

# 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(batchtm)
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)

#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 one 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)

```

## Distribution of age among cases and controls

```
ggplot(tel_all, aes(x=age_at_iview, group = is_case, fill=is_case))+  
  facet_wrap(~msex) +  
  geom_density(alpha=0.4) +  
  xlab("Age") + ylab("Density") +  
  ggtitle("Age Distribution of Cases and Controls") +  
  ggeasy::easy_center_title() +  
  labs(fill='Case Status') +  
  #scale_y_continuous(labels = scales::percent_format(scale = 1)) +  
  theme_bw() +  
  scale_fill_discrete(name = "Case Status", labels = c("Control", "Case"))
```



## Analysis

```
#LTL appears to be linear associated with Age (p<0.0001)
fit1<-lm(ltl~Age, data=tel_all)
fit2<-lm(ltl~Age+ I(Age^2), data=tel_all)
fit3<-lm(ltl~Age+ I(Age^0.5), data=tel_all)
fit4<-lm(ltl~Age+ I(Age^2)+I(Age^3), data=tel_all)
stargazer(fit1, fit2, fit3, fit4, type = "text",
          title="Age at interview polynomial associations with LTL")
```

```
##
## Age at interview polynomial associations with LTL
## =====
##                               Dependent variable:
##                               -----
##                               ltl
##                               (1)          (2)          (3)          (4)
## -----
## Age                -0.002*          0.002          -0.009          -0.002
##                   (0.001)          (0.009)          (0.018)          (0.065)
##
## I(Age2)                -0.0001          0.0001
##                   (0.0001)          (0.002)
##
## I(Age0.5)                0.096
##                   (0.218)
##
## I(Age3)                -0.00000
##                   (0.00001)
##
## Constant              0.022          -0.052          -0.271          0.003
##                   (0.035)          (0.172)          (0.664)          (0.814)
## -----
## Observations              971          971          971          971
## R2                0.003          0.003          0.003          0.003
## Adjusted R2          0.002          0.001          0.001          0.0002
## Residual Std. Error  0.226 (df = 969)  0.226 (df = 968)  0.226 (df = 968)  0.226 (df = 967)
## F Statistic          2.961* (df = 1; 969)  1.577 (df = 2; 968)  1.577 (df = 2; 968)  1.052 (df = 3; 967)
## =====
## Note:                               *p<0.1; **p<0.05; ***p<0.01
```

## Bivariate Analysis (Unadjusted)

Age was not statistically associated with LTL ( $p=0.0856$ )

Case status ( $p=0.034$ ) and sex ( $p=0.0012$ ) are associated with LTL.

Participants who completed College from educ\_ord were also significantly associated with LTL ( $p=3e-04$ )...  
I'm not sure what to make of this.

The following variables were shown to be associated with LTL:

### 1. Binary

- msex,  $p= 0.0012$
- is\_case,  $p= 0.034$

### 2. Ordinal

- educ\_ord

Secondary,  $p=0.1236$  College,  $p= 3e-04$

```
results=NULL
for (ii in 1:length(vars)) {
  # Create a formula for the linear regression
  fmla <- as.formula(paste("ltl ~", vars[ii]))

  # Fit the linear regression model
  model <- lm(fmla, data = tel_all)

  fit<-lm(fmla, data= tel_all)
  res<-coef(summary(fit))
  ci<-round(confint(fit),3)

  if(nrow(res)==2){
    rownames(res)[2]<-vars[ii]
    rownames(ci)[2]<-vars[ii]
    est<-round(coef(fit)[2:length(coef(fit))],3)
    p<-coef(summary(fit))[2:nrow(coef(summary(fit))), "Pr(>|t|)"]
    if(p<=0.05){est<-paste(est, "*", sep="")}
    if(p<=0.001){est<-paste(est, "*", sep="")}
    if(p<=0.0001){est<-paste(est, "*", sep="")}
    ci<-paste(ci[vars[ii],], collapse=", ")
    ci<-paste("(", ci, ")", sep="")
    r2<-round(summary(fit)$adj.r.squared,3)
    p<-round(p, 4)
    p[p<0.0001]<-"<0.001"
    results<-rbind(results, c(vars[ii], est, ci, p, r2))
  }

  if(nrow(res)>2){
    varLine<-c(vars[ii], rep("", 3), round(summary(fit)$adj.r.squared,3))
    res<-res[grepl(vars[ii], rownames(res)),]
```

```

rownames(res)<-gsub(vars[ii], "", rownames(res))
est<-round(res[, "Estimate"],3)
ind<-which(apply(res, 1, function(x) x["Pr(>|t|)"]<0.05))
est[ind]<-paste(est[ind], "*", sep="")
ind<-which(apply(res, 1, function(x) x["Pr(>|t|)"]<0.001))
est[ind]<-paste(est[ind], "*", sep="")
ind<-which(apply(res, 1, function(x) x["Pr(>|t|)"]<0.0001))
est[ind]<-paste(est[ind], "*", sep="")
ci<-ci[grep(vars[ii], rownames(ci)),]
rownames(ci)<-gsub(vars[ii], "", rownames(ci))
ci<-apply(ci, 1, function(x) paste(x, collapse=" "))
ci<-paste("(", ci, ")", sep="")
p<-res[, "Pr(>|t|)"]
p<-round(p,4)
p[which(p<=0.0001)]<-"<0.0001"
temp<-cbind(rownames(res), est)
temp<-cbind(temp, ci)
temp<-cbind(temp, p)
temp<-cbind(temp, rep("", nrow(res)))
temp<-rbind(varLine, temp)
results<-rbind(results, temp)
22}
}

rownames(results)<-c(1:nrow(results))
colnames(results)<-c("Variable", "Beta", "CI", "p", "Adj R2")

kable(results, booktabs = TRUE, longtable = TRUE,
  caption = "Bivariate associations between phenotypes and LTL") %>%
  kable_styling(latex_options = c("hold_position", "repeat_header"))

```

Table 1: Bivariate associations between phenotypes and LTL

Variable	Beta	CI	p	Adj R2
age_at_iview	-0.002	(-0.003, 0)	0.0856	0.002
bmi	0.001	(-0.002, 0.005)	0.4421	0
is_case	0.031*	(0.002, 0.059)	0.034	0.004
msex	0.049*	(0.019, 0.079)	0.0012	0.01
assist_khat	-0.018	(-0.049, 0.013)	0.249	0
khat_3cat				-0.001
Irregular Users	-0.019	(-0.055, 0.017)	0.3078	
Regular Users	-0.017	(-0.063, 0.029)	0.4743	
assist_alcohol	-0.016	(-0.045, 0.012)	0.2649	0
hbp				-0.001
Low BP	0.102	(-0.212, 0.416)	0.5247	
High BP	0.028	(-0.033, 0.089)	0.3642	
assist_tobacco	-0.02	(-0.057, 0.017)	0.2811	0
tobacco_3cat				0
Irregular Users	-0.01	(-0.055, 0.036)	0.6767	
Regular Users	-0.037	(-0.092, 0.019)	0.1934	
assist_cannabis	-0.036	(-0.108, 0.035)	0.3223	0

Table 1: Bivariate associations between phenotypes and LTL (*continued*)

Variable	Beta	CI	p	Adj R2
cannabis_3cat				-0.001
Irregular Users	-0.031	(-0.106, 0.045)	0.4254	
Regular Users	-0.086	(-0.308, 0.137)	0.4492	
cidi_q9	-0.061	(-0.156, 0.034)	0.2101	0.001
obese	0.027	(-0.025, 0.078)	0.3042	0
bmi_bin				-0.002
Underweight	0.014	(-0.048, 0.076)	0.6516	
Overweight	0.008	(-0.027, 0.043)	0.6548	
Obese	0.03	(-0.023, 0.082)	0.2651	
hbp				-0.001
Low BP	0.102	(-0.212, 0.416)	0.5247	
High BP	0.028	(-0.033, 0.089)	0.3642	
bpm_bin				0
Slow Heartrate	-0.023	(-0.124, 0.077)	0.6468	
Fast Heartrate	0.044	(-0.017, 0.105)	0.1612	
cidi_q15	0	(0, 0)	0.7996	-0.001
age_group				0.002
31-40	-0.014	(-0.049, 0.02)	0.4161	
41-50	-0.027	(-0.067, 0.013)	0.1909	
51-60	-0.071*	(-0.136, -0.007)	0.0308	
educ_ord				0.012
Secondary	-0.027	(-0.062, 0.007)	0.1236	
College	-0.066**	(-0.101, -0.03)	3e-04	
Set				0.099
B	-0.09***	(-0.129, -0.051)	<0.0001	
C	0.031	(-0.007, 0.07)	0.1125	
D	0.114***	(0.075, 0.152)	<0.0001	
E	-0.046	(-0.135, 0.042)	0.3066	

```
#kable(results, caption="Bivariate associations between phenotypes and LTL.")
```



## Rosner Batch Correction

```
#####  
# Rosner batch correction  
#####  
  
# Step 1: biomarker =  $\Sigma[i(\text{batch}_i)] + (\text{covariate1}) + (\text{covariate2})$   
  
fit<-lm(ltl~Set, data=tel_all)  
  
# Step 2: calculate the average for batch:  $\text{avgb}=(\Sigma i)/N$  (Note:  $N=0$ )  
  
avgb<-mean(coef(fit)[grep("Set", names(coef(fit)))])  
  
# Step 3: recalibrate biomarker levels  
# if batch=1 then  $\text{adj\_biomarker}=\text{orig\_biomarker}-(1-\text{avgb})$   
# if batch=2 then  $\text{adj\_biomarker}=\text{orig\_biomarker}-(2-\text{avgb})$   
# ...  
# if batch=N then  $\text{adj\_biomarker}=\text{orig\_biomarker}-(N-\text{avgb})$  (Note:  $N=0$ )  
  
#### with package  
#install.packages("batchtma")  
  
ros_df <- tel_all %>% select(is_case, subjid, ltl, Set, Age)  
  
ros <- adjust_batch(  
  data = ros_df,  
  markers = ltl,  
  batch = Set,  
  method = simple  
)
```

# Multivariate Analysis

After adjusting for age, the following variables were shown to be associated with LTL:

1. Binary

- is\_case, p=0.0177
- msex, p=8e-04

3. Ordinal

- educ\_ord

Secondary, p=0.1119

College p<0.0001

```
results=NULL
vars<-vars[!vars%in%"age_at_iview"]

for(ii in 1:length(vars)){
  fmla<-as.formula(paste("ltl ~ age_at_iview +", vars[ii]))
  fit<-lm(fmla, data=tel_all)
  res<-coef(summary(fit))
  res<-res[-2, ]
  ci<-round(confint(fit),3)
  ci<-ci[-2,]
  if(nrow(res)==2){
    rownames(res)[2]<-vars[ii]
    rownames(ci)[2]<-vars[ii]
    est<-round(res[2, "Estimate"], 3)
    p<-res[2, "Pr(>|t|)"]
    if(p<=0.05){est<-paste(est, "*", sep="")}
    if(p<=0.001){est<-paste(est, "*", sep="")}
    if(p<=0.0001){est<-paste(est, "*", sep="")}
    ci<-paste(ci[vars[ii],], collapse=", ")
    ci<-paste("(", ci, ")", sep="")
    r2<-round(summary(fit)$adj.r.squared,3)
    p<-round(p, 4)
    p[p<0.0001]<-"<0.001"
    results<-rbind(results, c(vars[ii], est, ci, p, r2))
  }

  if(nrow(res)>2){
    varLine<-c(vars[ii], rep("", 3), round(summary(fit)$adj.r.squared,3))
    res<-res[grep(vars[ii], rownames(res)),]
    rownames(res)<-gsub(vars[ii], "", rownames(res))
    est<-round(res[, "Estimate"],3)
    ind<-which(apply(res, 1, function(x) x["Pr(>|t|)"]<0.05))
    est[ind]<-paste(est[ind], "*", sep="")
    ind<-which(apply(res, 1, function(x) x["Pr(>|t|)"]<0.001))
    est[ind]<-paste(est[ind], "*", sep="")
  }
}
```

```

ind<-which(apply(res, 1, function(x) x["Pr(>|t|)"]<0.0001))
est[ind]<-paste(est[ind], "*", sep="")
ci<-ci[grep(vars[ii], rownames(ci)),]
rownames(ci)<-gsub(vars[ii], "", rownames(ci))
24
ci<-apply(ci, 1, function(x) paste(x, collapse=", "))
ci<-paste("(", ci, ")", sep="")
p<-res[, "Pr(>|t|)"]
p<-round(p,4)
p[which(p<=0.0001)]<-"<0.0001"
temp<-cbind(rownames(res), est)
temp<-cbind(temp, ci)
temp<-cbind(temp, p)
temp<-cbind(temp, rep("", nrow(res)))
temp<-rbind(varLine, temp)
results<-rbind(results, temp)
}
}

rownames(results)<-c(1:nrow(results))
colnames(results)<-c("Variable", "Beta", "CI", "p", "Adj R2")

kable(results, booktabs = TRUE, longtable = TRUE,
  caption = "Associations between phenotype and LTL adjusted for age" %>%
  kable_styling(latex_options = c("hold_position", "repeat_header"))

```

Table 2: Associations between phenotype and LTL adjusted for age

Variable	Beta	CI	p	Adj R2
bmi	0.002	(-0.002, 0.005)	0.2905	0.002
is_case	0.035*	(0.006, 0.063)	0.0177	0.007
msex	0.051**	(0.021, 0.081)	8e-04	0.013
assist_khat	-0.018	(-0.049, 0.012)	0.2389	0.002
khat_3cat				0.001
Irregular Users	-0.018	(-0.054, 0.018)	0.3254	
Regular Users	-0.019	(-0.065, 0.027)	0.4167	
assist_alcohol	-0.016	(-0.045, 0.013)	0.2701	0.002
hbp				0.002
Low BP	0.11	(-0.203, 0.424)	0.4901	
High BP	0.04	(-0.022, 0.102)	0.2049	
assist_tobacco	-0.02	(-0.057, 0.016)	0.2794	0.002
tobacco_3cat				0.002
Irregular Users	-0.011	(-0.056, 0.034)	0.6437	
Regular Users	-0.035	(-0.09, 0.02)	0.2122	
assist_cannabis	-0.042	(-0.114, 0.03)	0.2486	0.002
cannabis_3cat				0.002
Irregular Users	-0.037	(-0.112, 0.039)	0.3418	
Regular Users	-0.093	(-0.316, 0.129)	0.4108	
cidi_q9	-0.051	(-0.148, 0.045)	0.2972	0.001

Table 2: Associations between phenotype and LTL adjusted for age (*continued*)

Variable	Beta	CI	p	Adj R2
obese	0.033	(-0.019, 0.085)	0.2153	0.003
bmi_bin				0.001
Underweight	0.015	(-0.047, 0.077)	0.6312	
Overweight	0.012	(-0.023, 0.047)	0.5068	
Obese	0.037	(-0.016, 0.09)	0.1723	
hbp				0.002
Low BP	0.11	(-0.203, 0.424)	0.4901	
High BP	0.04	(-0.022, 0.102)	0.2049	
bpm_bin				0.003
Slow Heartrate	-0.02	(-0.12, 0.08)	0.6939	
Fast Heartrate	0.049	(-0.013, 0.11)	0.1203	
cidi_q15	0	(0, 0)	0.7982	0.001
age_group				0.002
31-40	-0.031	(-0.083, 0.022)	0.2489	
41-50	-0.067	(-0.171, 0.037)	0.2072	
51-60	-0.128	(-0.279, 0.023)	0.0955	
educ_ord				0.016
Secondary	-0.028	(-0.063, 0.007)	0.1119	
College	-0.071***	(-0.107, -0.035)	<0.0001	
Set				0.103
B	-0.093***	(-0.132, -0.054)	<0.0001	
C	0.032	(-0.007, 0.07)	0.1073	
D	0.112***	(0.074, 0.151)	<0.0001	
E	-0.048	(-0.137, 0.04)	0.2841	

## Tables

### Case Status

```
label(tel_all$bmi_bin) <- "BMI"
label(tel_all$alcohol_3cat) <- "Alcohol Use"
label(tel_all$khat_3cat) <- "Khat Use"
label(tel_all$cannabis_3cat) <- "Cannabis Use"
label(tel_all$tobacco_3cat) <- "Tobacco Use"
label(tel_all$cidi_q1) <- "Arthritis and Rheumatism"
label(tel_all$cidi_q2) <- "Chronic back or neck problems"
label(tel_all$cidi_q3) <- "Frequent or severe headaches"
label(tel_all$cidi_q4) <- "Any other chronic pain"
label(tel_all$cidi_q5) <- "Seasonal allergies"
label(tel_all$cidi_q6) <- "Stroke"
label(tel_all$cidi_q7) <- "Heart attack"
label(tel_all$cidi_q8) <- "Heart disease"
label(tel_all$cidi_q9) <- "High blood pressure"
label(tel_all$cidi_q10) <- "Asthma"
label(tel_all$cidi_q11) <- "Tuberculosis"
label(tel_all$cidi_q12) <- "Other chronic lung disease"
label(tel_all$cidi_q13) <- "Diabetes"
label(tel_all$cidi_q14) <- "Stomach or intestine"
label(tel_all$cidi_q15) <- "HIV/AIDS"
label(tel_all$cidi_q16) <- "Epilepsy/Seizure"
label(tel_all$cidi_q17) <- "Cancer"

tbl <- table1(~ age_at_iview + msex + educ_ord +
  bmi_bin + alcohol_3cat + khat_3cat + cannabis_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 | is_case_f, data=tel_all, overall=

tbl <- t1flex(tbl)

tbl
```

	Case (N=487)	Control (N=484)
<b>Age</b>		
Mean (SD)	37.8 (7.88)	35.7 (7.68)
Median [Min, Max]	36.0 [25.0, 55.0]	33.5 [25.0, 55.0]
<b>Sex</b>		
Male	317 (65.1%)	315 (65.1%)
Female	170 (34.9%)	169 (34.9%)
<b>Education</b>		
Primary or less	161 (33.1%)	135 (27.9%)

	Case (N=487)	Control (N=484)
Secondary	212 (43.5%)	138 (28.5%)
College	114 (23.4%)	211 (43.6%)
<b>BMI</b>		
Underweight	28 (5.7%)	28 (5.8%)
Normal Weight	311 (63.9%)	307 (63.4%)
Obese	48 (9.9%)	33 (6.8%)
Overweight	100 (20.5%)	116 (24.0%)
<b>Alcohol Use</b>		
Never Users	233 (47.8%)	201 (41.5%)
Irregular Users	215 (44.1%)	188 (38.8%)
Regular Users	39 (8.0%)	95 (19.6%)
<b>Khat Use</b>		
Never Users	262 (53.8%)	404 (83.5%)
Irregular Users	144 (29.6%)	52 (10.7%)
Regular Users	81 (16.6%)	28 (5.8%)
<b>Cannabis Use</b>		
Never Users	453 (93.0%)	478 (98.8%)
Irregular Users	30 (6.2%)	6 (1.2%)
Regular Users	4 (0.8%)	0 (0%)
<b>Tobacco Use</b>		
Never Users	307 (63.0%)	484 (100%)
Irregular Users	110 (22.6%)	0 (0%)
Regular Users	70 (14.4%)	0 (0%)
<b>Arthritis and Rheumatism</b>		
0	443 (91.0%)	484 (100%)
1	42 (8.6%)	0 (0%)
777	2 (0.4%)	0 (0%)
<b>Chronic back or neck problems</b>		
0	423 (86.9%)	484 (100%)
1	63 (12.9%)	0 (0%)
777	1 (0.2%)	0 (0%)
<b>Frequent or severe headaches</b>		
0	406 (83.4%)	484 (100%)
1	79 (16.2%)	0 (0%)

	Case (N=487)	Control (N=484)
777	2 (0.4%)	0 (0%)
<b>Any other chronic pain</b>		
0	472 (96.9%)	484 (100%)
1	15 (3.1%)	0 (0%)
<b>Seasonal allergies</b>		
0	461 (94.7%)	484 (100%)
1	26 (5.3%)	0 (0%)
<b>Stroke</b>		
0	484 (99.4%)	484 (100%)
1	1 (0.2%)	0 (0%)
777	2 (0.4%)	0 (0%)
<b>Heart attack</b>		
0	474 (97.3%)	484 (100%)
1	12 (2.5%)	0 (0%)
777	1 (0.2%)	0 (0%)
<b>Heart disease</b>		
0	476 (97.7%)	484 (100%)
1	8 (1.6%)	0 (0%)
777	3 (0.6%)	0 (0%)
<b>High blood pressure</b>		
0	458 (94.0%)	484 (100%)
1	22 (4.5%)	0 (0%)
777	7 (1.4%)	0 (0%)
<b>Asthma</b>		
0	477 (97.9%)	484 (100%)
1	8 (1.6%)	0 (0%)
777	2 (0.4%)	0 (0%)
<b>Tuberculosis</b>		
0	457 (93.8%)	484 (100%)
1	27 (5.5%)	0 (0%)
777	3 (0.6%)	0 (0%)
<b>Other chronic lung disease</b>		
0	475 (97.5%)	484 (100%)
1	9 (1.8%)	0 (0%)

	Case (N=487)	Control (N=484)
777	3 (0.6%)	0 (0%)
<b>Diabetes</b>		
0	465 (95.5%)	484 (100%)
1	9 (1.8%)	0 (0%)
777	13 (2.7%)	0 (0%)
<b>Stomach or intestine</b>		
0	451 (92.6%)	484 (100%)
1	34 (7.0%)	0 (0%)
777	2 (0.4%)	0 (0%)
<b>HIV/AIDS</b>		
0	440 (90.3%)	484 (100%)
1	10 (2.1%)	0 (0%)
777	37 (7.6%)	0 (0%)
<b>Epilepsy/Seizure</b>		
0	480 (98.6%)	484 (100%)
1	4 (0.8%)	0 (0%)
777	3 (0.6%)	0 (0%)
<b>Cancer</b>		
0	476 (97.7%)	484 (100%)
777	11 (2.3%)	0 (0%)

## Tables

### Set

```
tbl <- table1(~ age_at_iview + msex + educ_ord +
  bmi_bin + alcohol_3cat + khat_3cat + cannabis_3cat + tobacco_3cat +
  cidi_q1 + cidi_q2 + cidi_q3 + cidi_q4 + cidi_q5 + cidi_q6 +
  cidi_q7 + cidi_q8 + cidi_q9 | Set, data=tel_all, overall=F)
```

```
tbl <- kbl(tbl, booktabs = T) %>% kable_styling(full_width = F)
```

```
tbl
```

```
tbl <- table1(~ cidi_q10 + cidi_q11 + cidi_q12 +
  cidi_q13 + cidi_q14 + cidi_q15 + cidi_q16 + cidi_q17 | Set, data=tel_all, overall=F)
```

```
tbl <- kbl(tbl, booktabs = T) %>% kable_styling(full_width = F)
```



	A	B	C	D	E
	(N=235)	(N=236)	(N=238)	(N=237)	(N=25)
Age					
Mean (SD)	37.2 (7.73)	35.7 (7.56)	37.4 (8.29)	36.5 (7.79)	36.2 (7.22)
Median [Min, Max]	36.0 [25.0, 55.0]	34.0 [25.0, 55.0]	36.0 [25.0, 55.0]	35.0 [25.0, 55.0]	36.0 [25.0, 52.0]
Sex					
Male	142 (60.4%)	170 (72.0%)	145 (60.9%)	158 (66.7%)	17 (68.0%)
Female	93 (39.6%)	66 (28.0%)	93 (39.1%)	79 (33.3%)	8 (32.0%)
Education					
Primary or less	63 (26.8%)	75 (31.8%)	76 (31.9%)	78 (32.9%)	4 (16.0%)
Secondary	96 (40.9%)	83 (35.2%)	75 (31.5%)	87 (36.7%)	9 (36.0%)
College	76 (32.3%)	78 (33.1%)	87 (36.6%)	72 (30.4%)	12 (48.0%)
BMI					
Underweight	9 (3.8%)	23 (9.7%)	11 (4.6%)	11 (4.6%)	2 (8.0%)
Normal Weight	149 (63.4%)	159 (67.4%)	147 (61.8%)	151 (63.7%)	12 (48.0%)
Obese	28 (11.9%)	17 (7.2%)	17 (7.1%)	17 (7.2%)	2 (8.0%)
Overweight	49 (20.9%)	37 (15.7%)	63 (26.5%)	58 (24.5%)	9 (36.0%)
Alcohol Use					
Never Users	107 (45.5%)	93 (39.4%)	122 (51.3%)	98 (41.4%)	14 (56.0%)
Irregular Users	92 (39.1%)	106 (44.9%)	84 (35.3%)	111 (46.8%)	10 (40.0%)
Regular Users	36 (15.3%)	37 (15.7%)	32 (13.4%)	28 (11.8%)	1 (4.0%)
Khat Use					
Never Users	179 (76.2%)	150 (63.6%)	168 (70.6%)	153 (64.6%)	16 (64.0%)
Irregular Users	43 (18.3%)	48 (20.3%)	49 (20.6%)	52 (21.9%)	4 (16.0%)
Regular Users	13 (5.5%)	38 (16.1%)	21 (8.8%)	32 (13.5%)	5 (20.0%)
Cannabis Use					
Never Users	226 (96.2%)	224 (94.9%)	231 (97.1%)	225 (94.9%)	25 (100%)
Irregular Users	9 (3.8%)	10 (4.2%)	6 (2.5%)	11 (4.6%)	0 (0%)
Regular Users	0 (0%)	2 (0.8%)	1 (0.4%)	1 (0.4%)	0 (0%)
Tobacco Use					
Never Users	191 (81.3%)	183 (77.5%)	206 (86.6%)	191 (80.6%)	20 (80.0%)
Irregular Users	33 (14.0%)	29 (12.3%)	14 (5.9%)	33 (13.9%)	1 (4.0%)
Regular Users	11 (4.7%)	24 (10.2%)	18 (7.6%)	13 (5.5%)	4 (16.0%)
Arthritis and Rheumatism					
0	226 (96.2%)	229 (97.0%)	224 (94.1%)	223 (94.1%)	25 (100%)
1	9 (3.8%)	7 (3.0%)	12 (5.0%)	14 (5.9%)	0 (0%)
777	0 (0%)	0 (0%)	2 (0.8%)	0 (0%)	0 (0%)
Chronic back or neck problems					
0	218 (92.8%)	223 (94.5%)	222 (93.3%)	219 (92.4%)	25 (100%)
1	17 (7.2%)	13 (5.5%)	15 (6.3%)	18 (7.6%)	0 (0%)
777	0 (0%)	0 (0%)	1 (0.4%)	0 (0%)	0 (0%)
Frequent or severe headaches					
0	218 (92.8%)	218 (92.4%)	216 (90.8%)	214 (90.3%)	24 (96.0%)
1	17 (7.2%)	17 (7.2%)	21 (8.8%)	23 (9.7%)	1 (4.0%)
777	0 (0%)	1 (0.4%)	1 (0.4%)	0 (0%)	0 (0%)
Any other chronic pain					
0	232 (98.7%)	233 (98.7%)	233 (97.9%)	233 (98.3%)	25 (100%)
1	3 (1.3%)	3 (1.3%)	5 (2.1%)	4 (1.7%)	0 (0%)
Seasonal allergies					
0	228 (97.0%)	231 (97.9%)	233 (97.9%)	228 (96.2%)	25 (100%)
1	7 (3.0%)	5 (2.1%)	5 (2.1%)	9 (3.8%)	0 (0%)
Stroke		17			
0	234 (99.6%)	235 (99.6%)	238 (100%)	236 (99.6%)	25 (100%)
1	0 (0%)	0 (0%)	0 (0%)	1 (0.4%)	0 (0%)
777	1 (0.4%)	1 (0.4%)	0 (0%)	0 (0%)	0 (0%)
Heart attack					

	A	B	C	D	E
	(N=235)	(N=236)	(N=238)	(N=237)	(N=25)
Asthma					
0	235 (100%)	230 (97.5%)	236 (99.2%)	236 (99.6%)	24 (96.0%)
1	0 (0%)	4 (1.7%)	2 (0.8%)	1 (0.4%)	1 (4.0%)
777	0 (0%)	2 (0.8%)	0 (0%)	0 (0%)	0 (0%)
Tuberculosis					
0	229 (97.4%)	229 (97.0%)	229 (96.2%)	231 (97.5%)	23 (92.0%)
1	6 (2.6%)	5 (2.1%)	9 (3.8%)	5 (2.1%)	2 (8.0%)
777	0 (0%)	2 (0.8%)	0 (0%)	1 (0.4%)	0 (0%)
Other chronic lung disease					
0	232 (98.7%)	231 (97.9%)	237 (99.6%)	235 (99.2%)	24 (96.0%)
1	2 (0.9%)	4 (1.7%)	1 (0.4%)	2 (0.8%)	0 (0%)
777	1 (0.4%)	1 (0.4%)	0 (0%)	0 (0%)	1 (4.0%)
Diabetes					
0	229 (97.4%)	229 (97.0%)	233 (97.9%)	233 (98.3%)	25 (100%)
1	2 (0.9%)	4 (1.7%)	2 (0.8%)	1 (0.4%)	0 (0%)
777	4 (1.7%)	3 (1.3%)	3 (1.3%)	3 (1.3%)	0 (0%)
Stomach or intestine					
0	232 (98.7%)	224 (94.9%)	229 (96.2%)	226 (95.4%)	24 (96.0%)
1	2 (0.9%)	11 (4.7%)	9 (3.8%)	11 (4.6%)	1 (4.0%)
777	1 (0.4%)	1 (0.4%)	0 (0%)	0 (0%)	0 (0%)
HIV/AIDS					
0	218 (92.8%)	231 (97.9%)	227 (95.4%)	223 (94.1%)	25 (100%)
1	5 (2.1%)	1 (0.4%)	2 (0.8%)	2 (0.8%)	0 (0%)
777	12 (5.1%)	4 (1.7%)	9 (3.8%)	12 (5.1%)	0 (0%)
Epilepsy/Seizure					
0	233 (99.1%)	232 (98.3%)	237 (99.6%)	237 (100%)	25 (100%)
1	2 (0.9%)	1 (0.4%)	1 (0.4%)	0 (0%)	0 (0%)
777	0 (0%)	3 (1.3%)	0 (0%)	0 (0%)	0 (0%)
Cancer					
0	231 (98.3%)	234 (99.2%)	234 (98.3%)	236 (99.6%)	25 (100%)
777	4 (1.7%)	2 (0.8%)	4 (1.7%)	1 (0.4%)	0 (0%)

tbl

## Propensity score analysis

Predicting case status using all covariates

```
glm_fit <- glm(factor(is_case) ~ age_at_iview + factor(msex) + educ_ord +
  bmi_bin + alcohol_3cat + khat_3cat + cannabis_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,
  data=tel_all,
```

```

        family="binomial")

#tbl_regression(glm_fit, exp=TRUE)

glm_probs = data.frame(probs = predict(glm_fit, type="response"), tel_all$is_case_f)
glm_probs$is_case <- tel_all$is_case

res <- as.data.frame(coef(summary(glm_fit)))

res <- res %>% mutate(ci_ll = round(exp(Estimate - qnorm(0.975)*`Std. Error`),3))
res <- res %>% mutate(ci_ul = round(exp(Estimate + qnorm(0.975)*`Std. Error`),3))
res <- res %>% mutate(ci = paste0("(",ci_ll,",",ci_ul,")"))

res <- res %>% mutate(OR = exp(Estimate)) %>%
  mutate(OR_round = case_when(
    OR < 100 ~ sprintf("%.2f", OR),
    OR>100 ~ "Error"))

res <- res %>% mutate(pval = round(`Pr(>|z|)`,3)) %>%
  mutate(pval2 = case_when(
    pval<0.001 ~ paste0(pval,"***"),
    pval<0.01 & pval>0 ~ paste0(pval,"**"),
    pval<0.05 & pval>0 ~ paste0(pval,"*"),
    TRUE ~ as.character(pval)
  ))

res <- res %>% select(OR_round, ci, pval2)

rownames(res) <- c("Intercept", "Age", "Female",
  "Education - Secondary", "Education - College",
  "BMI - Underweight", "BMI- Overweight", "BMI - Obese",
  "Alcohol - Irregular", "Alcohol - Regular",
  "Khat - Irregular", "Khat - Regular",
  "Cannabis - Irregular", "Cannabis - Regular",
  "Tobacco - Irregular", "Tobacco - Regular",
  "Arthritis - Yes", "Arthritis - Missing",
  "Chronic back/neck problems - Yes",
  "Chronic back/neck problems - Missing",
  "Headaches - Yes", "Headaches - Missing",
  "Other chronic pain - Yes",
  "Allergies - Yes",
  "Stroke - Yes", "Stroke - Missing",
  "Heart attack - Yes", "Heart Attack - Missing",
  "Heart disease - Yes", "Heart Disase - Missing",
  "High blood pressure - Yes", "High blood pressure - Missing",
  "Asthma - Yes", "Asthma - Missing",
  "Tuberculosis - Yes", "Tuberculosis - Missing",
  "Other chronic lung disease - Yes",
  "Other chronic lung disease - Missing",

```

```

      "Diabetes - Yes", "Diabetes - Missing",
      "Ulcer - Yes", "Ulcer - Missing",
      "HIV/AIDS - Yes", "HIV/AIDS - Missing",
      "Epilepsy/Seizure - Yes",
      "Cancer - Missing")

kable(res, col.names = c("Coefficient", "OR", "95% CI", "p-value")) %>%
  kable_classic(full_width = F) %>% add_footnote(paste("Psuedo R2 = ", efronRSquared(glm_fit)))

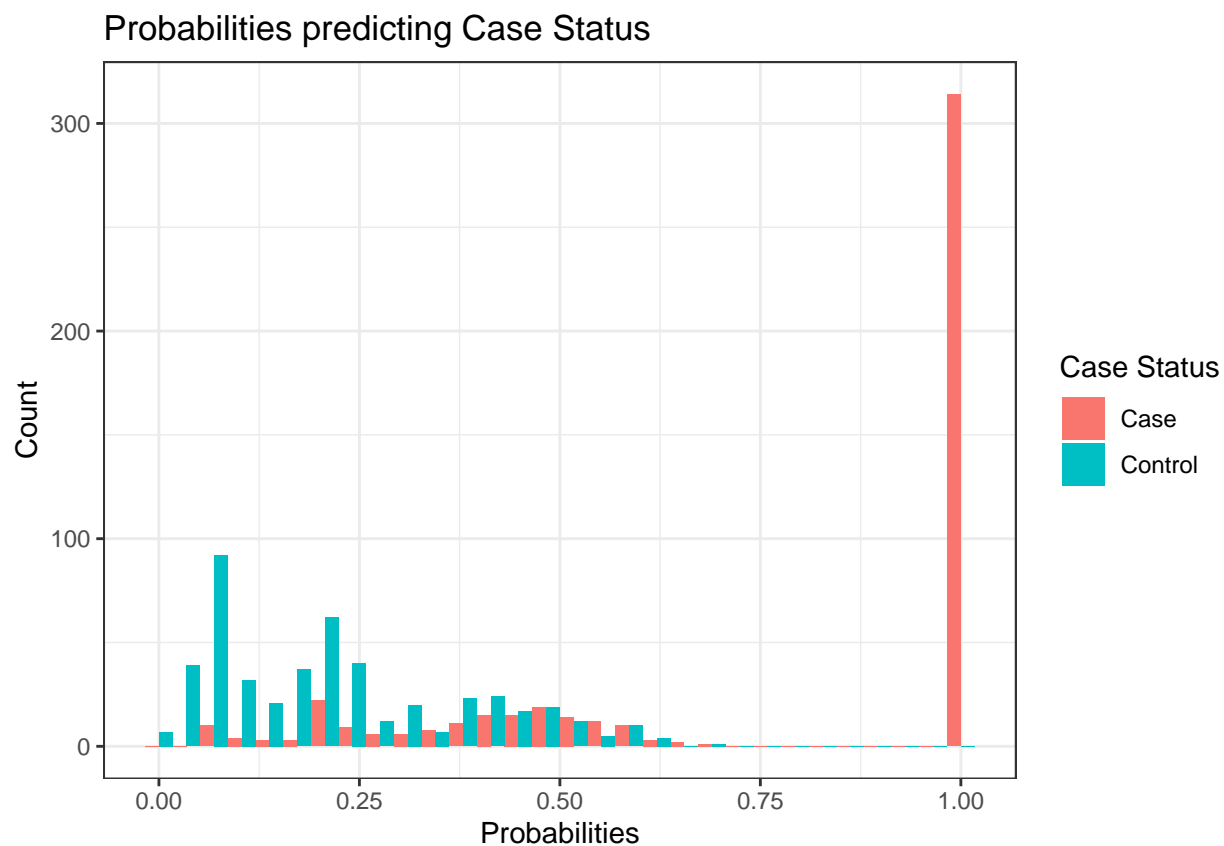
```

Coefficient	OR	95% CI	p-value
Intercept	0.27	(0.069,1.026)	0.055
Age	1.02	(0.995,1.045)	0.121
Female	0.88	(0.567,1.353)	0.551
Education - Secondary	1.33	(0.847,2.093)	0.215
Education - College	0.40	(0.24,0.654)	0***
BMI - Underweight	1.43	(0.572,3.577)	0.444
BMI- Overweight	2.55	(0.858,7.557)	0.092
BMI - Obese	1.47	(0.555,3.882)	0.44
Alcohol - Irregular	0.31	(0.199,0.471)	0***
Alcohol - Regular	0.08	(0.034,0.199)	0***
Khat - Irregular	1.65	(0.914,2.977)	0.096
Khat - Regular	0.94	(0.391,2.262)	0.891
Cannabis - Irregular	0.00	(0,Inf)	0.988
Cannabis - Regular	0.00	(0,Inf)	0.999
Tobacco - Irregular	Error	(0,Inf)	0.982
Tobacco - Regular	Error	(0,Inf)	0.985
Arthritis - Yes	Error	(0,Inf)	0.992
Arthritis - Missing	2.00	(0,Inf)	1
Chronic back/neck problems - Yes	Error	(0,Inf)	0.99
Chronic back/neck problems - Missing	0.00	(0,Inf)	0.997
Headaches - Yes	Error	(0,Inf)	0.99
Headaches - Missing	Error	(0,Inf)	0.997
Other chronic pain - Yes	3.47	(0,Inf)	1
Allergies - Yes	Error	(0,Inf)	0.993
Stroke - Yes	0.00	(0,Inf)	0.998
Stroke - Missing	0.00	(0,Inf)	0.996
Heart attack - Yes	0.30	(0,Inf)	1
Heart Attack - Missing	0.00	(0,Inf)	0.998
Heart disease - Yes	Error	(0,Inf)	0.997
Heart Disease - Missing	Error	(0,Inf)	0.999
High blood pressure - Yes	Error	(0,Inf)	0.994
High blood pressure - Missing	0.33	(0,Inf)	1
Asthma - Yes	Error	(0,Inf)	0.997
Asthma - Missing	0.00	(0,Inf)	0.998
Tuberculosis - Yes	Error	(0,Inf)	0.994
Tuberculosis - Missing	2.84	(0,Inf)	1
Other chronic lung disease - Yes	Error	(0,Inf)	0.996
Other chronic lung disease - Missing	4.74	(0,Inf)	1
Diabetes - Yes	Error	(0,Inf)	0.997

Diabetes - Missing	Error	(0,Inf)	0.996
Ulcer - Yes	Error	(0,Inf)	0.993
Ulcer - Missing	0.00	(0,Inf)	0.999
HIV/AIDS - Yes	Error	(0,Inf)	0.998
HIV/AIDS - Missing	Error	(0,Inf)	0.992
Epilepsy/Seizure - Yes	Error	(0,Inf)	0.997
Cancer - Missing	Error	(0,Inf)	0.996

```
ggplot(data=glm_probs, aes(x=probs, fill=tel_all.is_case_f)) +
  geom_histogram(position="dodge") +
  ggtitle("Probabilities predicting Case Status") +
  xlab("Probabilities") +
  ylab("Count") +
  guides(fill = guide_legend(title = "Case Status")) +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Propensity score analysis for set

```
#With chronic conditions
#Missingness in cidi_q9 and cidi_q13

glm_fit <- multinom(factor(Set) ~ age_at_iview + factor(msex) + educ_ord +
  bmi_bin + alcohol_3cat + khat_3cat + cannabis_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,
  data=tel_all)
```

```
## # weights: 240 (188 variable)
## initial value 1562.764213
## iter 10 value 1384.195516
## iter 20 value 1331.352221
## iter 30 value 1320.585962
## iter 40 value 1316.832197
## iter 50 value 1314.470208
## iter 60 value 1313.613017
## iter 70 value 1313.487470
## iter 80 value 1313.446385
## iter 90 value 1313.434546
## iter 100 value 1313.429003
## final value 1313.429003
## stopped after 100 iterations
```

```
# tidy(glm_fit, conf.int = TRUE) %>%
# kable() %>%
# kable_styling("basic", full_width = FALSE,
#               latex_options = "scale_down")
```

```
tbl_regression(glm_fit, exp = TRUE)
```

```
## i Multinomial models have a different underlying structure than the models
## gtsummary was designed for. Other gtsummary functions designed to work with
## tbl_regression objects may yield unexpected results.
```

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

Outcome	Characteristic	OR	95% CI	p-value
B	Age	0.97	0.95, 1.00	0.042
	factor(msex)			
	Male	—	—	
	Female	0.82	0.51, 1.32	0.4
	Education			
	Primary or less	—	—	
	Secondary	0.66	0.41, 1.07	0.092
	College	0.75	0.45, 1.23	0.3

Outcome	Characteristic	OR	95% CI	p-value
	BMI			
	Underweight	—	—	
	Normal Weight	0.34	0.15, 0.82	0.015
	Obese	0.26	0.09, 0.77	0.015
	Overweight	0.29	0.11, 0.75	0.010
	Alcohol Use			
	Never Users	—	—	
	Irregular Users	1.23	0.79, 1.90	0.4
	Regular Users	0.96	0.54, 1.72	>0.9
	Khat Use			
	Never Users	—	—	
	Irregular Users	1.70	0.91, 3.18	0.093
	Regular Users	3.96	1.67, 9.42	0.002
	Cannabis Use			
	Never Users	—	—	
	Irregular Users	0.74	0.25, 2.14	0.6
	Regular Users	131,020,801	31,787,706, 540,034,255	<0.001
	Tobacco Use			
	Never Users	—	—	
	Irregular Users	0.62	0.29, 1.33	0.2
	Regular Users	0.93	0.34, 2.59	0.9
	Arthritis and Rheumatism			
	Asthma	—	—	
	Tuberculosis	0.70	0.18, 2.73	0.6
	777	0.01	0.01, 0.01	<0.001
	Chronic back or neck problems			
	0	—	—	
	1	0.65	0.23, 1.82	0.4
	777	0.12	0.12, 0.12	<0.001
	Frequent or severe headaches			
	0	—	—	
	1	0.80	0.32, 2.00	0.6
	777	3.42		
	Any other chronic pain			
	0	—	—	
	1	0.58	0.07, 4.86	0.6
	Seasonal allergies			
	0	—	—	
	1	0.52	0.12, 2.26	0.4
	Stroke			
	0	—	—	
	1	0.01		
	777	0.00	0.00, 0.00	<0.001
	Heart attack			
	0	—	—	
	1	1.04	0.11, 9.96	>0.9
	777	0.00	0.00, 0.00	<0.001
	Heart disease			

Outcome	Characteristic	OR	95% CI	p-value
C	0	—	—	
	1	7,091,086	1,901,324, 26,446,574	<0.001
	777	3.59	0.03, 440	0.6
	High blood pressure			
	0	—	—	
	1	1.43	0.29, 7.12	0.7
	777	5.52	0.03, 883	0.5
	Asthma			
	0	—	—	
	1	12,854,804,039	3,694,455,664, 44,728,101,211	<0.001
	777	3,856,318,629,538,427	3,855,644,690,325,668, 3,856,992,686,550,957	<0.001
	Tuberculosis			
	0	—	—	
	1	1.44	0.36, 5.79	0.6
	777	1.19	1.19, 1.19	<0.001
	Other chronic lung disease			
	0	—	—	
	1	5.03	0.31, 82.7	0.3
	777	0.00	0.00, 0.00	<0.001
	Diabetes			
	0	—	—	
	1	3.08	0.45, 21.2	0.3
	777	0.73	0.06, 8.20	0.8
	Stomach or intestine			
	0	—	—	
	1	7.67	1.36, 43.2	0.021
	777	246,750	246,747, 246,753	<0.001
	HIV/AIDS			
	0	—	—	
	1	0.39	0.04, 4.18	0.4
	777	0.20	0.04, 0.98	0.047
	Epilepsy/Seizure			
	0	—	—	
	1	0.23	0.00, 11.4	0.5
	777	107,678,680,707,522,736	107,659,862,530,800,944, 107,697,502,173,527,712	<0.001
	Cancer			
	0	—	—	
	777	0.00	0.00, 0.00	<0.001
	Age factor(msex)	1.00	0.98, 1.03	>0.9
	Male	—	—	
	Female	1.01	0.64, 1.59	>0.9
	Education			
	Primary or less	—	—	
	Secondary	0.63	0.39, 1.02	0.059
	College	0.96	0.59, 1.57	0.9
	BMI			
	Underweight	—	—	
	Normal Weight	0.71	0.27, 1.86	0.5



Outcome	Characteristic	OR	95% CI	p-value
	Obese	0.49	0.15, 1.54	0.2
	Overweight	1.10	0.40, 3.05	0.8
	Alcohol Use			
	Never Users	—	—	
	Irregular Users	0.82	0.53, 1.26	0.4
	Regular Users	0.70	0.39, 1.25	0.2
	Khat Use			
	Never Users	—	—	
	Irregular Users	2.13	1.16, 3.91	0.015
	Regular Users	2.52	1.00, 6.33	0.049
	Cannabis Use			
	Never Users	—	—	
	Irregular Users	0.74	0.22, 2.47	0.6
	Regular Users	88,961,117	17,166,585, 461,016,574	<0.001
	Tobacco Use			
	Never Users	—	—	
	Irregular Users	0.22	0.09, 0.52	<0.001
	Regular Users	0.92	0.33, 2.57	0.9
	Arthritis and Rheumatism			
	Asthma	—	—	
	Tuberculosis	1.29	0.41, 4.01	0.7
	777	116,592,899	116,592,899, 116,592,900	<0.001
	Chronic back or neck problems			
	0	—	—	
	1	0.92	0.36, 2.35	0.9
	777	1,088	1,088, 1,088	<0.001
	Frequent or severe headaches			
	0	—	—	
	1	1.50	0.63, 3.54	0.4
	777	392		
	Any other chronic pain			
	0	—	—	
	1	1.48	0.26, 8.55	0.7
	Seasonal allergies			
	0	—	—	
	1	0.41	0.10, 1.79	0.2
	Stroke			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	0.00	0.00, 0.00	<0.001
	Heart attack			
	0	—	—	
	1	0.10	0.01, 2.00	0.13
	777	0.00	0.00, 0.00	<0.001
	Heart disease			
	0	—	—	
	1	15,578,354	4,328,565, 56,065,951	<0.001
	777	2.28	0.02, 323	0.7

Outcome	Characteristic	OR	95% CI	p-value
D	High blood pressure			
	0	—	—	
	1	0.55	0.10, 2.96	0.5
	777	1.58	0.01, 271	0.9
	Asthma			
	0	—	—	
	1	7,077,832,784	1,837,633,900, 27,260,988,671	<0.001
	777	0.08	0.08, 0.08	<0.001
	Tuberculosis			
	0	—	—	
	1	1.82	0.53, 6.27	0.3
	777	0.01		
	Other chronic lung disease			
	0	—	—	
	1	0.98	0.04, 26.2	>0.9
	777	0.00	0.00, 0.00	<0.001
	Diabetes			
	0	—	—	
	1	1.22	0.14, 10.5	0.9
	777	1.07	0.16, 7.14	>0.9
	Stomach or intestine			
	0	—	—	
	1	6.79	1.21, 38.1	0.029
	777	0.00	0.00, 0.00	<0.001
	HIV/AIDS			
	0	—	—	
	1	0.40	0.05, 3.29	0.4
	777	0.96	0.34, 2.70	>0.9
	Epilepsy/Seizure			
	0	—	—	
	1	1.17	0.05, 29.0	>0.9
	777	0.03	0.03, 0.03	<0.001
	Cancer			
	0	—	—	
	777	1.32	0.25, 7.12	0.7
	Age factor(msex)	0.99	0.96, 1.01	0.3
	Male	—	—	
	Female	0.87	0.55, 1.39	0.6
	Education			
	Primary or less	—	—	
	Secondary	0.75	0.47, 1.20	0.2
	College	0.71	0.43, 1.17	0.2
	BMI			
	Underweight	—	—	
	Normal Weight	0.77	0.29, 2.01	0.6
	Obese	0.52	0.16, 1.62	0.3
	Overweight	1.02	0.37, 2.80	>0.9
	Alcohol Use			
	Never Users	—	—	
	Irregular Users	1.33	0.87, 2.04	0.2

Outcome	Characteristic	OR	95% CI	p-value
	Regular Users	0.72	0.39, 1.31	0.3
	Khat Use			
	Never Users	—	—	
	Irregular Users	1.91	1.04, 3.50	0.036
	Regular Users	4.47	1.88, 10.6	<0.001
	Cannabis Use			
	Never Users	—	—	
	Irregular Users	0.98	0.34, 2.80	>0.9
	Regular Users	159,156,860	31,195,433, 812,006,883	<0.001
	Tobacco Use			
	Never Users	—	—	
	Irregular Users	0.48	0.23, 0.99	0.048
	Regular Users	0.35	0.12, 1.05	0.060
	Arthritis and Rheumatism			
	Asthma	—	—	
	Tuberculosis	1.62	0.55, 4.82	0.4
	777	0.01	0.01, 0.01	<0.001
	Chronic back or neck problems			
	0	—	—	
	1	0.82	0.34, 2.00	0.7
	777	0.25	0.25, 0.25	<0.001
	Frequent or severe headaches			
	0	—	—	
	1	1.20	0.52, 2.78	0.7
	777	0.12		
	Any other chronic pain			
	0	—	—	
	1	0.59	0.08, 4.21	0.6
	Seasonal allergies			
	0	—	—	
	1	0.84	0.24, 2.96	0.8
	Stroke			
	0	—	—	
	1	20,075,119,220,975	20,075,119,220,952, 20,075,119,220,998	<0.001
	777	0.03	0.03, 0.03	<0.001
	Heart attack			
	0	—	—	
	1	0.67	0.07, 6.19	0.7
	777	0.00	0.00, 0.00	<0.001
	Heart disease			
	0	—	—	
	1	11,227,360	3,288,907, 38,326,906	<0.001
	777	0.00	0.00, 0.00	<0.001
	High blood pressure			
	0	—	—	
	1	1.13	0.24, 5.37	0.9
	777	4.08	0.03, 606	0.6

Outcome	Characteristic	OR	95% CI	p-value
E	Asthma			
	0	—	—	
	1	2,214,169,781	467,895,478, 10,477,869,639	<0.001
	777	0.00	0.00, 0.00	<0.001
	Tuberculosis			
	0	—	—	
	1	1.13	0.29, 4.33	0.9
	777	9,104,847,967,347	9,104,847,967,162, 9,104,847,967,532	<0.001
	Other chronic lung disease			
	0	—	—	
	1	1.21	0.04, 33.7	>0.9
	777	0.00	0.00, 0.00	<0.001
	Diabetes			
	0	—	—	
	1	0.49	0.04, 6.45	0.6
	777	1.17	0.19, 7.17	0.9
	Stomach or intestine			
	0	—	—	
	1	7.24	1.33, 39.3	0.022
	777	0.00	0.00, 0.00	<0.001
	HIV/AIDS			
	0	—	—	
	1	0.34	0.05, 2.49	0.3
	777	1.11	0.43, 2.89	0.8
	Epilepsy/Seizure			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	0.00	0.00, 0.00	<0.001
	Cancer			
	0	—	—	
	777	0.35	0.03, 3.72	0.4
	Age factor(msex)	0.97	0.91, 1.04	0.4
	Male	—	—	
	Female	0.87	0.30, 2.58	0.8
	Education			
	Primary or less	—	—	
	Secondary	1.69	0.42, 6.83	0.5
	College	2.81	0.70, 11.2	0.14
	BMI			
	Underweight	—	—	
	Normal Weight	0.36	0.06, 2.24	0.3
	Obese	0.43	0.04, 4.53	0.5
	Overweight	1.29	0.19, 8.58	0.8
	Alcohol Use			
	Never Users	—	—	
	Irregular Users	0.95	0.36, 2.49	>0.9
	Regular Users	0.16	0.02, 1.44	0.10
	Khat Use			
	Never Users	—	—	
	Irregular Users	1.72	0.36, 8.08	0.5

Outcome	Characteristic	OR	95% CI	p-value
	Regular Users	6.63	1.16, 37.8	0.033
	Cannabis Use			
	Never Users	—	—	
	Irregular Users	0.00	0.00, 0.00	<0.001
	Regular Users	0.00	0.00, 0.00	<0.001
	Tobacco Use			
	Never Users	—	—	
	Irregular Users	0.34	0.03, 3.62	0.4
	Regular Users	1.91	0.27, 13.7	0.5
	Arthritis and Rheumatism			
	Asthma	—	—	
	Tuberculosis	0.00	0.00, 0.00	<0.001
	777	0.23	0.23, 0.23	<0.001
	Chronic back or neck problems			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	1.14	1.14, 1.14	<0.001
	Frequent or severe headaches			
	0	—	—	
	1	0.71	0.05, 10.2	0.8
	777	1.35	1.35, 1.35	<0.001
	Any other chronic pain			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	Seasonal allergies			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	Stroke			
	0	—	—	
	1	1.18	1.18, 1.18	<0.001
	777	1.37	1.37, 1.37	<0.001
	Heart attack			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	1.08	1.08, 1.08	<0.001
	Heart disease			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	0.02	0.02, 0.02	<0.001
	High blood pressure			
	0	—	—	
	1	12.5	0.92, 171	0.058
	777	2.11	2.11, 2.11	<0.001
	Asthma			
	0	—	—	
	1	10,847	10,847, 10,847	<0.001
	777	1.18	1.18, 1.18	<0.001
	Tuberculosis			

Outcome	Characteristic	OR	95% CI	p-value
	0	—	—	
	1	9.16	1.17, 71.8	0.035
	777	1.26	1.26, 1.26	<0.001
	Other chronic lung disease			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	6,650,678	6,650,673, 6,650,682	<0.001
	Diabetes			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	0.00	0.00, 0.00	<0.001
	Stomach or intestine			
	0	—	—	
	1	13.8	0.88, 217	0.062
	777	0.00	0.00, 0.00	<0.001
	HIV/AIDS			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	0.00	0.00, 0.00	<0.001
	Epilepsy/Seizure			
	0	—	—	
	1	0.00	0.00, 0.00	<0.001
	777	1.39	1.39, 1.39	<0.001
	Cancer			
	0	—	—	
	777	0.00	0.00, 0.00	<0.001

```

glm_probs = data.frame(probs = predict(glm_fit, type="probs"), tel_all$Set,
                        tel_all$is_case_f)

predA <- ggplot(data=glm_probs, aes(x=probs.A, fill=tel_all.Set)) + geom_histogram(position="dodge") +
facet_wrap(~tel_all.is_case_f) + ggtitle("Probabilities predicting Set A") +
  xlab("Probabilities") +
  ylab("Count") +
  guides(fill = guide_legend(title = "Set")) +
  theme_bw()

predB <- ggplot(data=glm_probs, aes(x=probs.B, fill=tel_all.Set)) + geom_histogram(position="dodge") +
facet_wrap(~tel_all.is_case_f) + ggtitle("Probabilities predicting Set B") +
  xlab("Probabilities") +
  ylab("Count") +
  guides(fill = guide_legend(title = "Set")) +
  theme_bw()

predC <- ggplot(data=glm_probs, aes(x=probs.C, fill=tel_all.Set)) + geom_histogram(position="dodge") +
facet_wrap(~tel_all.is_case_f) + ggtitle("Probabilities predicting Set C") +
  xlab("Probabilities") +
  ylab("Count") +
  guides(fill = guide_legend(title = "Set")) +
  theme_bw()

```

```

predD <- ggplot(data=glm_probs, aes(x=probs.D, fill=tel_all.Set)) + geom_histogram(position="dodge") +
facet_wrap(~tel_all.is_case_f) + ggtitle("Probabilities predicting Set D") +
  xlab("Probabilities") +
  ylab("Count") +
  guides(fill = guide_legend(title = "Set")) +
  theme_bw()

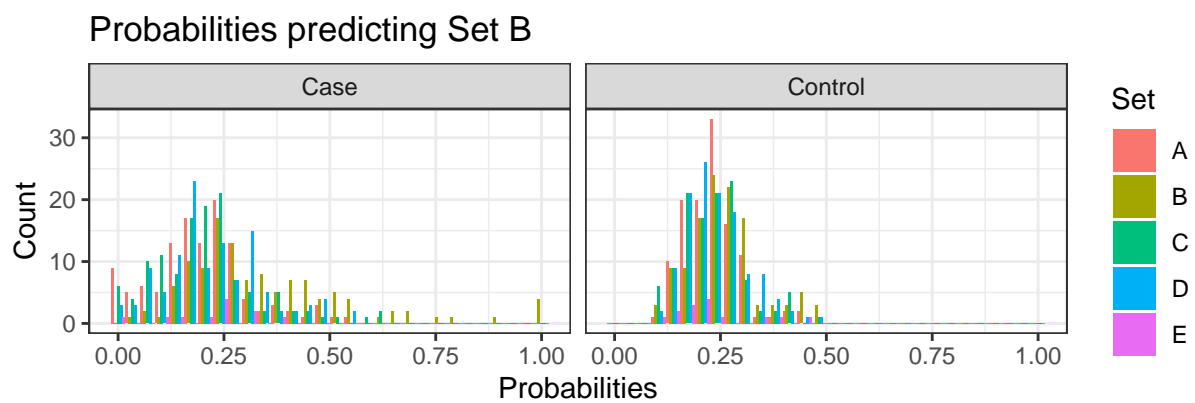
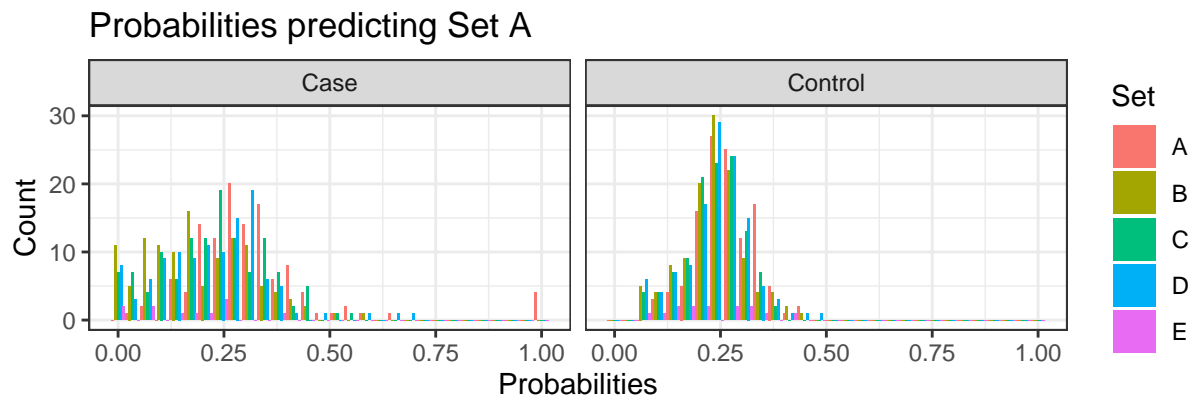
predE <- ggplot(data=glm_probs, aes(x=probs.E, fill=tel_all.Set)) + geom_histogram(position="dodge") +
facet_wrap(~tel_all.is_case_f) + ggtitle("Probabilities predicting Set E") +
  xlab("Probabilities") +
  ylab("Count") +
  guides(fill = guide_legend(title = "Set")) +
  theme_bw()

predF <- ggplot(data = NULL) +
  geom_blank()

```

```
predA / predB
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

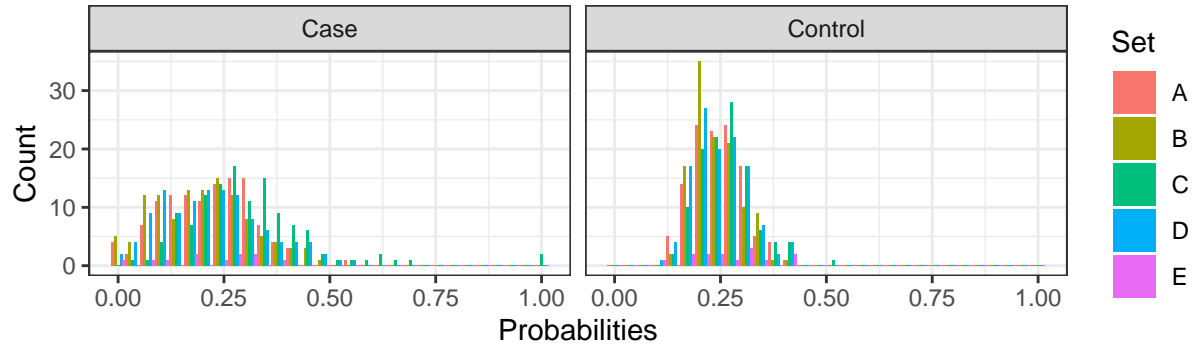


```
predC / predD
```

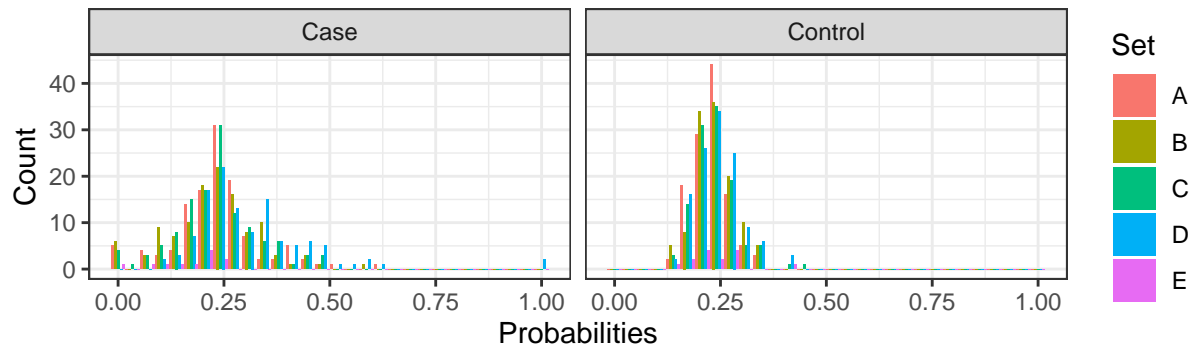
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.  
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



### Probabilities predicting Set C

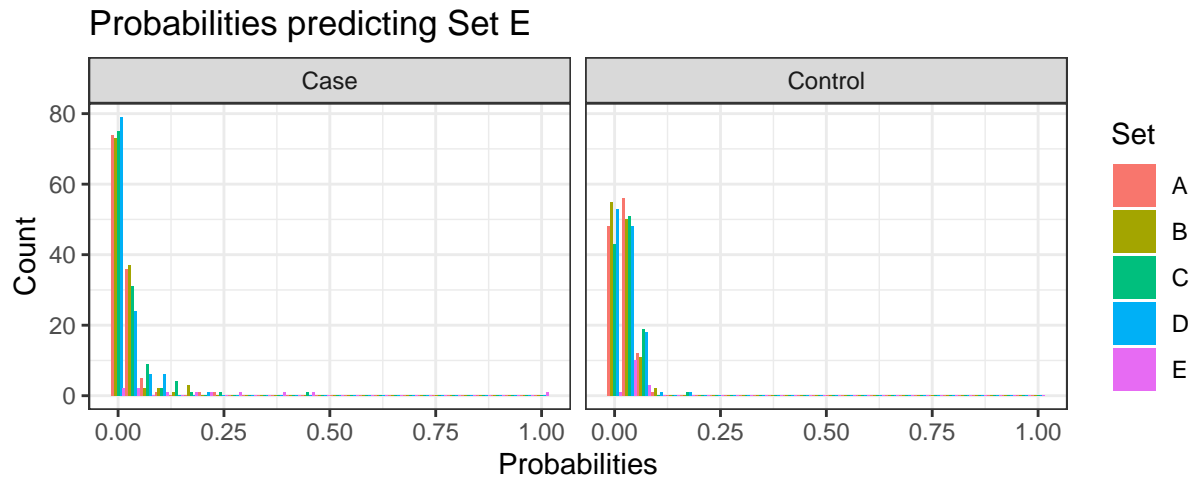


### Probabilities predicting Set D



```
predE / predF
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



## Comparing freeze 7 vs freeze 8

```
#High Blood Pressure

ngap_7$hbp<-0
ngap_7[which(ngap_7$bp_over<=90 | ngap_7$bp_under<=60),
        "hbp"]<-"Low BP"
ngap_7[which(ngap_7$bp_over>=140 | ngap_7$bp_under>=90),
        "hbp"]<-"High BP"
ngap_7[which((ngap_7$bp_over > 90 & ngap_7$bp_over < 140) |
              (ngap_7$bp_under > 60 & ngap_7$bp_under <= 90)),
        "hbp"] <- "Normal BP"

ngap_7$hbp <- factor(ngap_7$hbp, levels = c("Normal BP",
                                             "Low BP", "High BP"))

ngap_7$hbp <- relevel(ngap_7$hbp, ref = "Normal BP")

tab<-sumTab(ngap_7$hbp)
names(tab)[which(is.na(names(tab)))]<-"Missing"

#Obesity
```

```

ngap_7$bmi_bin<-0
ngap_7[which(ngap_7$bmi<=18.5), "bmi_bin"]<-"Underweight"
ngap_7[which(ngap_7$bmi>18.5 & ngap_7$bmi<=24.9),
      "bmi_bin"] <-"Normal Weight"
ngap_7[which(ngap_7$bmi>24.9 & ngap_7$bmi<30),
      "bmi_bin"] <- "Overweight"
ngap_7[which(ngap_7$bmi>=30),
      "bmi_bin"] <- "Obese"
ngap_7[which(is.na(ngap_7$bmi)), "bmi_bin"] <- NA

ngap_7$bmi_bin<-factor(ngap_7$bmi_bin, levels=c("Normal Weight",
                                                "Underweight",
                                                "Overweight",
                                                "Obese"
                                                ))
ngap_7$bmi_bin <- relevel(ngap_7$bmi_bin, ref = "Normal Weight")

tab<-sumTab(ngap_7$bmi_bin)
names(tab)[which(is.na(names(tab)))]<-"Missing"

ngap_7_new <- ngap_7 %>% select(subj_id, bmi_bin, hbp, cidi_q8, cidi_q9, cidi_q13) %>%
  rename(
    subj_id = subj_id,
    ngap_7_bmi_bin = bmi_bin,
    ngap_7_hbp = hbp,
    ngap_7_cidi_q8 = cidi_q8,
    ngap_7_cidi_q9 = cidi_q9,
    ngap_7_cidi_q13 = cidi_q13
  )

ngap_7_merge <- merge(ngap_7_new, tel_all, by="subj_id")
ngap_7_merge <- ngap_7_merge %>% select( subj_id,
    ngap_7_bmi_bin, bmi_bin,
    ngap_7_hbp, hbp,
    ngap_7_cidi_q8, cidi_q8,
    ngap_7_cidi_q9, cidi_q9,
    ngap_7_cidi_q13, cidi_q13)

```

## Comparing cardiometabolic variables

There are no differences in cardiometabolic conditions between Freeze 7 and Freeze 8.

```

#BMI
tab <- table(ngap_7_merge$ngap_7_bmi_bin, ngap_7_merge$bmi_bin)
kable(tab)

```

	Underweight	Normal Weight	Obese	Overweight
Normal Weight	0	618	0	0
Underweight	56	0	0	0
Overweight	0	0	0	216
Obese	0	0	81	0

*#High Blood Pressure - Calculated*

```
tab <- table(ngap_7_merge$ngap_7_hbp, ngap_7_merge$ngap_7_hbp)
kable(tab)
```

	Normal BP	Low BP	High BP
Normal BP	913	0	0
Low BP	0	2	0
High BP	0	0	56

*#Heart Disease*

```
tab <- table(ngap_7_merge$ngap_7_cidi_q8, ngap_7_merge$cidi_q8)
kable(tab)
```

	0	1	777
0	960	0	0
1	0	8	0
777	0	0	3

*#High Blood Pressure - CIDI*

```
tab <- table(ngap_7_merge$ngap_7_cidi_q9, ngap_7_merge$cidi_q9)
kable(tab)
```

	0	1	777
0	942	0	0
1	0	22	0
777	0	0	7

*#Diabetes or high blood sugar*

```
tab <- table(ngap_7_merge$ngap_7_cidi_q13, ngap_7_merge$cidi_q13)
kable(tab)
```

	0	1	777
0	949	0	0
1	0	9	0
777	0	0	13

## 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") |

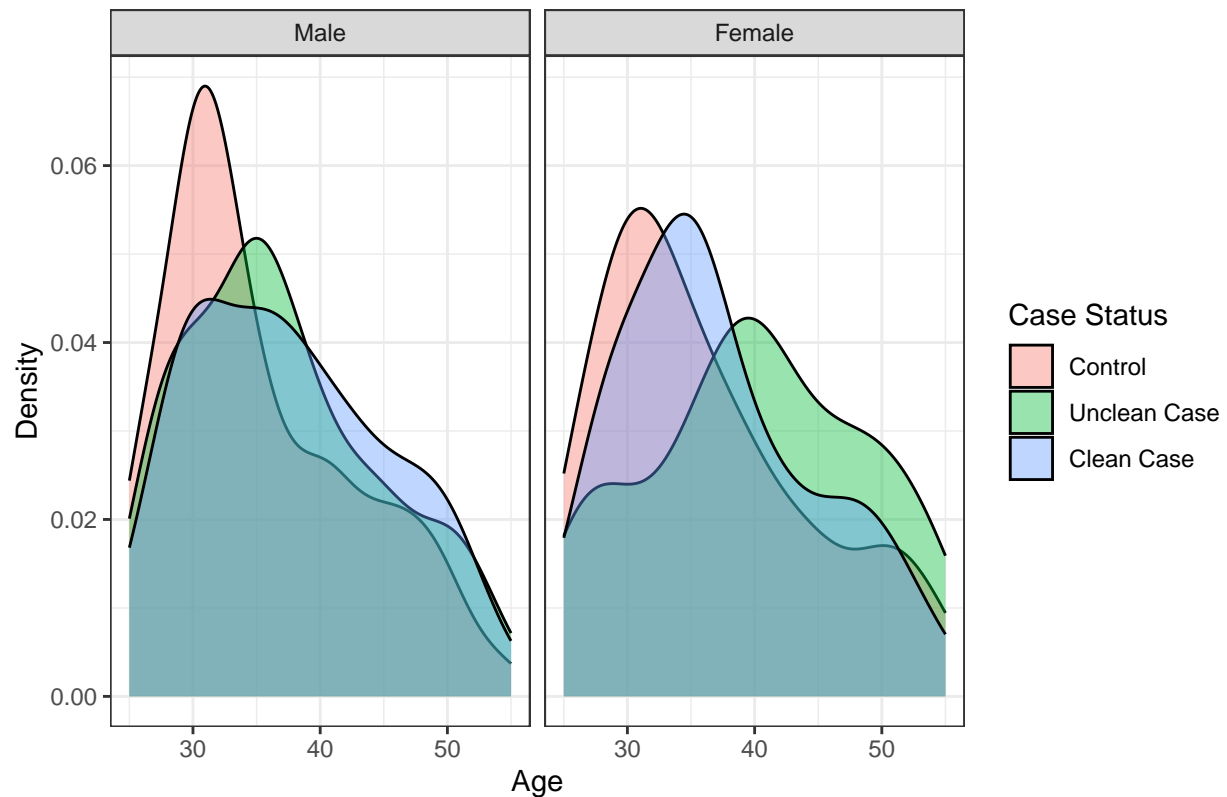
  cidi_q1 == 1 |
  cidi_q2 == 1 |
  cidi_q3 == 1 |
  cidi_q4 == 1 |
  cidi_q5 == 1 |
  cidi_q6 == 1 |
  cidi_q7 == 1 |
  cidi_q8 == 1 |
  cidi_q9 == 1 |
  cidi_q10 == 1 |
  cidi_q11 == 1 |
  cidi_q12 == 1 |
  cidi_q13 == 1 |
  cidi_q14 == 1 |
  cidi_q15 == 1 |
  cidi_q16 == 1 |
  cidi_q17 == 1) ~ "Unclean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

Age Distribution of Clean/Unclean Cases and Controls



```
tbl <- table1(~ age_at_iview + educ_ord +
  bmi_bin + alcohol_3cat + khat_3cat + cannabis_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,
  data=tel_all, overall=F)

tbl <- tbl %>%
  kbl() %>%
  kable_styling(full_width = FALSE)

tbl
```

Rosner batch correction with clean/unclean cases

Models

	Control	Unclean Case	Clean Case	Control	Unclean Case
	(N=315)	(N=214)	(N=103)	(N=169)	(N=85)
Age					
Mean (SD)	35.4 (7.41)	37.1 (7.70)	37.7 (7.66)	36.2 (8.15)	40.0 (8.60)
Median [Min, Max]	33.0 [25.0, 55.0]	36.0 [25.0, 55.0]	36.0 [25.0, 55.0]	34.0 [25.0, 55.0]	40.0 [25.0, 55.0]
Education					
Primary or less	83 (26.3%)	61 (28.5%)	35 (34.0%)	52 (30.8%)	34 (40.0%)
Secondary	75 (23.8%)	94 (43.9%)	39 (37.9%)	63 (37.3%)	38 (44.7%)
College	157 (49.8%)	59 (27.6%)	29 (28.2%)	54 (32.0%)	13 (15.3%)
BMI					
Underweight	19 (6.0%)	15 (7.0%)	6 (5.8%)	9 (5.3%)	5 (5.9%)
Normal Weight	229 (72.7%)	148 (69.2%)	72 (69.9%)	78 (46.2%)	47 (55.3%)
Obese	5 (1.6%)	10 (4.7%)	6 (5.8%)	28 (16.6%)	17 (20.0%)
Overweight	62 (19.7%)	41 (19.2%)	19 (18.4%)	54 (32.0%)	16 (18.8%)
Alcohol Use					
Never Users	118 (37.5%)	49 (22.9%)	74 (71.8%)	83 (49.1%)	46 (54.1%)
Irregular Users	116 (36.8%)	136 (63.6%)	23 (22.3%)	72 (42.6%)	36 (42.4%)
Regular Users	81 (25.7%)	29 (13.6%)	6 (5.8%)	14 (8.3%)	3 (3.5%)
Khat Use					
Never Users	240 (76.2%)	42 (19.6%)	73 (70.9%)	164 (97.0%)	68 (80.0%)
Irregular Users	47 (14.9%)	102 (47.7%)	22 (21.4%)	5 (3.0%)	16 (18.8%)
Regular Users	28 (8.9%)	70 (32.7%)	8 (7.8%)	0 (0%)	1 (1.2%)
Cannabis Use					
Never Users	311 (98.7%)	181 (84.6%)	103 (100%)	167 (98.8%)	84 (98.8%)
Irregular Users	4 (1.3%)	29 (13.6%)	0 (0%)	2 (1.2%)	1 (1.2%)
Regular Users	0 (0%)	4 (1.9%)	0 (0%)	0 (0%)	0 (0%)
Tobacco Use					
Never Users	315 (100%)	43 (20.1%)	103 (100%)	169 (100%)	76 (89.4%)
Irregular Users	0 (0%)	102 (47.7%)	0 (0%)	0 (0%)	8 (9.4%)
Regular Users	0 (0%)	69 (32.2%)	0 (0%)	0 (0%)	1 (1.2%)
Arthritis and Rheumatism					
0	315 (100%)	193 (90.2%)	103 (100%)	169 (100%)	62 (72.9%)
1	0 (0%)	20 (9.3%)	0 (0%)	0 (0%)	22 (25.9%)
777	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	1 (1.2%)
Chronic back or neck problems					
0	315 (100%)	175 (81.8%)	103 (100%)	169 (100%)	60 (70.6%)
1	0 (0%)	38 (17.8%)	0 (0%)	0 (0%)	25 (29.4%)
777	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	0 (0%)
Frequent or severe headaches					
0	315 (100%)	172 (80.4%)	103 (100%)	169 (100%)	46 (54.1%)
1	0 (0%)	40 (18.7%)	0 (0%)	0 (0%)	39 (45.9%)
777	0 (0%)	2 (0.9%)	0 (0%)	0 (0%)	0 (0%)
Any other chronic pain					
0	315 (100%)	206 (96.3%)	103 (100%)	169 (100%)	78 (91.8%)
1	0 (0%)	8 (3.7%)	0 (0%)	0 (0%)	7 (8.2%)
Seasonal allergies					
0	315 (100%)	200 (93.5%)	103 (100%)	169 (100%)	73 (85.9%)
1	0 (0%)	14 (6.5%)	0 (0%)	0 (0%)	12 (14.1%)
Stroke					
0	315 (100%)	211 (98.6%)	103 (100%)	169 (100%)	85 (100%)
1	0 (0%)	1 (0.5%)	0 (0%)	0 (0%)	0 (0%)
777	0 (0%)	2 (0.9%)	0 (0%)	0 (0%)	0 (0%)
Heart attack					
0	315 (100%)	208 (97.2%)	103 (100%)	169 (100%)	78 (91.8%)
1	0 (0%)	39 (2.8%)	0 (0%)	0 (0%)	6 (7.1%)
777	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.2%)
Heart disease					
0	315 (100%)	209 (97.7%)	103 (100%)	169 (100%)	80 (94.1%)
1	0 (0%)	5 (2.3%)	0 (0%)	0 (0%)	3 (3.5%)