
cidi_q15	—	—	—	

0	—	—	—	
1	-0.17	0.078	-0.33, -0.02	0.027*
777	-0.02	0.042	-0.10, 0.06	0.6
cidi_q16				
0	—	—	—	
1	-0.18	0.155	-0.48, 0.12	0.2
cidi_q17				
0	—	—	—	
777	-0.16	0.081	-0.32, 0.00	0.054
is_case_clean * educ_ord				
Healthy Case * Secondary	0.03	0.044	-0.06, 0.12	0.5
Unhealthy Case * Secondary	-0.02	0.041	-0.10, 0.06	0.7
Healthy Case * College	0.03	0.049	-0.06, 0.13	0.5
Unhealthy Case * College	0.00	0.044	-0.08, 0.09	>0.9

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

```
nobs(model11)
```

```
## [1] 939
```

Create table with values

```
#summary_table <- data.frame()

row_names <- c("Age", "Filtered Case", "Unfiltered Case")
summary_table <- data.frame(matrix(NA, nrow = length(row_names), ncol = 0))
for (i in 1:6) {

  model_name <- paste0("model",i)

  model_name <- eval(parse(text = model_name))

  beta_age <- round(coef(model_name)["age_at_iview"],3)
  beta_filt <- round(coef(model_name)["is_case_cleanFiltered Case"],3)
  beta_unfilt <- round(coef(model_name)["is_case_cleanUnfiltered Case"], 3)

  ci_age_ll <- round(confint(model_name)["age_at_iview", 1], 3)
  ci_age_ul <- round(confint(model_name)["age_at_iview", 2], 3)
  ci_filt_ll <- round(confint(model_name)["is_case_cleanHealthy Case", 1],3)
  ci_filt_ul <- round(confint(model_name)["is_case_cleanHealthy Case", 2],3)
  ci_unfilt_ll <- round(confint(model_name)["is_case_cleanUnhealthy Case", 1],3)
  ci_unfilt_ul <- round(confint(model_name)["is_case_cleanUnhealthy Case", 2],3)

  df <- data.frame(
    beta = c(beta_age, beta_filt, beta_unfilt),
    ci_ll = c(ci_age_ll, ci_filt_ll, ci_unfilt_ll),
    ci_ul = c(ci_age_ul, ci_filt_ul, ci_unfilt_ul))
}
```

```

df <- df %>%
  mutate( together = paste0(beta," (",ci_ll, ", ", ci_ul,")")) %>%
  select(together) %>%
  `rownames<-`(c("Beta Age", "Beta Filtered", "Beta Unfiltered"))

  # Assign new row names
  rownames(df) <- row_names

summary_table <- cbind(df, summary_table)
}

col_names <- c("Model 1", "Model 2", "Model 3", "Model 4", "Model 5", "Model 6")
colnames(summary_table) <- col_names

# Function to create new dataframe
insertRow <- function(data, new_row, r) {
  data_new <- rbind(data[1:r, ],
                    new_row,
                    data[-(1:r), ])
  rownames(data_new) <- 1:nrow(data_new)
  return(data_new)
}

index <- 1
newrow <- c(rep("Ref",6))
summary_table_new=insertRow(summary_table, newrow, index)

# index <- 0
# nob = c(rep(945, 6))
# summary_table_new=insertRow(summary_table, nob, 0)

row_names <- c("Age", "Control","Filtered Case", "Unfiltered Case")
rownames(summary_table_new) <- row_names

kable(summary_table_new)

```

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Age	-0.002 (-0.004, 0)	-0.002 (-0.004, 0)	-0.002 (-0.004, 0)	-0.002 (-0.004, -0.001)	-0.002 (-0.004, 0)	-0.002 (-0.004, 0)
Control	Ref	Ref	Ref	Ref	Ref	Ref
Filtered Case	NA (-0.018, 0.111)	NA (0.025, 0.124)	NA (0.03, 0.107)	NA (0.021, 0.095)	NA (0.028, 0.101)	NA (0.032, 0.105)
Unfiltered Case	NA (-0.051, 0.119)	NA (-0.075, 0.096)	NA (-0.04, 0.096)	NA (-0.015, 0.05)	NA (-0.007, 0.057)	NA (-0.01, 0.054)

2.2.7 Analysis of sets

```
ros_merged_A <- ros_merged %>% filter(Set == "A")
```

2.2.7.1 Set A Model 1

```
model1 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_A)
#summary(model)

tbl_regression(model1) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.038*
is_case_clean				
Control	—	—	—	
Healthy Case	0.02	0.034	-0.05, 0.09	0.5
Unhealthy Case	0.02	0.032	-0.04, 0.08	0.5

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 2

```
model2 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged_A)
#summary(model)

tbl_regression(model2) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.029*
is_case_clean				
Control	—	—	—	
Healthy Case	0.02	0.034	-0.05, 0.09	0.6
Unhealthy Case	0.02	0.032	-0.04, 0.09	0.5
msex				
Male	—	—	—	
Female	0.04	0.027	-0.02, 0.09	0.2

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
             educ_ord + bmi_bin, ros_merged_A)

tbl_regression(model3) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.024*
is_case_clean				
Control	—	—	—	
Healthy Case	0.01	0.034	-0.05, 0.08	0.7
Unhealthy Case	0.02	0.032	-0.05, 0.08	0.6
msex				
Male	—	—	—	
Female	0.02	0.029	-0.04, 0.08	0.5
educ_ord				
Primary or less	—	—	—	
Secondary	-0.02	0.033	-0.09, 0.04	0.5
College	-0.08	0.035	-0.14, -0.01	0.034*
bmi_bin				
Normal Weight (18.6 - 24.9)	—	—	—	
Underweight (<= 18.5)	0.05	0.071	-0.09, 0.19	0.5
Overweight (25.0 - 29.9)	-0.07	0.034	-0.14, 0.00	0.048*
Obese (>= 30)	0.06	0.043	-0.03, 0.14	0.2

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 4

```
#
# model4 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
#             educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
#             alcohol_3cat + tobacco_3cat +
#             cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
#             cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
#             cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
#             cidi_q17, ros_merged_A)
#
# tbl_regression(model4) %>% add_significance_stars(
#   hide_p = FALSE, hide_ci = FALSE,
#   pattern = "{p.value}{stars}"
# )
```

Create table with values

```

#summary_table <- data.frame()

row_names <- c("Age", "Filtered Case", "Unfiltered Case")
summary_table <- data.frame(matrix(NA, nrow = length(row_names), ncol = 0))
for (i in 1:3) {

  model_name <- paste0("model",i)

  model_name <- eval(parse(text = model_name))

  beta_age <- round(coef(model_name)["age_at_iview"],3)
  beta_filt <- round(coef(model_name)["is_case_cleanFiltered Case"],3)
  beta_unfilt <- round(coef(model_name)["is_case_cleanUnfiltered Case"], 3)

  ci_age_ll <- round(confint(model_name)["age_at_iview", 1], 3)
  ci_age_ul <- round(confint(model_name)["age_at_iview", 2], 3)
  ci_filt_ll <- round(confint(model_name)["is_case_cleanHealthy Case", 1],3)
  ci_filt_ul <- round(confint(model_name)["is_case_cleanHealthy Case", 2],3)
  ci_unfilt_ll <- round(confint(model_name)["is_case_cleanUnhealthy Case", 1],3)
  ci_unfilt_ul <- round(confint(model_name)["is_case_cleanUnhealthy Case", 2],3)

  df <- data.frame(
    beta = c(beta_age, beta_filt, beta_unfilt),
    ci_ll = c(ci_age_ll, ci_filt_ll, ci_unfilt_ll),
    ci_ul = c(ci_age_ul, ci_filt_ul, ci_unfilt_ul))

  df <- df %>%
    mutate( together = paste0(beta, " (",ci_ll, ", ", ", ci_ul,")")) %>%
    select(together) %>%
    `rownames<-`(c("Beta Age", "Beta Filtered", "Beta Unfiltered"))

  # Assign new row names
  rownames(df) <- row_names

  summary_table <- cbind(df, summary_table)
}

col_names <- c("Model 1", "Model 2", "Model 3")
colnames(summary_table) <- col_names

# Function to create new dataframe
insertRow <- function(data, new_row, r) {
  data_new <- rbind(data[1:r, ],
                    new_row,
                    data[-(1:r), ])
  rownames(data_new) <- 1:nrow(data_new)
  return(data_new)
}

index <- 1

```

```

newrow <- c(rep("Ref",3))
summary_table_new=insertRow(summary_table, newrow, index)

# index <- 0
# nobs = c(rep(945, 6))
# summary_table_new=insertRow(summary_table, nobs, 0)

row_names <- c("Age", "Control","Filtered Case", "Unfiltered Case")
rownames(summary_table_new) <- row_names

kable(summary_table_new, caption="Models for Set A (n=227)",
)

```

Table 19: Models for Set A (n=227)

	Model 1	Model 2	Model 3
Age	-0.004 (-0.007, -0.001)	-0.004 (-0.007, 0)	-0.004 (-0.007, 0)
Control	Ref	Ref	Ref
Filtered Case	NA (-0.053, 0.083)	NA (-0.049, 0.085)	NA (-0.045, 0.089)
Unfiltered Case	NA (-0.048, 0.079)	NA (-0.039, 0.086)	NA (-0.043, 0.081)

```

ros_merged_B <- ros_merged %>% filter(Set == "B")

```

2.2.7.2 Set B Model 1

```

model1 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_B)
#summary(model)

tbl_regression(model1) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)

```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.062
is_case_clean				
Control	—	—	—	
Healthy Case	0.09	0.040	0.01, 0.17	0.022*
Unhealthy Case	0.06	0.031	0.00, 0.12	0.046*

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 2


```

model2 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged_B)
#summary(model)

tbl_regression(model2) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)

```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.040*
is_case_clean				
Control	—	—	—	
Healthy Case	0.09	0.040	0.01, 0.17	0.025*
Unhealthy Case	0.07	0.031	0.01, 0.13	0.030*
msex				
Male	—	—	—	
Female	0.05	0.031	-0.01, 0.12	0.080

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```

model3 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin, ros_merged_B)

tbl_regression(model3) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)

```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.024*
is_case_clean				
Control	—	—	—	
Healthy Case	0.08	0.041	0.00, 0.17	0.044*
Unhealthy Case	0.06	0.032	0.00, 0.13	0.049*
msex				
Male	—	—	—	
Female	0.06	0.033	-0.01, 0.12	0.085
educ_ord				
Primary or less	—	—	—	
Secondary	-0.03	0.034	-0.10, 0.04	0.4
College	-0.05	0.035	-0.12, 0.02	0.2
bmi_bin				

Normal Weight (18.6 - 24.9)	—	—	—	
Underweight (≤ 18.5)	0.08	0.047	-0.02, 0.17	0.11
Overweight (25.0 - 29.9)	0.00	0.041	-0.08, 0.08	>0.9
Obese (≥ 30)	0.02	0.058	-0.10, 0.13	0.8

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 4

```
# model4 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
#             educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
#             alcohol_3cat + tobacco_3cat +
#             cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
#             cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
#             cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
#             cidi_q17, ros_merged_B)
#
# tbl_regression(model4) %>% add_significance_stars(
#   hide_p = FALSE, hide_ci = FALSE,
#   pattern = "{p.value}{stars}"
# )
```

Create table with values

```
#summary_table <- data.frame()

row_names <- c("Age", "Filtered Case", "Unfiltered Case")
summary_table <- data.frame(matrix(NA, nrow = length(row_names), ncol = 0))
for (i in 1:3) {

  model_name <- paste0("model",i)

  model_name <- eval(parse(text = model_name))

  beta_age <- round(coef(model_name)["age_at_iview"],3)
  beta_filt <- round(coef(model_name)["is_case_cleanFiltered Case"],3)
  beta_unfilt <- round(coef(model_name)["is_case_cleanUnfiltered Case"], 3)

  ci_age_ll <- round(confint(model_name)["age_at_iview", 1], 3)
  ci_age_ul <- round(confint(model_name)["age_at_iview", 2], 3)
  ci_filt_ll <- round(confint(model_name)["is_case_cleanHealthy Case", 1],3)
  ci_filt_ul <- round(confint(model_name)["is_case_cleanHealthy Case", 2],3)
  ci_unfilt_ll <- round(confint(model_name)["is_case_cleanUnhealthy Case", 1],3)
  ci_unfilt_ul <- round(confint(model_name)["is_case_cleanUnhealthy Case", 2],3)

  df <- data.frame(
    beta = c(beta_age, beta_filt, beta_unfilt),
    ci_ll = c(ci_age_ll, ci_filt_ll, ci_unfilt_ll),
    ci_ul = c(ci_age_ul, ci_filt_ul, ci_unfilt_ul))

  df <- df %>%
    mutate(toString = paste0(beta, " (",ci_ll, ", ", ci_ul,")")) %>%
```

```

select(together) %>%
`rownames<-`(c("Beta Age", "Beta Filtered", "Beta Unfiltered"))

# Assign new row names
rownames(df) <- row_names

summary_table <- cbind(df, summary_table)
}

col_names <- c("Model 1", "Model 2", "Model 3")
colnames(summary_table) <- col_names

# Function to create new dataframe
insertRow <- function(data, new_row, r) {
  data_new <- rbind(data[1:r, ],
                    new_row,
                    data[-(1:r), ])
  rownames(data_new) <- 1:nrow(data_new)
  return(data_new)
}

index <- 1
newrow <- c(rep("Ref",3))
summary_table_new=insertRow(summary_table, newrow, index)

# index <- 0
# nobs = c(rep(945, 6))
# summary_table_new=insertRow(summary_table, nobs, 0)

row_names <- c("Age", "Control", "Filtered Case", "Unfiltered Case")
rownames(summary_table_new) <- row_names

kable(summary_table_new, caption="Models for Set B (n=229)")

```

Table 23: Models for Set B (n=229)

	Model 1	Model 2	Model 3
Age	-0.004 (-0.008, -0.001)	-0.004 (-0.008, 0)	-0.004 (-0.007, 0)
Control	Ref	Ref	Ref
Filtered Case	NA (0.002, 0.166)	NA (0.011, 0.169)	NA (0.013, 0.172)
Unfiltered Case	NA (0, 0.126)	NA (0.007, 0.13)	NA (0.001, 0.124)

```

ros_merged_C <- ros_merged %>% filter(Set == "C")

```

2.2.7.3 Set C Model 1

```
model1 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_C)
#summary(model)

tbl_regression(model1) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	0.00, 0.00	0.6
is_case_clean				
Control	—	—	—	
Healthy Case	0.06	0.035	-0.01, 0.13	0.090
Unhealthy Case	0.03	0.033	-0.03, 0.10	0.3

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 2

```
model2 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged_C)
#summary(model)

tbl_regression(model2) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	0.00, 0.00	0.6
is_case_clean				
Control	—	—	—	
Healthy Case	0.06	0.035	-0.01, 0.13	0.10
Unhealthy Case	0.03	0.033	-0.04, 0.09	0.4
msex				
Male	—	—	—	
Female	0.03	0.029	-0.03, 0.08	0.3

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin, ros_merged_C)
```

```
tbl_regression(model3) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.4
is_case_clean				
Control	—	—	—	
Healthy Case	0.06	0.036	-0.02, 0.13	0.13
Unhealthy Case	0.02	0.034	-0.05, 0.09	0.5
msex				
Male	—	—	—	
Female	0.03	0.030	-0.03, 0.09	0.3
educ_ord				
Primary or less	—	—	—	
Secondary	-0.02	0.035	-0.09, 0.05	0.6
College	-0.06	0.035	-0.13, 0.01	0.084
bmi_bin				
Normal Weight (18.6 - 24.9)	—	—	—	
Underweight (<= 18.5)	-0.03	0.066	-0.16, 0.10	0.7
Overweight (25.0 - 29.9)	0.02	0.033	-0.04, 0.09	0.5
Obese (>= 30)	-0.05	0.057	-0.16, 0.06	0.4

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 4

```
# model4 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
#               educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
#               alcohol_3cat + tobacco_3cat +
#               cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
#               cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
#               cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
#               cidi_q17, ros_merged_B)
#
# tbl_regression(model4) %>% add_significance_stars(
#   hide_p = FALSE, hide_ci = FALSE,
#   pattern = "{p.value}{stars}"
# )
```

Create table with values

```
#summary_table <- data.frame()

row_names <- c("Age", "Filtered Case", "Unfiltered Case")
summary_table <- data.frame(matrix(NA, nrow = length(row_names), ncol = 0))
```

```

for (i in 1:3) {

  model_name <- paste0("model",i)

  model_name <- eval(parse(text = model_name))

  beta_age <- round(coef(model_name)["age_at_iview"],3)
  beta_filt <- round(coef(model_name)["is_case_cleanFiltered Case"],3)
  beta_unfilt <- round(coef(model_name)["is_case_cleanUnfiltered Case"], 3)

  ci_age_ll <- round(confint(model_name)["age_at_iview", 1], 3)
  ci_age_ul <- round(confint(model_name)["age_at_iview", 2], 3)
  ci_filt_ll <- round(confint(model_name)["is_case_cleanHealthy Case", 1],3)
  ci_filt_ul <- round(confint(model_name)["is_case_cleanHealthy Case", 2],3)
  ci_unfilt_ll <- round(confint(model_name)["is_case_cleanUnhealthy Case", 1],3)
  ci_unfilt_ul <- round(confint(model_name)["is_case_cleanUnhealthy Case", 2],3)

  df <- data.frame(
    beta = c(beta_age, beta_filt, beta_unfilt),
    ci_ll = c(ci_age_ll, ci_filt_ll, ci_unfilt_ll),
    ci_ul = c(ci_age_ul, ci_filt_ul, ci_unfilt_ul))

  df <- df %>%
    mutate( together = paste0(beta, " (",ci_ll, ", ", ci_ul,")")) %>%
    select(together) %>%
    `rownames<-`(c("Beta Age", "Beta Filtered", "Beta Unfiltered"))

    # Assign new row names
    rownames(df) <- row_names

  summary_table <- cbind(df, summary_table)
}

col_names <- c("Model 1", "Model 2", "Model 3")
colnames(summary_table) <- col_names

# Function to create new dataframe
insertRow <- function(data, new_row, r) {
  data_new <- rbind(data[1:r, ],
                    new_row,
                    data[- (1:r), ])
  rownames(data_new) <- 1:nrow(data_new)
  return(data_new)
}

index <- 1
newrow <- c(rep("Ref",3))
summary_table_new=insertRow(summary_table, newrow, index)

# index <- 0

```

```
# nobs = c(rep(945, 6))
# summary_table_new=insertRow(summary_table, nobs, 0)

row_names <- c("Age", "Control", "Filtered Case", "Unfiltered Case")
rownames(summary_table_new) <- row_names

kable(summary_table_new, caption="Models for Set C (n=233)")
```

Table 27: Models for Set C (n=233)

	Model 1	Model 2	Model 3
Age	-0.002 (-0.005, 0.002)	-0.001 (-0.004, 0.002)	-0.001 (-0.004, 0.003)
Control	Ref	Ref	Ref
Filtered Case	NA (-0.016, 0.127)	NA (-0.011, 0.128)	NA (-0.009, 0.13)
Unfiltered Case	NA (-0.046, 0.09)	NA (-0.035, 0.094)	NA (-0.033, 0.096)

```
ros_merged_D <- ros_merged %>% filter(Set == "D")
```

2.2.7.4 Set D Model 1

```
model1 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_D)
#summary(model)

tbl_regression(model1) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.3
is_case_clean				
Control	—	—	—	
Healthy Case	0.08	0.041	0.00, 0.16	0.046*
Unhealthy Case	-0.01	0.035	-0.08, 0.06	0.8

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 2

```
model2 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged_D)
#summary(model)

tbl_regression(model2) %>% add_significance_stars(
```

```
hide_p = FALSE, hide_ci = FALSE,
pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.3
is_case_clean				
Control	—	—	—	
Healthy Case	0.08	0.041	0.00, 0.16	0.063
Unhealthy Case	0.00	0.035	-0.07, 0.07	>0.9
msex				
Male	—	—	—	
Female	0.05	0.032	-0.02, 0.11	0.14

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
educ_ord + bmi_bin, ros_merged_D)

tbl_regression(model3) %>% add_significance_stars(
  hide_p = FALSE, hide_ci = FALSE,
  pattern = "{p.value}{stars}"
)
```

Characteristic	Beta	SE ¹	95% CI ¹	p-value ²
age_at_iview	0.00	0.002	-0.01, 0.00	0.3
is_case_clean				
Control	—	—	—	
Healthy Case	0.07	0.042	-0.01, 0.16	0.088
Unhealthy Case	0.00	0.036	-0.07, 0.07	>0.9
msex				
Male	—	—	—	
Female	0.04	0.034	-0.02, 0.11	0.2
educ_ord				
Primary or less	—	—	—	
Secondary	-0.02	0.037	-0.10, 0.05	0.5
College	-0.03	0.039	-0.11, 0.05	0.4
bmi_bin				
Normal Weight (18.6 - 24.9)	—	—	—	
Underweight (<= 18.5)	-0.01	0.073	-0.15, 0.14	>0.9
Overweight (25.0 - 29.9)	0.02	0.037	-0.06, 0.09	0.6

Obese (≥ 30)	-0.01	0.062	-0.13, 0.12	>0.9
---------------------	-------	-------	-------------	------

¹SE = Standard Error, CI = Confidence Interval

²*p<0.05; **p<0.01; ***p<0.001

Model 4

```
# model4 <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
#               educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
#               alcohol_3cat + tobacco_3cat +
#               cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
#               cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
#               cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
#               cidi_q17, ros_merged_B)
#
# tbl_regression(model4) %>% add_significance_stars(
#   hide_p = FALSE, hide_ci = FALSE,
#   pattern = "{p.value}{stars}"
# )
```

Create table with values

```
#summary_table <- data.frame()

row_names <- c("Age", "Filtered Case", "Unfiltered Case")
summary_table <- data.frame(matrix(NA, nrow = length(row_names), ncol = 0))
for (i in 1:3) {

  model_name <- paste0("model",i)

  model_name <- eval(parse(text = model_name))

  beta_age <- round(coef(model_name)["age_at_iview"],3)
  beta_filt <- round(coef(model_name)["is_case_cleanFiltered Case"],3)
  beta_unfilt <- round(coef(model_name)["is_case_cleanUnfiltered Case"], 3)

  ci_age_ll <- round(confint(model_name)["age_at_iview", 1], 3)
  ci_age_ul <- round(confint(model_name)["age_at_iview", 2], 3)
  ci_filt_ll <- round(confint(model_name)["is_case_cleanHealthy Case", 1],3)
  ci_filt_ul <- round(confint(model_name)["is_case_cleanHealthy Case", 2],3)
  ci_unfilt_ll <- round(confint(model_name)["is_case_cleanUnhealthy Case", 1],3)
  ci_unfilt_ul <- round(confint(model_name)["is_case_cleanUnhealthy Case", 2],3)

  df <- data.frame(
    beta = c(beta_age, beta_filt, beta_unfilt),
    ci_ll = c(ci_age_ll, ci_filt_ll, ci_unfilt_ll),
    ci_ul = c(ci_age_ul, ci_filt_ul, ci_unfilt_ul))

  df <- df %>%
    mutate( together = paste0(beta, " (",ci_ll, ", ", ci_ul,")")) %>%
    select(together) %>%
    `rownames<-`(c("Beta Age", "Beta Filtered", "Beta Unfiltered"))
```

```

# Assign new row names
rownames(df) <- row_names

summary_table <- cbind(df, summary_table)
}

col_names <- c("Model 1", "Model 2", "Model 3")
colnames(summary_table) <- col_names

# Function to create new dataframe
insertRow <- function(data, new_row, r) {
  data_new <- rbind(data[1:r, ],
                    new_row,
                    data[-(1:r), ])
  rownames(data_new) <- 1:nrow(data_new)
  return(data_new)
}

index <- 1
newrow <- c(rep("Ref",3))
summary_table_new=insertRow(summary_table, newrow, index)

# index <- 0
# nob = c(rep(945, 6))
# summary_table_new=insertRow(summary_table, nob, 0)

row_names <- c("Age", "Control", "Filtered Case", "Unfiltered Case")
rownames(summary_table_new) <- row_names

kable(summary_table_new, caption="Models for Set D (n=233)")

```

Table 31: Models for Set D (n=233)

	Model 1	Model 2	Model 3
Age	-0.002 (-0.006, 0.002)	-0.002 (-0.006, 0.002)	-0.002 (-0.006, 0.002)
Control	Ref	Ref	Ref
Filtered Case	NA (-0.011, 0.156)	NA (-0.004, 0.156)	NA (0.001, 0.161)
Unfiltered Case	NA (-0.072, 0.068)	NA (-0.072, 0.065)	NA (-0.078, 0.059)

3 Stratified analysis

3.1 By batch

3.1.1 Set ACE

```
ros_merged_ACE <- ros_merged %>% filter(Set %in% c("A", "C", "E"))
```

Model 1

```
model1_ACE <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_ACE)
```

Model 2

```
model2_ACE <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged_ACE)
```

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3_ACE <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +  
educ_ord + bmi_bin, ros_merged_ACE)
```

```
summary_table = NULL  
  
model_list <- c("model1_ACE", "model2_ACE", "model3_ACE")  
  
var <- c("age_at_iview", "control", "is_case_cleanHealthy Case",  
        "is_case_cleanUnhealthy Case")  
  
results <- NULL  
  
for (i in 1:length(model_list)) {  
  
  model <- get(model_list[i])  
  nobs <- nobs(model)  
  cont <- "Ref"  
  
  for (j in 1:length(var)) {  
  
    if (var[j] == "control") {  
  
      results<- rbind(results, c(model_list[i], var[j], nobs, rep("Ref", 3)))  
  
    } else {  
      beta <- round(coef(model)[var[j]], 3)  
  
      ci_ll <- round(confint(model)[var[j], 1], 3)  
      ci_ul <- round(confint(model)[var[j], 2], 3)  
      ci <- paste0("(", ci_ll, ", ", ci_ul, ")")  
  
      p <- round(summary(model)$coefficients[var[j], "Pr(>|t|)"], 3)  
  
      pval <- p  
      if(p<=0.05){pval<-paste(p, "*", sep="")}  
      if(p<=0.01){pval<-paste(p, "*", sep="")}  
      if(p<=0.001){pval<-paste(p, "*", sep="")}
```

```

    results <- rbind(results, c(model_list[i], var[j], nobs, beta, ci, pval))
  }

}

}

df <- as.data.frame(results)
colnames(df) <- c("Model", "Coefficient", "N", "Beta", "95% CI", "P")

kable(df, caption="Sets ACE")

```

Table 32: Sets ACE

Model	Coefficient	N	Beta	95% CI	P
model1_ACE	age_at_iview	480	-0.001	(-0.004, 0.001)	0.261
model1_ACE	control	480	Ref	Ref	Ref
model1_ACE	is_case_cleanHealthy Case	480	0.054	(0.005, 0.102)	0.03*
model1_ACE	is_case_cleanUnhealthy Case	480	0.017	(-0.027, 0.062)	0.451
model2_ACE	age_at_iview	480	-0.002	(-0.004, 0.001)	0.206
model2_ACE	control	480	Ref	Ref	Ref
model2_ACE	is_case_cleanHealthy Case	480	0.051	(0.003, 0.099)	0.038*
model2_ACE	is_case_cleanUnhealthy Case	480	0.018	(-0.026, 0.062)	0.426
model3_ACE	age_at_iview	480	-0.002	(-0.004, 0.001)	0.138
model3_ACE	control	480	Ref	Ref	Ref
model3_ACE	is_case_cleanHealthy Case	480	0.038	(-0.01, 0.087)	0.122
model3_ACE	is_case_cleanUnhealthy Case	480	0.004	(-0.041, 0.049)	0.864

3.1.2 Set B

```
ros_merged_B <- ros_merged %>% filter(Set == "B")
```

Model 1

```
model1_B <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_B)
```

Model 2

```
model2_B <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged_B)
```

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3_B <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin, ros_merged_B)
```

```
model_list <- c("model1_B", "model2_B", "model3_B")
```

```
var <- c("age_at_iview", "control", "is_case_cleanHealthy Case",
```

```

      "is_case_cleanUnhealthy Case")

results <- NULL

for (i in 1:length(model_list)) {

  model <- get(model_list[i])
  nobs <- nobs(model)
  cont <- "Ref"

  for (j in 1:length(var)) {

    if (var[j] == "control") {

      results<- rbind(results,c(model_list[i], var[j], nobs, rep("Ref",3)))

    } else {

      beta <- round(coef(model)[var[j]], 3)

      ci_ll <- round(confint(model)[var[j], 1],3)
      ci_ul <- round(confint(model)[var[j], 2],3)
      ci <- paste0("(", ci_ll, ", ", ci_ul, ")")

      p <- round(summary(model)$coefficients[var[j], "Pr(>|t|)"],3)

      pval <- p
      if(p<=0.05){pval<-paste(p, "*",sep="")}
      if(p<=0.01){pval<-paste(p, "**", sep="")}
      if(p<=0.001){pval<-paste(p, "***", sep="")}

      results <- rbind(results,c(model_list[i], var[j], nobs, beta, ci, pval))
    }

  }

}

df <- as.data.frame(results)
colnames(df)<-c("Model", "Coefficient", "N", "Beta", "95% CI", "P")

kable(df)

```

Model	Coefficient	N	Beta	95% CI	P
model1_B	age_at_iview	228	-0.004	(-0.007, 0)	0.062
model1_B	control	228	Ref	Ref	Ref
model1_B	is_case_cleanHealthy Case	228	0.093	(0.013, 0.172)	0.022*
model1_B	is_case_cleanUnhealthy Case	228	0.063	(0.001, 0.124)	0.046*
model2_B	age_at_iview	228	-0.004	(-0.008, 0)	0.04*
model2_B	control	228	Ref	Ref	Ref
model2_B	is_case_cleanHealthy Case	228	0.09	(0.011, 0.169)	0.025*
model2_B	is_case_cleanUnhealthy Case	228	0.068	(0.007, 0.13)	0.03*

Model	Coefficient	N	Beta	95% CI	P
model3_B	age_at_iview	228	-0.004	(-0.008, -0.001)	0.024*
model3_B	control	228	Ref	Ref	Ref
model3_B	is_case_cleanHealthy Case	228	0.084	(0.002, 0.166)	0.044*
model3_B	is_case_cleanUnhealthy Case	228	0.063	(0, 0.126)	0.049*

3.1.3 Set D

```
ros_merged_D <- ros_merged %>% filter(Set == "D")
```

Model 1

```
model1_D <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_D)
```

Model 2

```
model2_D <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged_D)
```

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3_D <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +  
educ_ord + bmi_bin, ros_merged_D)
```

```
summary_table = NULL

model_list <- c("model1_D", "model2_D", "model3_D")

var <- c("age_at_iview", "control", "is_case_cleanHealthy Case",  
        "is_case_cleanUnhealthy Case")

results <- NULL

for (i in 1:length(model_list)) {

  model <- get(model_list[i])
  nobs <- nobs(model)
  cont <- "Ref"

  for (j in 1:length(var)) {

    if (var[j] == "control") {

      results<- rbind(results,c(model_list[i], var[j], nobs, rep("Ref",3)))

    } else {
      beta    <- round(coef(model)[var[j]], 3)

      ci_ll    <- round(confint(model)[var[j], 1],3)
    }
  }
}
```

```

ci_ul  <- round(confint(model)[var[j], 2],3)
ci     <- paste0("(", ci_ll, ", ", ci_ul, ")")

p      <- round(summary(model)$coefficients[var[j], "Pr(>|t|)"],3)

pval <- p
if(p<=0.05){pval<-paste(p, "*",sep="")}
if(p<=0.01){pval<-paste(p, "**", sep="")}
if(p<=0.001){pval<-paste(p, "***", sep="")}

results <- rbind(results,c(model_list[i], var[j], nobs, beta, ci, pval))
}

}

df <- as.data.frame(results)
colnames(df)<-c("Model", "Coefficient", "N", "Beta", "95% CI", "P")

kable(df)

```

Model	Coefficient	N	Beta	95% CI	P
model1_D	age_at_iview	231	-0.002	(-0.006, 0.002)	0.314
model1_D	control	231	Ref	Ref	Ref
model1_D	is_case_cleanHealthy Case	231	0.081	(0.001, 0.161)	0.046*
model1_D	is_case_cleanUnhealthy Case	231	-0.009	(-0.078, 0.059)	0.785
model2_D	age_at_iview	231	-0.002	(-0.006, 0.002)	0.327
model2_D	control	231	Ref	Ref	Ref
model2_D	is_case_cleanHealthy Case	231	0.076	(-0.004, 0.156)	0.063
model2_D	is_case_cleanUnhealthy Case	231	-0.003	(-0.072, 0.065)	0.922
model3_D	age_at_iview	231	-0.002	(-0.006, 0.002)	0.296
model3_D	control	231	Ref	Ref	Ref
model3_D	is_case_cleanHealthy Case	231	0.072	(-0.011, 0.156)	0.088
model3_D	is_case_cleanUnhealthy Case	231	-0.002	(-0.072, 0.068)	0.951

3.1.4 All sets

Model 1

```
model1_all <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged)
```

Model 2

```
model2_all <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex, ros_merged)
```

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3_all <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
educ_ord + bmi_bin, ros_merged)
```

```

model4_all <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
  alcohol_3cat + tobacco_3cat +
  cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
  cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
  cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
  cidi_q17, ros_merged)

model4_all_truescd <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
  alcohol_3cat + tobacco_3cat +
  cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
  cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
  cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
  cidi_q17, ros_merged)

```

3.2 Models (all sets) with K10 + antipsychotics

Basic Model

```

model_antipsych <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  antipsych_meds_count + miniK_count
  , ros_merged)
summary(model_antipsych)

##
## Call:
## lm(formula = ltl_adj2 ~ age_at_iview + is_case_clean + msex +
##   antipsych_meds_count + miniK_count, data = ros_merged)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63582 -0.14858 -0.01388  0.13408  0.74805
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.0004808  0.0333827  -0.014  0.98851
## age_at_iview   -0.0021460  0.0008926  -2.404  0.01640 *
## is_case_cleanHealthy Case    0.1504015  0.0459627   3.272  0.00111 **
## is_case_cleanUnhealthy Case  0.1105551  0.0474937   2.328  0.02014 *
## msexFemale     0.0383157  0.0147129   2.604  0.00935 **
## antipsych_meds_count   -0.0793820  0.0282941  -2.806  0.00513 **
## miniK_count      0.0004104  0.0051817   0.079  0.93689
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2118 on 932 degrees of freedom
## Multiple R-squared:  0.03444,    Adjusted R-squared:  0.02822
## F-statistic:  5.54 on 6 and 932 DF,  p-value: 1.182e-05

```



```
model_antipsych <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
                      antipsych_meds_cat + miniK_count
                      , ros_merged)
summary(model_antipsych)
```

```
##
## Call:
## lm(formula = ltl_adj2 ~ age_at_iview + is_case_clean + msex +
##     antipsych_meds_cat + miniK_count, data = ros_merged)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.63614 -0.14980 -0.01349  0.13392  0.74773
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.0010456   0.0334243    0.031  0.97505
## age_at_iview   -0.0021904   0.0008938   -2.451  0.01445 *
## is_case_cleanHealthy Case    0.0703035   0.0988737    0.711  0.47724
## is_case_cleanUnhealthy Case  0.0296005   0.1001874    0.295  0.76771
## msexFemale      0.0384839   0.0147147    2.615  0.00906 **
## antipsych_meds_cat1    0.0052871   0.0957190    0.055  0.95596
## antipsych_meds_cat2+  -0.0853483   0.0995062   -0.858  0.39127
## miniK_count      0.0001013   0.0051850    0.020  0.98441
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2119 on 931 degrees of freedom
## Multiple R-squared:  0.03525,    Adjusted R-squared:  0.028
## F-statistic:  4.86 on 7 and 931 DF,  p-value: 2.135e-05
```

Full Model

```
model_antipsych <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
                      educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
                      alcohol_3cat + tobacco_3cat + antipsych_meds_count + miniK_count +
                      cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
                      cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
                      cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
                      cidi_q17, ros_merged)
summary(model_antipsych)
```

```
##
## Call:
## lm(formula = ltl_adj2 ~ age_at_iview + is_case_clean + msex +
##     educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
##     alcohol_3cat + tobacco_3cat + antipsych_meds_count + miniK_count +
##     cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
##     cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 + cidi_q12 +
##     cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 + cidi_q17, data = ros_merged)
##
```

```
## Residuals:
##      Min        1Q      Median        3Q        Max
## -0.60672 -0.14594 -0.01246  0.13551  0.69664
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.0196315  0.0394976   0.497  0.61929
## age_at_iview     -0.0021428  0.0009391  -2.282  0.02274 *
## is_case_cleanHealthy Case    0.1412111  0.0481491   2.933  0.00344 **
## is_case_cleanUnhealthy Case  0.0959210  0.0549846   1.745  0.08141 .
## msexFemale       0.0382477  0.0172397   2.219  0.02676 *
## educ_ordSecondary -0.0262874  0.0175167  -1.501  0.13378
## educ_ordCollege  -0.0492405  0.0181715  -2.710  0.00686 **
## bmi_binUnderweight (<= 18.5) 0.0312587  0.0299610   1.043  0.29708
## bmi_binOverweight (25.0 - 29.9) -0.0047764  0.0176068  -0.271  0.78624
## bmi_binObese (>= 30)      0.0194911  0.0270399   0.721  0.47120
## khat_3catIrregular Users    -0.0081794  0.0213251  -0.384  0.70140
## khat_3catRegular Users     -0.0048374  0.0274999  -0.176  0.86041
## alcohol_3catIrregular Users  0.0053228  0.0160371   0.332  0.74004
## alcohol_3catRegular Users    0.0301085  0.0225362   1.336  0.18189
## cannabis_3catIrregular Users -0.0246772  0.0396152  -0.623  0.53349
## cannabis_3catRegular Users   -0.0423302  0.1112466  -0.381  0.70366
## tobacco_3catIrregular Users  0.0159612  0.0383646   0.416  0.67748
## tobacco_3catRegular Users    0.0100476  0.0436934   0.230  0.81818
## antipsych_meds_count      -0.0698612  0.0296480  -2.356  0.01867 *
## miniK_count             0.0005858  0.0053519   0.109  0.91286
## cidi_q11              0.0046743  0.0423915   0.110  0.91222
## cidi_q21             -0.0230531  0.0352711  -0.654  0.51354
## cidi_q31              0.0211409  0.0317239   0.666  0.50533
## cidi_q41              0.1707147  0.0642711   2.656  0.00804 **
## cidi_q51             -0.0477606  0.0492151  -0.970  0.33209
## cidi_q61             -0.1917295  0.2335456  -0.821  0.41189
## cidi_q71              0.1490282  0.0764277   1.950  0.05150 .
## cidi_q81             -0.0095406  0.0945240  -0.101  0.91963
## cidi_q91             -0.0079007  0.0544651  -0.145  0.88470
## cidi_q101            -0.2292768  0.0867307  -2.644  0.00835 **
## cidi_q111            0.0491146  0.0473009   1.038  0.29939
## cidi_q121            0.1272377  0.0914622   1.391  0.16452
## cidi_q131            -0.0301131  0.0773933  -0.389  0.69730
## cidi_q141            -0.0451688  0.0422323  -1.070  0.28512
## cidi_q151            -0.1688360  0.0773784  -2.182  0.02937 *
## cidi_q15777          -0.0210039  0.0413653  -0.508  0.61174
## cidi_q161            -0.1780817  0.1538133  -1.158  0.24726
## cidi_q17777          -0.1321911  0.0814781  -1.622  0.10506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2103 on 901 degrees of freedom
## Multiple R-squared:  0.08024,    Adjusted R-squared:  0.04247
## F-statistic: 2.124 on 37 and 901 DF,  p-value: 0.0001324
```

```
model_antipsych <- lm(ltl_adj2 ~ age_at_iview + is_case_clean + msex +
  educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
  alcohol_3cat + tobacco_3cat + antipsych_meds_cat + miniK_count +
```

```

cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
cidi_q17, ros_merged)

```

```
summary(model_antipsych)
```

```

##
## Call:
## lm(formula = ltl_adj2 ~ age_at_iview + is_case_clean + msex +
##     educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
##     alcohol_3cat + tobacco_3cat + antipsych_meds_cat + miniK_count +
##     cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
##     cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 + cidi_q12 +
##     cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 + cidi_q17, data = ros_merged)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60974 -0.14739 -0.01233  0.13435  0.69659
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.199e-02  3.954e-02   0.556  0.57836
## age_at_iview     -2.231e-03  9.417e-04  -2.369  0.01805 *
## is_case_cleanHealthy Case    3.491e-02  1.024e-01   0.341  0.73325
## is_case_cleanUnhealthy Case -1.343e-02  1.081e-01  -0.124  0.90114
## msexFemale       3.865e-02  1.724e-02   2.242  0.02518 *
## educ_ordSecondary -2.583e-02  1.751e-02  -1.475  0.14053
## educ_ordCollege  -4.953e-02  1.817e-02  -2.725  0.00655 **
## bmi_binUnderweight (<= 18.5)  3.128e-02  2.996e-02   1.044  0.29660
## bmi_binOverweight (25.0 - 29.9) -3.168e-03  1.769e-02  -0.179  0.85795
## bmi_binObese (>= 30)    1.935e-02  2.704e-02   0.716  0.47429
## khat_3catIrregular Users  -7.994e-03  2.132e-02  -0.375  0.70783
## khat_3catRegular Users   -4.859e-03  2.749e-02  -0.177  0.85977
## alcohol_3catIrregular Users  5.556e-03  1.604e-02   0.346  0.72906
## alcohol_3catRegular Users   3.102e-02  2.254e-02   1.376  0.16901
## cannabis_3catIrregular Users -2.630e-02  3.959e-02  -0.664  0.50666
## cannabis_3catRegular Users  -3.768e-02  1.114e-01  -0.338  0.73522
## tobacco_3catIrregular Users  1.718e-02  3.837e-02   0.448  0.65432
## tobacco_3catRegular Users   1.431e-02  4.385e-02   0.326  0.74423
## antipsych_meds_cat1        4.308e-02  9.978e-02   0.432  0.66607
## antipsych_meds_cat2+     -4.095e-02  1.036e-01  -0.395  0.69280
## miniK_count          8.562e-05  5.362e-03   0.016  0.98726
## cidi_q11            2.910e-03  4.242e-02   0.069  0.94533
## cidi_q21           -2.410e-02  3.529e-02  -0.683  0.49490
## cidi_q31            2.188e-02  3.175e-02   0.689  0.49087
## cidi_q41            1.820e-01  6.495e-02   2.802  0.00518 **
## cidi_q51           -4.704e-02  4.921e-02  -0.956  0.33941
## cidi_q61           -2.039e-01  2.338e-01  -0.872  0.38331
## cidi_q71            1.564e-01  7.662e-02   2.041  0.04158 *
## cidi_q81           -1.296e-02  9.457e-02  -0.137  0.89100
## cidi_q91           -8.414e-04  5.480e-02  -0.015  0.98775
## cidi_q101          -2.132e-01  8.777e-02  -2.429  0.01532 *

```

```
## cidi_q111          4.977e-02  4.730e-02   1.052  0.29291
## cidi_q121          1.267e-01  9.147e-02   1.385  0.16647
## cidi_q131         -3.197e-02  7.740e-02  -0.413  0.67966
## cidi_q141         -4.922e-02  4.237e-02  -1.162  0.24562
## cidi_q151         -1.723e-01  7.741e-02  -2.226  0.02625 *
## cidi_q15777        -2.474e-02  4.134e-02  -0.598  0.54976
## cidi_q161         -1.769e-01  1.538e-01  -1.151  0.25024
## cidi_q17777        -1.286e-01  8.154e-02  -1.577  0.11525
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2102 on 900 degrees of freedom
## Multiple R-squared:  0.08157,    Adjusted R-squared:  0.0428
## F-statistic: 2.104 on 38 and 900 DF,  p-value: 0.0001371
```

```
summary_table = NULL

model_list <- c("model1_all", "model2_all", "model3_all", "model4_all")

results <- NULL

for (i in 1:length(model_list)) {

  model <- get(model_list[i])
  nobs <- nobs(model)
  cont <- "Ref"

  var <- names(model$coefficients)[-1]

  for (j in 1:length(var)) {

    beta    <- round(coef(model)[var[j]], 4)

    ci_ll    <- round(confint(model)[var[j], 1],4)
    ci_ul    <- round(confint(model)[var[j], 2],4)
    ci       <- paste0("(", ci_ll, ", ", ci_ul, ")")

    p        <- round(summary(model)$coefficients[var[j], "Pr(>|t|)"],3)

    pval <- p
    if(p<= 0.001){beta<-paste(beta, "***", sep="")}
    else if (p<= 0.01){beta<-paste(beta, "**", sep="")}
    else if(p<= 0.05){beta<-paste(beta, "*", sep="")}

    results <- rbind(results,c(model_list[i], var[j], nobs, beta, ci, p))
  }
}

df <- as.data.frame(results)
colnames(df)<-c("Model", "Coefficient", "N", "Beta", "95% CI", "P")
```

`kable(df)`

Model	Coefficient	N	Beta	95% CI	P
model1_all	age_at_iview	939	-0.002*	(-0.0037, -2e-04)	0.028
model1_all	is_case_cleanHealthy Case	939	0.0682***	(0.0317, 0.1048)	0
model1_all	is_case_cleanUnhealthy Case	939	0.0222	(-0.0096, 0.054)	0.171
model2_all	age_at_iview	939	-0.0021*	(-0.0039, -4e-04)	0.017
model2_all	is_case_cleanHealthy Case	939	0.0648***	(0.0283, 0.1013)	0.001
model2_all	is_case_cleanUnhealthy Case	939	0.0252	(-0.0065, 0.0569)	0.12
model2_all	msexFemale	939	0.0417**	(0.0129, 0.0706)	0.005
model3_all	age_at_iview	939	-0.0025**	(-0.0043, -7e-04)	0.006
model3_all	is_case_cleanHealthy Case	939	0.0578**	(0.021, 0.0946)	0.002
model3_all	is_case_cleanUnhealthy Case	939	0.0176	(-0.0146, 0.0498)	0.283
model3_all	msexFemale	939	0.0335*	(0.0032, 0.0639)	0.03
model3_all	educ_ordSecondary	939	-0.025	(-0.0587, 0.0087)	0.146
model3_all	educ_ordCollege	939	-0.0557**	(-0.0906, -0.0208)	0.002
model3_all	bmi_binUnderweight (<= 18.5)	939	0.0382	(-0.0204, 0.0968)	0.201
model3_all	bmi_binOverweight (25.0 - 29.9)	939	-9e-04	(-0.0352, 0.0335)	0.959
model3_all	bmi_binObese (>= 30)	939	0.021	(-0.0315, 0.0734)	0.433
model4_all	age_at_iview	939	-0.0022*	(-0.004, -3e-04)	0.022
model4_all	is_case_cleanHealthy Case	939	0.0682***	(0.0298, 0.1066)	0.001
model4_all	is_case_cleanUnhealthy Case	939	0.0278	(-0.0403, 0.0959)	0.424
model4_all	msexFemale	939	0.0398*	(0.0059, 0.0737)	0.021
model4_all	educ_ordSecondary	939	-0.0255	(-0.0599, 0.0089)	0.146
model4_all	educ_ordCollege	939	-0.0487**	(-0.0844, -0.0129)	0.008
model4_all	bmi_binUnderweight (<= 18.5)	939	0.0321	(-0.0267, 0.091)	0.284
model4_all	bmi_binOverweight (25.0 - 29.9)	939	-0.0037	(-0.0383, 0.0309)	0.834
model4_all	bmi_binObese (>= 30)	939	0.0194	(-0.0338, 0.0725)	0.475
model4_all	khat_3catIrregular Users	939	-0.0093	(-0.0513, 0.0326)	0.662
model4_all	khat_3catRegular Users	939	-0.0065	(-0.0604, 0.0475)	0.814
model4_all	alcohol_3catIrregular Users	939	0.005	(-0.0265, 0.0365)	0.755
model4_all	alcohol_3catRegular Users	939	0.0316	(-0.0127, 0.0759)	0.161
model4_all	cannabis_3catIrregular Users	939	-0.0281	(-0.1058, 0.0495)	0.477
model4_all	cannabis_3catRegular Users	939	-0.0803	(-0.2967, 0.1362)	0.467
model4_all	tobacco_3catIrregular Users	939	0.0156	(-0.0598, 0.091)	0.685
model4_all	tobacco_3catRegular Users	939	-0.0016	(-0.0865, 0.0833)	0.97
model4_all	cidi_q11	939	0.0066	(-0.0767, 0.09)	0.876
model4_all	cidi_q21	939	-0.0239	(-0.0929, 0.0451)	0.496
model4_all	cidi_q31	939	0.0196	(-0.0427, 0.082)	0.537
model4_all	cidi_q41	939	0.1784**	(0.0522, 0.3046)	0.006
model4_all	cidi_q51	939	-0.0513	(-0.148, 0.0454)	0.298
model4_all	cidi_q61	939	-0.1734	(-0.6318, 0.2851)	0.458
model4_all	cidi_q71	939	0.1672*	(0.018, 0.3164)	0.028
model4_all	cidi_q81	939	-0.0322	(-0.217, 0.1527)	0.733
model4_all	cidi_q91	939	-0.012	(-0.1188, 0.0948)	0.826
model4_all	cidi_q101	939	-0.222*	(-0.3923, -0.0517)	0.011
model4_all	cidi_q111	939	0.048	(-0.045, 0.1409)	0.312
model4_all	cidi_q121	939	0.1127	(-0.0667, 0.292)	0.218
model4_all	cidi_q131	939	-0.0256	(-0.1778, 0.1265)	0.741
model4_all	cidi_q141	939	-0.0426	(-0.1255, 0.0403)	0.314
model4_all	cidi_q151	939	-0.1775*	(-0.3294, -0.0256)	0.022
model4_all	cidi_q15777	939	-0.025	(-0.1062, 0.0563)	0.547

Model	Coefficient	N	Beta	95% CI	P
model4_all	cidi_q161	939	-0.1765	(-0.4783, 0.1254)	0.252
model4_all	cidi_q17777	939	-0.1545	(-0.3136, 0.0046)	0.057

```
summary_table = NULL

model_list <- c("model4_all")

var <- c("age_at_iview","control", "is_case_cleanHealthy Case",
        "is_case_cleanUnhealthy Case", "msexFemale", "educ_ordSecondary",
        "educ_ordCollege", "cidi_q41", "cidi_q71", "cidi_q101", "cidi_q151")

results <- NULL

for (i in 1:length(model_list)) {

  model <- get(model_list[i])
  nobs <- nobs(model)
  cont <- "Ref"

  for (j in 1:length(var)) {

    if (var[j] == "control") {

      results<- rbind(results,c(model_list[i], var[j], nobs, rep("Ref",3)))

    } else {
      beta <- round(coef(model)[var[j]], 3)

      ci_ll <- round(confint(model)[var[j], 1],6)
      ci_ul <- round(confint(model)[var[j], 2],6)
      ci <- paste0("(", ci_ll, ", ", ci_ul, ")")

      p <- round(summary(model)$coefficients[var[j], "Pr(>|t|)"],3)

      pval <- p
      if(p<=0.05){pval<-paste(p, "*",sep="")}
      if(p<=0.01){pval<-paste(p, "**", sep="")}
      if(p<=0.001){pval<-paste(p, "***", sep="")}

      results <- rbind(results,c(model_list[i], var[j], nobs, beta, ci, pval))
    }

  }

}

df <- as.data.frame(results)
colnames(df)<-c("Model", "Coefficient", "N", "Beta", "95% CI", "P")

kable(df)
```

Model	Coefficient	N	Beta	95% CI	P
model4_all	age_at_iview	939	-0.002	(-0.004005, -0.000316)	0.022*
model4_all	control	939	Ref	Ref	Ref
model4_all	is_case_cleanHealthy Case	939	0.068	(0.029754, 0.106582)	0.001***
model4_all	is_case_cleanUnhealthy Case	939	0.028	(-0.040325, 0.095892)	0.424
model4_all	msexFemale	939	0.04	(0.005933, 0.073668)	0.021*
model4_all	educ_ordSecondary	939	-0.026	(-0.059935, 0.008915)	0.146
model4_all	educ_ordCollege	939	-0.049	(-0.084387, -0.012931)	0.008**
model4_all	cidi_q41	939	0.178	(0.052159, 0.304606)	0.006**
model4_all	cidi_q71	939	0.167	(0.018007, 0.316418)	0.028*
model4_all	cidi_q101	939	-0.222	(-0.392338, -0.051662)	0.011*
model4_all	cidi_q151	939	-0.178	(-0.329448, -0.025624)	0.022*

3.3 Stratified by sex

```
ros_merged_male <- ros_merged %>% filter(msex=="Male")
ros_merged_female <- ros_merged %>% filter(msex=="Female")
```

3.3.1 Male

Model 1

```
model1_male <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_male)
```

Model 2 Not needed as no sex

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3_male <- lm(ltl_adj2 ~ age_at_iview + is_case_clean +
  educ_ord + bmi_bin, ros_merged_male)
```

```
model4_male <- lm(ltl_adj2 ~ age_at_iview + is_case_clean +
  educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +
  alcohol_3cat + tobacco_3cat +
  cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
  cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +
  cidi_q12 + cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +
  cidi_q17, ros_merged_male)
```

```
summary_table = NULL
```

```
model_list <- c("model1_male", "model3_male", "model4_male")
```

```
results <- NULL
```

```
for (i in 1:length(model_list)) {
  model <- get(model_list[i])
  nob <- nob(model)
```

```

cont <- "Ref"

var <- names(model$coefficients)[-1]

for (j in 1:length(var)) {

  beta    <- round(coef(model)[var[j]], 3)

  ci_ll    <- round(confint(model)[var[j], 1],3)
  ci_ul    <- round(confint(model)[var[j], 2],3)
  ci       <- paste0("(", ci_ll, ", ", ci_ul, ")")

  p        <- round(summary(model)$coefficients[var[j], "Pr(>|t|)"],3)

  pval <- p
  if(p<= 0.001){beta<-paste(beta, "***", sep="")}
  else if (p<= 0.01){beta<-paste(beta, "**", sep="")}
  else if(p<= 0.05){beta<-paste(beta, "*", sep="")}

  results <- rbind(results,c(model_list[i], var[j], nobs, beta, ci, p))
}

}

df <- as.data.frame(results)
colnames(df)<-c("Model", "Coefficient", "N", "Beta", "95% CI", "P")

kable(df)

```

Model	Coefficient	N	Beta	95% CI	P
model1_male	age_at_iview	614	-0.003**	(-0.005, -0.001)	0.009
model1_male	is_case_cleanHealthy Case	614	0.074***	(0.029, 0.119)	0.001
model1_male	is_case_cleanUnhealthy Case	614	0.024	(-0.012, 0.06)	0.195
model3_male	age_at_iview	614	-0.003**	(-0.005, -0.001)	0.01
model3_male	is_case_cleanHealthy Case	614	0.067**	(0.021, 0.112)	0.004
model3_male	is_case_cleanUnhealthy Case	614	0.018	(-0.019, 0.054)	0.35
model3_male	educ_ordSecondary	614	-0.025	(-0.066, 0.017)	0.24
model3_male	educ_ordCollege	614	-0.049*	(-0.089, -0.009)	0.017
model3_male	bmi_binUnderweight (<= 18.5)	614	0.012	(-0.053, 0.078)	0.712
model3_male	bmi_binOverweight (25.0 - 29.9)	614	-0.034	(-0.076, 0.007)	0.106
model3_male	bmi_binObese (>= 30)	614	-0.001	(-0.092, 0.09)	0.984
model4_male	age_at_iview	614	-0.002*	(-0.004, 0)	0.048
model4_male	is_case_cleanHealthy Case	614	0.069**	(0.021, 0.117)	0.005
model4_male	is_case_cleanUnhealthy Case	614	0.038	(-0.06, 0.136)	0.445
model4_male	educ_ordSecondary	614	-0.018	(-0.061, 0.024)	0.398
model4_male	educ_ordCollege	614	-0.043*	(-0.084, -0.002)	0.042
model4_male	bmi_binUnderweight (<= 18.5)	614	0.017	(-0.05, 0.083)	0.625
model4_male	bmi_binOverweight (25.0 - 29.9)	614	-0.031	(-0.074, 0.011)	0.148
model4_male	bmi_binObese (>= 30)	614	0.023	(-0.073, 0.119)	0.639

Model	Coefficient	N	Beta	95% CI	P
model4_male	khat_3catIrregular Users	614	-0.021	(-0.065, 0.023)	0.357
model4_male	khat_3catRegular Users	614	-0.01	(-0.063, 0.044)	0.718
model4_male	alcohol_3catIrregular Users	614	-0.023	(-0.062, 0.016)	0.243
model4_male	alcohol_3catRegular Users	614	0.009	(-0.038, 0.057)	0.704
model4_male	cannabis_3catIrregular Users	614	-0.011	(-0.091, 0.068)	0.78
model4_male	cannabis_3catRegular Users	614	-0.085	(-0.292, 0.122)	0.419
model4_male	tobacco_3catIrregular Users	614	0.011	(-0.088, 0.11)	0.823
model4_male	tobacco_3catRegular Users	614	-0.003	(-0.107, 0.101)	0.953
model4_male	cidi_q11	614	-0.009	(-0.125, 0.108)	0.884
model4_male	cidi_q21	614	0.005	(-0.08, 0.09)	0.911
model4_male	cidi_q31	614	0.011	(-0.067, 0.089)	0.781
model4_male	cidi_q41	614	0.065	(-0.101, 0.231)	0.439
model4_male	cidi_q51	614	-0.069	(-0.199, 0.061)	0.3
model4_male	cidi_q61	614	-0.088	(-0.55, 0.374)	0.709
model4_male	cidi_q71	614	0.269**	(0.089, 0.448)	0.003
model4_male	cidi_q81	614	-0.077	(-0.367, 0.213)	0.601
model4_male	cidi_q91	614	0.005	(-0.156, 0.166)	0.951
model4_male	cidi_q101	614	-0.219	(-0.446, 0.008)	0.058
model4_male	cidi_q111	614	0.009	(-0.123, 0.142)	0.889
model4_male	cidi_q121	614	0.122	(-0.058, 0.301)	0.183
model4_male	cidi_q131	614	-0.086	(-0.309, 0.137)	0.45
model4_male	cidi_q141	614	-0.007	(-0.114, 0.1)	0.898
model4_male	cidi_q151	614	-0.321*	(-0.606, -0.037)	0.027
model4_male	cidi_q15777	614	-0.061	(-0.153, 0.031)	0.191
model4_male	cidi_q161	614	-0.147	(-0.566, 0.272)	0.492
model4_male	cidi_q17777	614	-0.185	(-0.386, 0.016)	0.071

3.3.2 Female

Model 1

```
model1_female <- lm(ltl_adj2 ~ age_at_iview + is_case_clean, ros_merged_female)
```

Model 2 Not needed as no sex

Model 3 3. lm(Rosner LTL ~ Age + case + sex + education + BMI

```
model3_female <- lm(ltl_adj2 ~ age_at_iview + is_case_clean +  
educ_ord + bmi_bin, ros_merged_female)
```

Model 4

```
model4_female <- lm(ltl_adj2 ~ age_at_iview + is_case_clean +  
educ_ord + bmi_bin + khat_3cat + alcohol_3cat + cannabis_3cat +  
alcohol_3cat + tobacco_3cat +  
cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 +  
#cidi_q6 +  
cidi_q7 + cidi_q8 + cidi_q9 + cidi_q10 + cidi_q11 +  
#cidi_q12  
cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 +  
cidi_q17, ros_merged_female)
```

```

summary_table = NULL

model_list <- c("model1_female", "model3_female", "model4_female")

results <- NULL

for (i in 1:length(model_list)) {

  model <- get(model_list[i])
  nobs <- nobs(model)
  cont <- "Ref"

  var <- names(model$coefficients)[-1]

  for (j in 1:length(var)) {

    beta <- round(coef(model)[var[j]], 3)

    ci_ll <- round(confint(model)[var[j], 1], 3)
    ci_ul <- round(confint(model)[var[j], 2], 3)
    ci <- paste0("(", ci_ll, ", ", ci_ul, ")")

    p <- round(summary(model)$coefficients[var[j], "Pr(>|t|)"], 3)

    pval <- p
    if(p<= 0.001){beta<-paste(beta, "***", sep="")}
    else if (p<= 0.01){beta<-paste(beta, "**", sep="")}
    else if(p<= 0.05){beta<-paste(beta, "*", sep="")}

    results <- rbind(results,c(model_list[i], var[j], nobs, beta, ci, p))
  }

}

df <- as.data.frame(results)
colnames(df)<-c("Model", "Coefficient", "N", "Beta", "95% CI", "P")

kable(df)

```

Model	Coefficient	N	Beta	95% CI	P
model1_female	age_at_iview	325	-0.001	(-0.004, 0.002)	0.489
model1_female	is_case_cleanHealthy Case	325	0.054	(-0.009, 0.116)	0.091
model1_female	is_case_cleanUnhealthy Case	325	0.028	(-0.035, 0.091)	0.384
model3_female	age_at_iview	325	-0.002	(-0.005, 0.002)	0.305
model3_female	is_case_cleanHealthy Case	325	0.051	(-0.012, 0.114)	0.109
model3_female	is_case_cleanUnhealthy Case	325	0.029	(-0.036, 0.093)	0.381
model3_female	educ_ordSecondary	325	-0.024	(-0.083, 0.034)	0.415
model3_female	educ_ordCollege	325	-0.062	(-0.131, 0.006)	0.074

Model	Coefficient	N	Beta	95% CI	P
model3_female	bmi_binUnderweight (≤ 18.5)	325	0.111	(-0.013, 0.234)	0.079
model3_female	bmi_binOverweight (25.0 - 29.9)	325	0.057	(-0.004, 0.117)	0.069
model3_female	bmi_binObese (≥ 30)	325	0.046	(-0.025, 0.116)	0.203
model4_female	age_at_iview	325	-0.001	(-0.004, 0.003)	0.658
model4_female	is_case_cleanHealthy Case	325	0.068*	(0.002, 0.134)	0.043
model4_female	is_case_cleanUnhealthy Case	325	0.077	(-0.04, 0.194)	0.194
model4_female	educ_ordSecondary	325	-0.03	(-0.091, 0.031)	0.333
model4_female	educ_ordCollege	325	-0.049	(-0.12, 0.023)	0.183
model4_female	bmi_binUnderweight (≤ 18.5)	325	0.114	(-0.011, 0.24)	0.074
model4_female	bmi_binOverweight (25.0 - 29.9)	325	0.049	(-0.014, 0.112)	0.124
model4_female	bmi_binObese (≥ 30)	325	0.035	(-0.037, 0.108)	0.336
model4_female	khat_3catIrregular Users	325	0.007	(-0.11, 0.124)	0.909
model4_female	khat_3catRegular Users	325	0.132	(-0.194, 0.459)	0.426
model4_female	alcohol_3catIrregular Users	325	0.053	(-0.003, 0.109)	0.064
model4_female	alcohol_3catRegular Users	325	0.093	(-0.025, 0.21)	0.123
model4_female	cannabis_3catIrregular Users	325	-0.154	(-0.431, 0.123)	0.276
model4_female	tobacco_3catIrregular Users	325	0.066	(-0.144, 0.277)	0.536
model4_female	tobacco_3catRegular Users	325	-0.416	(-0.986, 0.154)	0.152
model4_female	cidi_q11	325	-0.022	(-0.154, 0.109)	0.738
model4_female	cidi_q21	325	-0.05	(-0.176, 0.077)	0.438
model4_female	cidi_q31	325	0.01	(-0.108, 0.129)	0.862
model4_female	cidi_q41	325	0.25*	(0.023, 0.477)	0.031
model4_female	cidi_q51	325	-0.059	(-0.227, 0.11)	0.494
model4_female	cidi_q71	325	-0.01	(-0.302, 0.282)	0.946
model4_female	cidi_q81	325	0.04	(-0.259, 0.339)	0.793
model4_female	cidi_q91	325	-0.01	(-0.175, 0.155)	0.907
model4_female	cidi_q101	325	-0.311*	(-0.608, -0.014)	0.04
model4_female	cidi_q111	325	0.01	(-0.139, 0.159)	0.897
model4_female	cidi_q131	325	-0.008	(-0.241, 0.224)	0.944
model4_female	cidi_q141	325	-0.125	(-0.278, 0.029)	0.112
model4_female	cidi_q151	325	-0.131	(-0.356, 0.094)	0.254
model4_female	cidi_q15777	325	0.025	(-0.149, 0.199)	0.778
model4_female	cidi_q161	325	-0.327	(-0.799, 0.144)	0.173
model4_female	cidi_q17777	325	-0.073	(-0.341, 0.195)	0.591