

Telomere Exploration

Hayden Mountcastle and Andrew Ratanatharathorn

November 3rd, 2023

Contents

Analyses

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)

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/Analytical Data"
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)

```

Analysis with clean/unclean cases and controls

Distribution of ltl/age with clean cases , unclean cases, controls Update powerpoint with controls as reference group

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

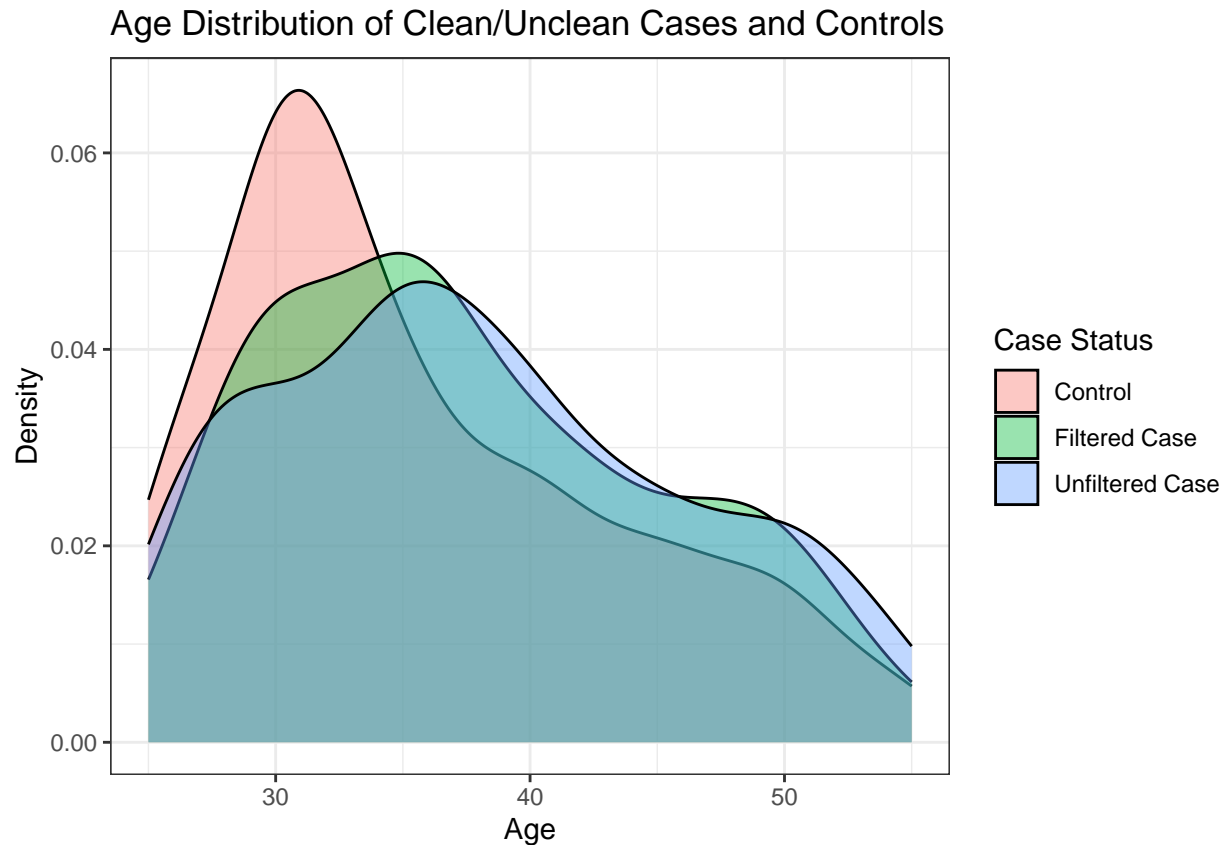
  (tobacco_3cat %in% c("Irregular Users",
                      "Regular Users") |

  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) ~ "Unfiltered Case",
  is_case == 0   ~ "Control",
  TRUE          ~ "Filtered Case"
)) %>% mutate(is_case_clean = factor(is_case_clean,
  levels = c("Control", "Filtered Case",
             "Unfiltered Case")))
```

Age distribution of clean/unclean cases and controls.

The unclean cases among Females screw older than the rest of the population

```
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()
```



New 3-29-23

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

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)
                  ) %>% summarise(`Count with missing` = n())
```

```
## Count with missing
## 1 26
```

2

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
                  ))
```

Make new rosner batch corrected dataframe

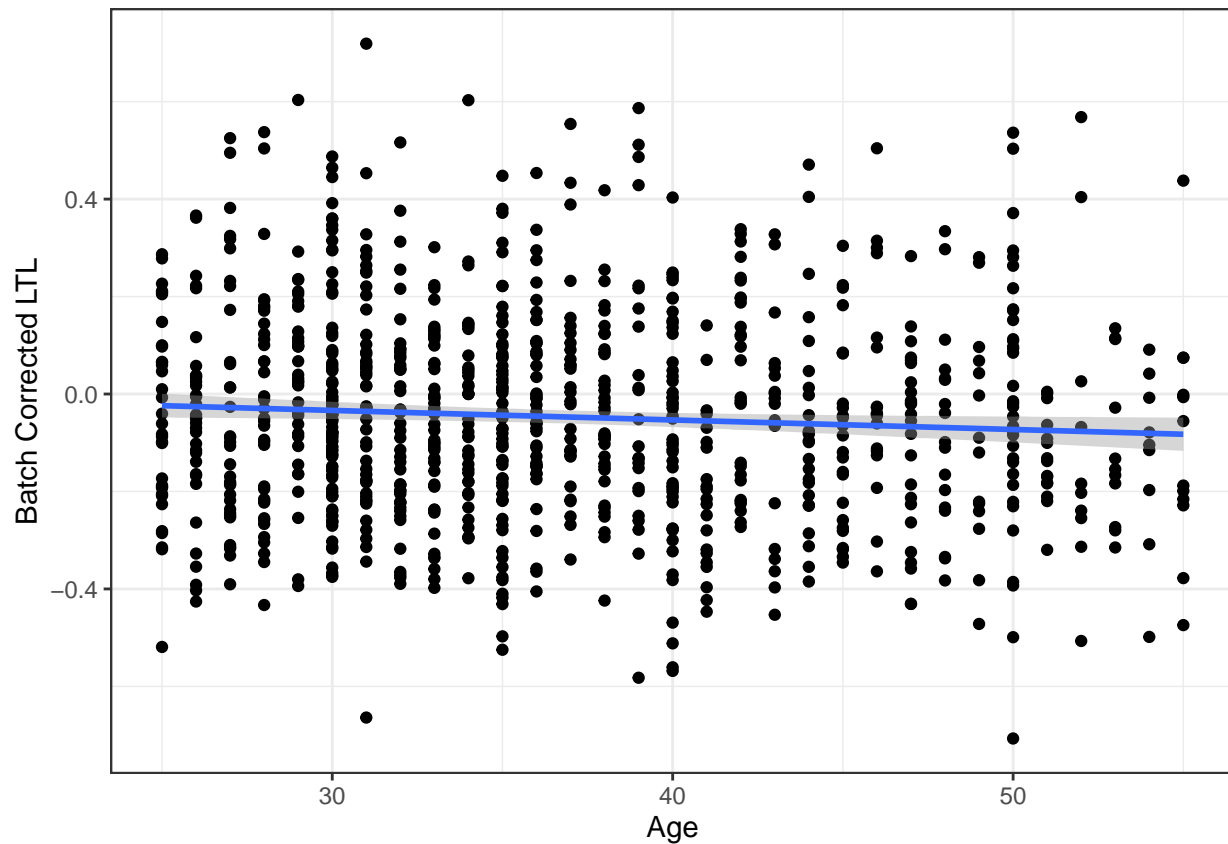
```
# ros <- adjust_batch(
#   data = tel_all_nomissing,
#   markers = ltl,
#   batch = Set,
#   method = simple
# )
```

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")
```

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



Boxplots of LTL and filtered/unfiltered cases

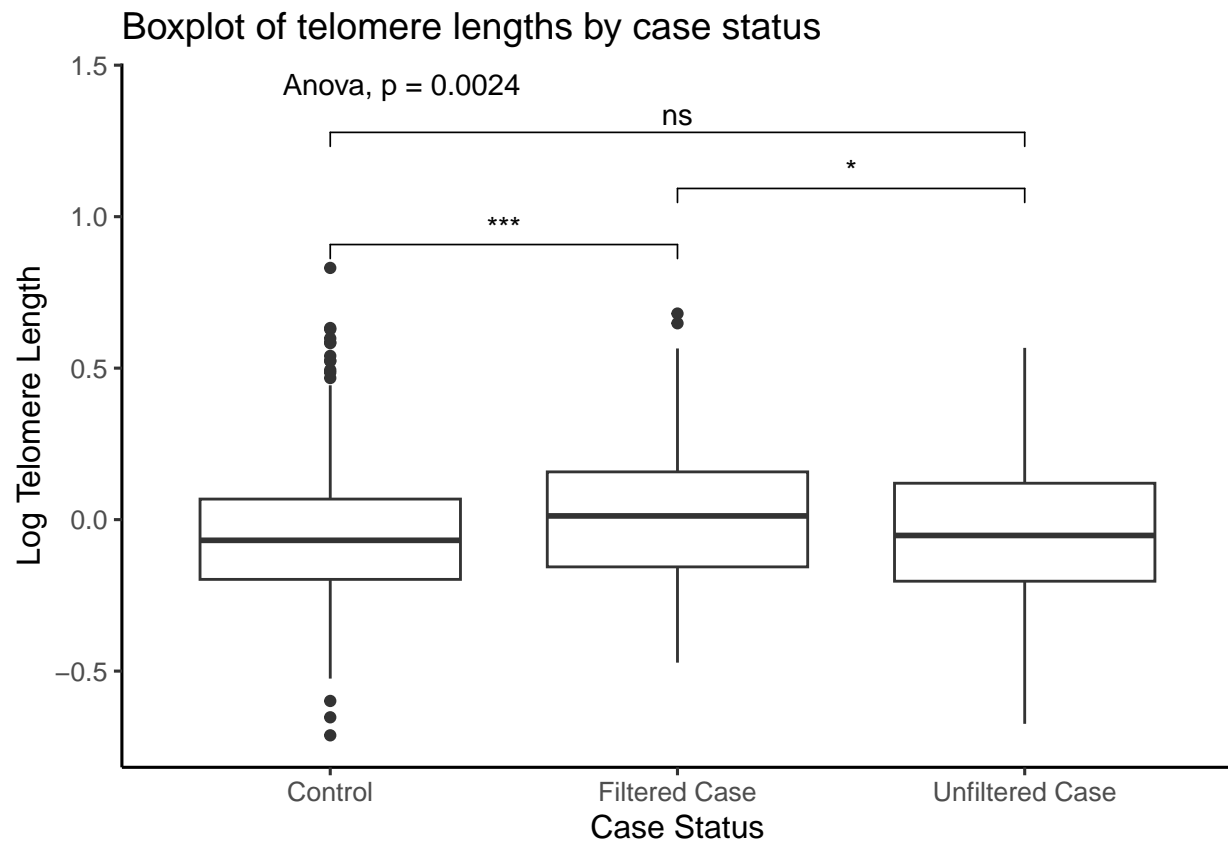
```
compare_means(ltl ~ is_case_clean, data = tel_all_nomissing, method = "t.test")
```

```
## # 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    Filtered Case 0.000578 0.0017 0.00058 ***      T-test  
## 2 ltl   Control    Unfiltered Case 0.411    0.41  0.41080 ns       T-test  
## 3 ltl   Filtered Case Unfiltered Case 0.0119    0.024 0.01194 *       T-test
```



```
my_comparisons <- list( c("Control", "Filtered Case"),
                        c("Filtered Case", "Unfiltered Case"),
                        c("Control", "Unfiltered Case") )

tel_all_nomissing %>%
  ggplot(aes(x=is_case_clean, y=l1)) +
  geom_boxplot() + xlab("Case Status") +
  ylab("Log Telomere Length") + ggtitle("Boxplot of telomere lengths by case status") +
  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()
```



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

```
ros %>% group_by(is_case_clean) %>%
  summarise(R_age_adj.ltl = cor(Age, ltl_adj2))
```

```
## # A tibble: 3 x 2
##   is_case_clean  R_age_adj.ltl
##   <fct>         <dbl>
## 1 Control      -0.0788
## 2 Filtered Case -0.140
## 3 Unfiltered Case -0.0481
```

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)  
}  
  
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
```

```
tbl <- table1(~ 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 1: Demographics by Case Status

	Control (N=484)	Filtered Case (N=183)	Unfiltered Case (N=278)
Log Telomere Length			
Mean (SD)	-0.0516 (0.226)	0.0163 (0.225)	-0.0377 (0.224)
Age			
Mean (SD)	35.7 (7.68)	37.5 (7.52)	37.8 (8.10)
Sex			
Male	315 (65.1%)	103 (56.3%)	199 (71.6%)
Female	169 (34.9%)	80 (43.7%)	79 (28.4%)
Education			
Primary or less	135 (27.9%)	64 (35.0%)	90 (32.4%)
Secondary	138 (28.5%)	78 (42.6%)	120 (43.2%)
College	211 (43.6%)	41 (22.4%)	68 (24.5%)
BMI			
Underweight	28 (5.8%)	8 (4.4%)	19 (6.8%)
Normal Weight	307 (63.4%)	113 (61.7%)	183 (65.8%)
Obese	33 (6.8%)	20 (10.9%)	26 (9.4%)
Overweight	116 (24.0%)	42 (23.0%)	50 (18.0%)
Alcohol Use			
Never Users	201 (41.5%)	136 (74.3%)	89 (32.0%)
Irregular Users	188 (38.8%)	40 (21.9%)	157 (56.5%)
Regular Users	95 (19.6%)	7 (3.8%)	32 (11.5%)
Khat Use			
Never Users	404 (83.5%)	147 (80.3%)	100 (36.0%)
Irregular Users	52 (10.7%)	26 (14.2%)	108 (38.8%)
Regular Users	28 (5.8%)	10 (5.5%)	70 (25.2%)
Cannabis Use			
Never Users	478 (98.8%)	183 (100%)	244 (87.8%)
Irregular Users	6 (1.2%)	0 (0%)	30 (10.8%)
Regular Users	0 (0%)	0 (0%)	4 (1.4%)
Tobacco Use			
Never Users	484 (100%)	183 (100%)	105 (37.8%)
Irregular Users	0 (0%)	0 (0%)	107 (38.5%)
Regular Users	0 (0%)	0 (0%)	66 (23.7%)
Arthritis and Rheumatism			
0	484 (100%)	183 (100%)	241 (86.7%)
1	0 (0%)	0 (0%)	37 (13.3%)
777	0 (0%)	0 (0%)	0 (0%)
Chronic back or neck problems			
0	484 (100%)	183 (100%)	225 (80.9%)
1	0 (0%)	0 (0%)	53 (19.1%)
777	0 (0%)	0 (0%)	0 (0%)
Frequent or severe headaches			
0	484 (100%)	183 (100%)	202 (72.7%)
1	0 (0%)	0 (0%)	76 (27.3%)
777	0 (0%)	0 (0%)	0 (0%)
Any other chronic pain			
0	484 (100%)	183 (100%)	264 (95.0%)
1	0 (0%)	0 (0%)	14 (5.0%)

Table 2: Demographics by Case Status

	Control	Filtered Case	Unfiltered Case
	(N=484)	(N=183)	(N=278)
Seasonal allergies			
0	484 (100%)	183 (100%)	255 (91.7%)
1	0 (0%)	0 (0%)	23 (8.3%)
Stroke			
0	484 (100%)	183 (100%)	277 (99.6%)
1	0 (0%)	0 (0%)	1 (0.4%)
777	0 (0%)	0 (0%)	0 (0%)
Heart attack			
0	484 (100%)	183 (100%)	268 (96.4%)
1	0 (0%)	0 (0%)	10 (3.6%)
777	0 (0%)	0 (0%)	0 (0%)
Heart disease			
0	484 (100%)	183 (100%)	272 (97.8%)
1	0 (0%)	0 (0%)	6 (2.2%)
777	0 (0%)	0 (0%)	0 (0%)
High blood pressure			
0	484 (100%)	183 (100%)	257 (92.4%)
1	0 (0%)	0 (0%)	21 (7.6%)
777	0 (0%)	0 (0%)	0 (0%)
Asthma			
0	484 (100%)	183 (100%)	271 (97.5%)
1	0 (0%)	0 (0%)	7 (2.5%)
777	0 (0%)	0 (0%)	0 (0%)
Tuberculosis			
0	484 (100%)	183 (100%)	253 (91.0%)
1	0 (0%)	0 (0%)	25 (9.0%)
777	0 (0%)	0 (0%)	0 (0%)
Other chronic lung disease			
0	484 (100%)	183 (100%)	272 (97.8%)
1	0 (0%)	0 (0%)	6 (2.2%)
777	0 (0%)	0 (0%)	0 (0%)
Diabetes			
0	484 (100%)	183 (100%)	269 (96.8%)
1	0 (0%)	0 (0%)	9 (3.2%)
777	0 (0%)	0 (0%)	0 (0%)
Stomach or intestine			
0	484 (100%)	183 (100%)	245 (88.1%)
1	0 (0%)	0 (0%)	33 (11.9%)
777	0 (0%)	0 (0%)	0 (0%)
HIV/AIDS			
0	484 (100%)	177 (96.7%)	246 (88.5%)
1	0 (0%)	0 (0%)	9 (3.2%)
777	0 (0%)	6 (3.3%)	23 (8.3%)
Epilepsy/Seizure			
0	484 (100%)	183 (100%)	276 (99.3%)
1	0 (0%)	0 (0%)	2 (0.7%)
777	0 (0%)	0 (0%)	0 (0%)
Cancer			
0	484 (100%)	179 ¹² (97.8%)	275 (98.9%)
777	0 (0%)	4 (2.2%)	3 (1.1%)

6 models with new data

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

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	95% CI	p-value
age_at_iview	0.00	0.001	0.00, 0.00	0.021*
is_case_clean				
Control	—	—	—	
Filtered Case	0.07	0.019	0.03, 0.10	<0.001***
Unfiltered Case	0.02	0.016	-0.01, 0.05	0.2

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	95% CI	p-value
age_at_iview	0.00	0.001	0.00, 0.00	0.012*
is_case_clean				
Control	—	—	—	
Filtered Case	0.06	0.019	0.03, 0.10	<0.001***
Unfiltered Case	0.02	0.016	-0.01, 0.06	0.14
msex				
Male	—	—	—	
Female	0.04	0.015	0.01, 0.07	0.004**

Model 3 3. $\text{lm}(\text{Rosner LTL} \sim \text{Age} + \text{case} + \text{sex} + \text{education} + \text{BMI})$

```
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}"
)
```

Table printed with `knitr::kable()`, not {gt}. Learn why at
 ## <https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html>
 ## To suppress this message, include `message = FALSE` in code chunk header.

Characteristic	Beta	SE	95% CI	p-value
age_at_iview	0.00	0.001	0.00, 0.00	0.005**
is_case_clean				
Control	—	—	—	
Filtered Case	0.06	0.019	0.02, 0.09	0.003**
Unfiltered Case	0.02	0.016	-0.02, 0.05	0.3
msex				
Male	—	—	—	
Female	0.03	0.015	0.00, 0.06	0.025*
educ_ord				
Primary or less	—	—	—	
Secondary	-0.02	0.017	-0.06, 0.01	0.2
College	-0.05	0.018	-0.09, -0.02	0.002**
bmi_bin				
Underweight	—	—	—	
Normal Weight	-0.04	0.030	-0.10, 0.02	0.2
Obese	-0.02	0.038	-0.09, 0.06	0.6
Overweight	-0.04	0.032	-0.10, 0.02	0.2

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	95% CI	p-value
age_at_iview	0.00	0.001	0.00, 0.00	0.017*
is_case_clean				

Characteristic	Beta	SE	95% CI	p-value
Control	—	—	—	
Filtered Case	0.07	0.020	0.03, 0.11	<0.001***
Unfiltered Case	0.03	0.035	-0.04, 0.09	0.5
msex				
Male	—	—	—	
Female	0.04	0.017	0.01, 0.07	0.018*
educ_ord				
Primary or less	—	—	—	
Secondary	-0.02	0.018	-0.06, 0.01	0.2
College	-0.05	0.018	-0.08, -0.01	0.010**
bmi_bin				
Underweight	—	—	—	
Normal Weight	-0.03	0.030	-0.09, 0.03	0.3
Obese	-0.01	0.038	-0.09, 0.06	0.7
Overweight	-0.04	0.033	-0.10, 0.03	0.3
khat_3cat				
Never Users	—	—	—	
Irregular Users	-0.01	0.021	-0.05, 0.03	0.6
Regular Users	-0.01	0.027	-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, 0.07	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, 0.04	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, 0.31	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*

Characteristic	Beta	SE	95% CI	p-value
cidi_q8				
0	—	—	—	
1	-0.03	0.094	-0.22, 0.15	0.7
cidi_q9				
0	—	—	—	
1	-0.01	0.054	-0.12, 0.10	0.8
cidi_q10				
0	—	—	—	
1	-0.22	0.087	-0.39, -0.05	0.011*
cidi_q11				
0	—	—	—	
1	0.05	0.047	-0.04, 0.14	0.3
cidi_q12				
0	—	—	—	
1	0.11	0.091	-0.07, 0.29	0.2
cidi_q13				
0	—	—	—	
1	-0.02	0.078	-0.18, 0.13	0.8
cidi_q14				
0	—	—	—	
1	-0.04	0.042	-0.13, 0.04	0.3
cidi_q15				
0	—	—	—	
1	-0.18	0.077	-0.33, -0.03	0.022*
777	-0.03	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

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}"
)
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```


Characteristic	Beta	SE	95% CI	p-value
age_at_iview	0.00	0.001	0.00, 0.00	0.017*
is_case_clean				
Control	—	—	—	
Filtered Case	0.07	0.025	0.02, 0.12	0.003**
Unfiltered Case	0.01	0.044	-0.08, 0.09	0.8
msex				
Male	—	—	—	
Female	0.04	0.022	0.00, 0.08	0.063
educ_ord				
Primary or less	—	—	—	
Secondary	-0.02	0.018	-0.06, 0.01	0.2
College	-0.05	0.018	-0.08, -0.01	0.010**
bmi_bin				
Underweight	—	—	—	
Normal Weight	-0.03	0.030	-0.09, 0.03	0.3
Obese	-0.01	0.038	-0.09, 0.06	0.7
Overweight	-0.03	0.033	-0.10, 0.03	0.3
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	0.7
cidi_q1				
cidi_q10	—	—	—	
cidi_q11	0.01	0.043	-0.08, 0.09	0.9
cidi_q2				
0	—	—	—	
1	-0.02	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	0.064	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				

Characteristic	Beta	SE	95% CI	p-value
0	—	—	—	
1	0.17	0.076	0.02, 0.32	0.024*
cidi_q8				
0	—	—	—	
1	-0.04	0.095	-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	-0.04, 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	-0.17, 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				
Filtered Case * Female	-0.02	0.038	-0.09, 0.06	0.7
Unfiltered Case * Female	0.03	0.045	-0.06, 0.12	0.5

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}"
)

```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	Beta	SE	95% CI	p-value
age_at_iview	0.00	0.001	0.00, 0.00	0.015*
is_case_clean				
Control	—	—	—	
Filtered Case	0.05	0.033	-0.02, 0.11	0.2
Unfiltered Case	0.03	0.043	-0.05, 0.12	0.4
msex				
Male	—	—	—	
Female	0.04	0.017	0.01, 0.07	0.021*
educ_ord				
Primary or less	—	—	—	
Secondary	-0.03	0.026	-0.08, 0.02	0.3
College	-0.05	0.024	-0.10, -0.01	0.027*
bmi_bin				
Underweight	—	—	—	
Normal Weight	-0.03	0.030	-0.09, 0.03	0.3
Obese	-0.01	0.039	-0.09, 0.06	0.8
Overweight	-0.03	0.033	-0.10, 0.03	0.3
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	0.01	0.016	-0.03, 0.04	0.7
Regular Users	0.03	0.023	-0.02, 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	—	—	—	

Characteristic	Beta	SE	95% CI	p-value
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, -0.05	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, 0.14	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	0.026*
777	-0.02	0.042	-0.10, 0.06	0.6
cidi_q16	—	—	—	
0	—	—	—	
1	-0.18	0.155	-0.48, 0.13	0.2
cidi_q17	—	—	—	
0	—	—	—	
777	-0.16	0.081	-0.32, 0.00	0.055
is_case_clean * educ_ord				
Filtered Case * Secondary	0.03	0.044	-0.06, 0.11	0.5
Unfiltered Case * Secondary	-0.02	0.041	-0.10, 0.06	0.7
Filtered Case * College	0.03	0.049	-0.06, 0.13	0.5
Unfiltered Case * College	0.00	0.044	-0.09, 0.09	>0.9

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_cleanFiltered Case", 1],3)
ci_filt_ul <- round(confint(model_name)["is_case_cleanFiltered Case", 2],3)
ci_unfilt_ll <- round(confint(model_name)["is_case_cleanUnfiltered Case", 1],3)
ci_unfilt_ul <- round(confint(model_name)["is_case_cleanUnfiltered 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)

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)	-0.003 (-0.004, -0.001)	-0.002 (-0.004, 0)	-0.002 (-0.004, 0)
Control	Ref	Ref	Ref	Ref	Ref	Ref
Filtered Case	0.047 (-0.018, 0.112)	0.074 (0.025, 0.123)	0.067 (0.029, 0.105)	0.057 (0.02, 0.093)	0.064 (0.027, 0.1)	0.067 (0.03, 0.103)
Unfiltered Case	0.034 (-0.051, 0.118)	0.009 (-0.077, 0.094)	0.026 (-0.042, 0.094)	0.016 (-0.016, 0.049)	0.024 (-0.008, 0.056)	0.021 (-0.011, 0.053)