Differences in Resting-State Functional MRI Quality Control Recommendations Alter the Representativeness of the Adolescent Brain Cognitive Development (ABCD) Study

Supplemental Methods and Results

Matthew Peverill1, Justin D. Russell, Taylor J. Keding, Hailey M. Rich, Max A. Halvorson, Kevin M. King, Rasmus Birn, & Ryan Herringa

2024-04-12

Contents

List of Tables

1 Supplemental Methods

This section includes R code used to generate tables, figures, and statistics used in the main manuscript. Additional code (specifically, for pre-processing and coding the raw data) is available on osf: https://osf.io/57xer/.

```r
recodesid <- function(x) {
  paste0("NDAR",substr(x, start = 6, stop = nchar(x)))
}


t1<-readRDS("ABCD-Data-Import/tabulateddata.Rds")
t2<-readRDS("ABCC_download/ABCCsubjectdata.RDS")
df<-left_join(readRDS("ABCD-Data-Import/tabulateddata.Rds") %>%
                mutate(subjectkey=recodesid(as.character(subjectkey))),
             readRDS("ABCC_download/ABCCsubjectdata.RDS"),
             by=c("subjectkey"="sname"))

df<-df %>%
  mutate(subjectkeychar=subjectkey,
         subjectkey=as.factor(subjectkey)) %>%
  # Pt's without ABCC data should have FALSE for the 'Scan Exists' booleans
  mutate(across(starts_with("Exists"),
                ~ case_when(. == "True" ~ TRUE,
                            . == "False" ~ FALSE,
                            is.na(.) ~ FALSE))) %>%
  # Pt's without ABCC data should have TRUE for > 5 minute data booleans.
  mutate(across(starts_with("ltfive"),
                ~ replace_na(.,TRUE))) %>%
  #Standardize psychopathology scores
  mutate(across(pfactor:EXT,scale)) %>%
```

3

```
  # Are there any rsfmri images in ABCC?
  mutate(any_rest = (Exists_rest_1 |
                      Exists_rest_2 |
                      Exists_rest_3 |
                      Exists_rest_4)) %>%
  #Are there >375 low motion frames per threshold
  mutate(across(starts_with("gframes"),\(x) {x>375},.names = "{.col}_375")) %>%
  mutate(across(ends_with("375"),~replace_na(.,FALSE)))

imgincl<-readRDS("ABCD-fastqc01/imginclusion.Rds")


df<-left_join(df,imgincl[,1:2],by=join_by(subjectkeychar==subjectkey)) %>%
  mutate(fastqcok=replace_na(fastqcok,FALSE))

# N.B. A few (8) Pt's have neuroimaging data but not tabular data.
# Possibly they have removed consent? They are not included in this
# analysis
```

```
options(tinytex.clean = FALSE)
# Seed for random number generation
set.seed(42)
knitr::opts_chunk$set(cache=TRUE)
knitr::opts_chunk$set(cache.extra = knitr::rand_seed)
```

1.1 Additional Coding and Releveling

```
cvars <- c(
    "adi",
    "coi",
    "nihtbx_flanker_agecorrected",
    "nihtbx_cryst_agecorrected",
    "nihtbx_totalcomp_agecorrected",
    "pea_wiscv_tss",
    "pfactor",
    "INT",
    "EXT",
    "interview_age",
    "bmi"
  )


# Z scores
df <- df %>% mutate(across(ends_with(cvars),scale,.names = "{.col}.z"))
```

```r
# Relevel factors to set comparison to modal levels

df$p.edu<-relevel(factor(df$p.edu,ordered=FALSE),4)
df$household.income<-relevel(df$household.income,5)
```

1.2 Descriptives

1.2.1 Figure 1 Table 1: Inclusion counts and inconsistencies

```r
setsdf<- df %>%
  select(c(subjectkey,ABCD_rsfmri_QC1,ABCD_rsfmri_QC2,any_rest)) %>%
  rename("T"="ABCD_rsfmri_QC1",
         "R"="ABCD_rsfmri_QC2",
         "C"="any_rest") %>%
  rowwise %>%
  mutate(sets = list(names(select(.,
          where(is.logical)))[c_across(where(is.logical))])) %>%
      ungroup
```

Figure 1 panel b:

```r
# This code also generates table 1
levelsdf<-data.frame(
  Level = c(
    "Full (F)",
    "ABCD 4 Tabulated (T)",
    "ABCC (C)",
    "ABCC < .5mm",
    "ABCC < .4mm",
    "ABCC < .3mm",
    "ABCD 4 Recommended (R)",
    "ABCD < .2mm",
    "ABCD < .1mm"
  ),
  Description = c(
    "Entire sample",
    "rsFMRI tabulated data available (at least 1 T1 and 1 rsFMRI were complete and passed visual inspection)",
    "rs-fMRI data in the ABCD Community Collection",
    "ABCC + censoring at threshold*",
    "\"",
    "\"",
```

```r
      "T1 and rs-fMRI recommended in 'ABCD Recommended Imaging Inclusion' table (includes 375 frames at <.2mm FD)",
      "ABCC + censoring at .2mm threshold",
      "\""
    ),
    Count = c(
      nrow(df),
      sum(df$ABCD_rsfmri_QC1),
      sum(df$any_rest),
      sum(df$gframes_0.5_375),
      sum(df$gframes_0.4_375),
      sum(df$gframes_0.3_375),
      sum(df$ABCD_rsfmri_QC2),
      sum(df$gframes_0.2_375),
      sum(df$gframes_0.1_375)
    )
)


levelsdf$Levelf<-factor(levelsdf$Level,levels=c(
      "Full (F)",
      "ABCD 4 Tabulated (T)",
      "ABCC (C)",
      "ABCC < .5mm",
      "ABCC < .4mm",
      "ABCC < .3mm",
      "ABCD 4 Recommended (R)",
      "ABCD < .2mm",
      "ABCD < .1mm"
    ),
    labels = c(
      "Full (**F**)",
      "ABCD 4 Tabulated (**T**)",
      "ABCC (**C**)",
      "ABCC < **.5**mm",
      "ABCC < **.4**mm",
      "ABCC < **.3**mm",
      "ABCD 4 Recommended (**R**)",
      "ABCC < **.2**mm",
      "ABCC < **.1**mm"
    ))

library(ggtext)
samplecounts<-ggplot(levelsdf,aes(x=Levelf,y=Count,label=Count)) +
  labs(title="B. n by Condition") +
  geom_point(color="#c5050c",size=4,alpha=0.6) +
```

```
    geom_segment(aes(x=Levelf,xend=Levelf,y=0,yend=Count),color="#646569") +
    scale_x_discrete(limits = rev(levels(levelsdf$Levelf))) +
    coord_flip() +
    theme(axis.ticks.y=element_blank(),
          axis.title.y=element_blank(),
          axis.title.x=element_blank(),
          axis.text.y=element_markdown(),
          plot.title=element_text()) +
    geom_text(nudge_x = .25,nudge_y=-1500,family="Times New Roman")
```

Figure 1 panel C:

```
upsetfig<-ggplot(setsdf,aes(x=sets)) +
  theme(axis.title.x = element_blank()) +
  ggtitle("C. Inclusion Inconsistencies ") +
  ylab("n") +
  geom_bar() +
  geom_text(stat='count',
            aes(label=after_stat(count)),
            vjust=-.5,
            size=10 / .pt,
            family="Times New Roman") +
  expand_limits(y=c(0,8800)) +
  scale_x_upset()
```

```
ggsave("Manuscript/Figures/inclusioncounts.svg",
       (plot_spacer() + (samplecounts / upsetfig + plot_layout(heights=c(3,2)))) +
         plot_layout(widths=c(3,2)),
       width=7.5,height=5)
```

(panel A is added using a vector graphics program.)

Table 1:

```
tables.xlsx<-openxlsx::createWorkbook()
levelstablehux<-as_hux(levelsdf[,1:2]) %>%
  theme_basic() %>%
  set_font_size(10) %>%
  set_width(1.8) %>%
  set_col_width(c(.25,.5,.25)) %>%
  set_background_color(NA)
tables.xlsx<-as_Workbook(levelstablehux,
                Workbook=tables.xlsx,
                sheet="LevelsTable")
```

1.2.2 Categorical Variable Figure (Figure 2)

```r
bools<-c("full",
         "ABCD_rsfmri_QC1",
         "any_rest",
         "gframes_0.5_375",
         "gframes_0.4_375",
         "gframes_0.3_375",
         "ABCD_rsfmri_QC2",
         "gframes_0.2_375",
         "gframes_0.1_375")


catvars<-c("sex",
           "household.income",
           "p.edu",
           "race_ethnicity.factor",
           "ksads_factor",
           "pds_category")

#This replaces NA with 'missing' for tabulation and reverses the scrubbing vars.

dfcatna <- df %>% select(c(ends_with(catvars), ends_with(bools))) %>%
  mutate(across(
    ends_with(catvars),
    ~ factor(
      .x,
      levels = levels(addNA(.x)),
      labels = c(levels(.x), "Missing"),
      exclude = NULL
    )
  ))

catvar375figdf<-data.frame(variable=character(),
                           level=character())
for (var in catvars) {
  levels <- levels(dfcatna[[var]])
  var_df <- data.frame(
    variable = var,
    level = levels
  )
  # Append the data frame to the levels_df
  catvar375figdf <- rbind(catvar375figdf, var_df)
}
```

```r
dfcatna$full=TRUE

for(b in bools) {
  n<-sum(dfcatna[,b])
  v<-sapply(1:nrow(catvar375figdf),\(x) {
    vname<-catvar375figdf[x,"variable"]
    tlevel<-catvar375figdf[x,"level"]
    vmatch<-dfcatna[[vname]]==tlevel
    sum(vmatch & dfcatna[[b]])/n
  })
  catvar375figdf[,b]<-v*100
}

catvar375figdflong <- catvar375figdf %>%
  pivot_longer(full:gframes_0.1_375) %>%
  mutate(thresh=factor(name,levels=bools,labels=c("F",
                                                  "T",
                                                  "C",
                                                  ".5",
                                                  ".4",
                                                  ".3",
                                                  "R",
                                                  ".2",
                                                  ".1"))) %>%
  mutate(levelf=as.factor(level)) %>%
  mutate(variable=as.factor(variable))

cp1<-catvar375figdflong %>%
  filter(variable=="sex") %>%
  mutate(level=as.factor(level)) %>%
  ggplot(aes(x=thresh,y=value,fill=level)) +
      geom_bar(position="fill",stat="identity") +
      scale_fill_viridis(discrete = TRUE) +
      labs(title="Sex Assigned at Birth") +
      theme(axis.title.y=element_blank(),
            axis.text.y=element_blank(),
            axis.title.x=element_blank(),
            axis.ticks.y=element_blank(),
            legend.title=element_blank())

cp2<-catvar375figdflong %>%
  filter(variable=="household.income") %>%
  mutate(level=factor(level,levels=levels(dfcatna$household.income))) %>%
  ggplot(aes(x=thresh,y=value,fill=level)) +
```

```r
      geom_bar(position="fill",stat="identity") +
      scale_fill_viridis(discrete = TRUE) +
      labs(title="Household Income") +
      theme(axis.title.y=element_blank(),
            axis.text.y=element_blank(),
            axis.title.x=element_blank(),
            axis.ticks.y=element_blank(),
            legend.title=element_blank())

cp3<-catvar375figdflong %>%
  filter(variable=="race_ethnicity.factor") %>%
  mutate(level=factor(level,levels=levels(dfcatna$race_ethnicity.factor))) %>%
  ggplot(aes(x=thresh,y=value,fill=level)) +
      geom_bar(position="fill",stat="identity") +
      scale_fill_viridis(discrete = TRUE) +
      labs(title="Census Race/Ethnicity") +
      theme(axis.title.y=element_blank(),
            axis.text.y=element_blank(),
            axis.title.x=element_blank(),
            axis.ticks.y=element_blank(),
            legend.title=element_blank())

cp4<-catvar375figdflong %>%
  filter(variable=="ksads_factor") %>%
  mutate(level=as.factor(level)) %>%
  ggplot(aes(x=thresh,y=value,fill=level)) +
      geom_bar(position="fill",stat="identity") +
      scale_fill_viridis(discrete = TRUE) +
      labs(title="Trauma Count") +
      theme(axis.title.y=element_blank(),
            axis.text.y=element_blank(),
            axis.title.x=element_blank(),
            axis.ticks.y=element_blank(),
            legend.title=element_blank())


cp5<-catvar375figdflong %>%
  filter(variable=="pds_category") %>%
  mutate(level=factor(level,levels=levels(dfcatna$pds_category))) %>%
  ggplot(aes(x=thresh,y=value,fill=level)) +
      geom_bar(position="fill",stat="identity") +
      scale_fill_viridis(discrete = TRUE) +
      labs(title="Pubertal Status") +
      theme(axis.title.y=element_blank(),
```

```r
        axis.text.y=element_blank(),
        axis.title.x=element_blank(),
        axis.ticks.y=element_blank(),
        legend.title=element_blank())

cp6<-catvar375figdflong %>%
  filter(variable=="p.edu") %>%
  mutate(level=factor(level,levels=levels(dfcatna$p.edu))) %>%
  ggplot(aes(x=thresh,y=value,fill=level)) +
      geom_bar(position="fill",stat="identity") +
      scale_fill_viridis(discrete = TRUE) +
      labs(title="Parent Education") +
      theme(axis.title.y=element_blank(),
            axis.title.x=element_blank(),
            axis.text.y=element_blank(),
            axis.ticks.y=element_blank(),
            legend.title=element_blank())
```

```r
svg(filename="Manuscript/Figures/CatVarFig.svg",height=9,width=7.5,family="Times New Roman")

((cp1 / cp4 / cp3) |
  (cp2 / cp6 / cp5))
grid::grid.draw(grid::textGrob("Proportion at Threshold",x=.01,rot=90,gp = grid::gpar(fontfamily = "Times New Roman")))
grid::grid.draw(grid::textGrob("QC Threshold",y=.01,gp = grid::gpar(fontfamily = "Times New Roman")))

dev.off()
```

```
## cairo_pdf
##         2
```

1.2.3 Continuous Variables Figure (Figure 3)

```r
# we need ymin, ymax, lower, upper, and middle for each variable in each condition.
#we need ymin, ymax, lower, upper, and middle for each variable in each condition.
bpstats <- function(var,bool) {
  #message(paste0("var is ",var))
  #message(paste0("bool is ",bool))
  x<-df[df[[bool]],var]
  minvalue=min(x,na.rm=TRUE)
  maxvalue=max(x,na.rm=TRUE)
  median=median(x,na.rm = TRUE)
  low=quantile(x,.25,na.rm=TRUE)
```

```r
  high=quantile(x,.75,na.rm=TRUE)
  IQR=high-low
  mn=max(low-1.5*IQR,minvalue)
  mx=min(high+1.5*IQR,maxvalue)
  return(data.frame(lwhisk=mn,low=low,median=median,high=high,uwhisk=mx))
}

cvars <- c(
    "adi.z",
    "coi.z",
    "nihtbx_flanker_agecorrected.z",
    "nihtbx_cryst_agecorrected.z",
    "nihtbx_totalcomp_agecorrected.z",
    "pea_wiscv_tss.z",
    "pfactor.z",
    "INT.z",
    "EXT.z",
    "interview_age.z",
    "bmiAgeZ"
    )

  df$full=TRUE

ContFig375df<-expand.grid(as.factor(cvars),unique(bools)) %>%
  mutate(Var1 = as.character(Var1),
         Var2 = as.character(Var2)) %>%
  rowwise() %>%
  mutate(bpstats(Var1,Var2))

ContFig375df$measp<-factor(ContFig375df$Var1,
                           levels=cvars,
                           labels=c("Area Disadvantage",
                                    "Child Opportunity",
                                    "NIHTB Flanker",
                                    "NIHTB Crystalized",
                                    "NIHTB Total",
                                    "WISC V Matrix",
                                    "Psychopathology (p)",
                                    "Internalizing",
                                    "Externalizing",
                                    "Age",
                                    "BMI"))

ContFig375df$thresh <- factor(
```

```
    ContFig375df$Var2,
    levels = bools,
    labels = c("F",
               "T",
               "C",
               ".5",
               ".4",
               ".3",
               "R",
               ".2",
               ".1")
)

ggsave("Manuscript/Figures/ContVarFig.svg",
       ggplot(
         ContFig375df,
         aes(
           x = thresh,
           ymin = lwhisk,
           lower = low,
           middle = median,
           upper = high,
           ymax = uwhisk
         )
       ) +
         geom_boxplot(stat = "identity") +
         ylab("Z") +
         theme(axis.title.x = element_blank()) +
         facet_wrap( ~ measp, scales = "free"),
       height = 5, width = 7.5)
```

1.2.4   Trauma Count

We describe trauma count statistics in our methods:

7723, 3004, 1149

1.3   Table: Associations of behavioral data with framewise displacement (Table 2 and 3)

```
ccor<-df %>% select("adi.z",
    "coi.z",
    "nihtbx_flanker_agecorrected.z",
```

```r
      "nihtbx_cryst_agecorrected.z",
      "nihtbx_totalcomp_agecorrected.z",
      "pea_wiscv_tss.z",
      "pfactor.z",
      "INT.z",
      "EXT.z",
      "interview_age.z",
      "bmiAgeZ",
      filtered_fd_0.1:filtered_fd_0.5) %>%
  sdlabFunctions::sdcorr()

#This is the version in the manuscript.
write.csv(ccor[1:11,19:30],"Manuscript/Tables/FDcorr_continuousDVs2.csv",row.names=TRUE)

# This will write tables of bivariate linear models of FD with both continuous and discrete variables::

# Updated colnames for vars
y_names = c("filtered_fd_0.1","filtered_fd_0.2","filtered_fd_0.3","filtered_fd_0.4","filtered_fd_0.5")
x_names = c("adi","coi","nihtbx_flanker_agecorrected","nihtbx_cryst_agecorrected","nihtbx_totalcomp_agecorrected",
            "pea_wiscv_tss","pfactor","INT","EXT","interview_age","bmiAgeZ","sex","household.income","p.edu",
            "race_ethnicity.factor","ksads_factor","pds_category")

contOut = data.frame("IV"=c(),"DV"=c(),"BETA"=c(),"STD_BETA"=c(),"R_SQU"=c(),"T"=c(),"P"=c())
discOut = data.frame("IV"=c(),"DV"=c(),"F"=c(),"DF_1"=c(),"DF_2"=c(),"MODEL_P"=c(),"CONTRAST"=c(),"BETA"=c(),"SE"=c(),"T-RATIO"=c(),"P"=c())

# Iterate dependent variables (motion thresholds)
for(y_name in y_names){
  # Iterate independent variables
  for(x_name in x_names){

    # Create dataframe with only variables of interest, remove NAs
    temp_df = na.omit(data.frame(y_name=df[y_name],x_name=df[x_name]))

    # For continuous predictors:
    if( !(x_name %in% c("sex","household.income","p.edu","race_ethnicity.factor","ksads_factor","pds_category"))){

      # Scale
      temp_df[x_name] = scale(temp_df[x_name],center=TRUE,scale=TRUE)

      # Build a model
      mod = lm(data=temp_df,paste0(y_name,"~",x_name))
      modTemp = summary(mod)

      # Append statistics
```

```r
      contOut = rbind(contOut,data.frame(
        "IV"=y_name,"DV"= x_name,"BETA"=modTemp$coefficients[2],"STD_BETA"=as.numeric(QuantPsyc::lm.beta(mod)),"R_SQU"=modTemp$r.squared,
        "T"=modTemp$coefficients[6],"P"=modTemp$coefficients[8]
      ))
    }

    # For discrete predictors:
    else{

      # Build a model
      mod = lm(data=temp_df,paste0(y_name,"~",x_name))
      modTemp = summary(mod)
      #post_hoc = data.frame(emmeans::emmeans(mod,reformulate(x_name, 'pairwise'))$contrasts)[,c(1,2,3,5,6)]
      post_hoc = data.frame(emmeans::emmeans(mod,reformulate(x_name, 'trt.vs.ctrl1'))$contrasts)[,c(1,2,3,5,6)]


      # Format and append statistics
      colnames(post_hoc) = c("CONTRAST","BETA","SE","T-RATIO","P")
      mod_stats = data.frame("IV"=rep(y_name,nrow(post_hoc)),
                             "DV"=rep(x_name,nrow(post_hoc)),
                             "F"=rep(as.numeric(modTemp$fstatistic[1]),nrow(post_hoc)),
                             "DF_1"=rep(as.numeric(modTemp$fstatistic[2]),nrow(post_hoc)),
                             "DF_2"=rep(modTemp$fstatistic[3],nrow(post_hoc)),
                             "MODEL_P"=rep(pf(as.numeric(modTemp$fstatistic[1]),as.numeric(modTemp$fstatistic[2]),
                                             as.numeric(modTemp$fstatistic[3]),lower.tail = FALSE),nrow(post_hoc)))
      discOut = rbind(discOut,cbind(mod_stats,post_hoc))
    }

  }
}


sdfd <- t(df %>%
  summarize(
    sd_fd_0.1 = sd(filtered_fd_0.1, na.rm = TRUE),
    sd_fd_0.2 = sd(filtered_fd_0.2, na.rm = TRUE),
    sd_fd_0.3 = sd(filtered_fd_0.3, na.rm = TRUE),
    sd_fd_0.4 = sd(filtered_fd_0.4, na.rm = TRUE),
    sd_fd_0.5 = sd(filtered_fd_0.5, na.rm = TRUE)
  ))

discOut$fdsd<-rep(sdfd,each=sum(discOut$IV=="filtered_fd_0.1"))
discOut$sBeta<-discOut$BETA/discOut$fdsd

# Write tables to csv This continuous table did not end up included in the manuscirpt.
```

```
write.csv(contOut,"Manuscript/Tables/FDcorr_continuousDVs.csv",row.names=FALSE)
write.csv(discOut,"Manuscript/Tables/FDcorr_discreteDVs.csv",row.names=FALSE)
```

In the manuscript, F tests associated with the discrete variables tested are omitted for brevity. These are printed here:

```
discOut %>%
  filter(IV %in% c("filtered_fd_0.1","filtered_fd_0.2","filtered_fd_0.3")) %>%
  select(DV:MODEL_P) %>%
  distinct() %>%
  as_hux() %>%
  insert_column(c("Threshold",rep(c(".1mm",".2mm",".3mm"),each=6))) %>%
  set_caption("F tests for relation of discrete DVs with FD after scrubbing, by threshold (pt. 1)")
```

```
discOut %>%
  filter(IV %in% c("filtered_fd_0.4","filtered_fd_0.5")) %>%
  select(DV:MODEL_P) %>%
  distinct() %>%
  as_hux() %>%
  insert_column(c("Threshold",rep(c(".4mm",".5mm"),each=6))) %>%
  set_caption("F tests for relation of discrete DVs with FD after scrubbing, by threshold (pt. 2)")
```

1.4 Bivariate Models and Table 4

```
ORformat<-\(x) {DescTools::Format(exp(x),digits=2,ldigits=0)}
pformat<-\(x) {DescTools::Format(x,digits=3,ldigits=0,sci=30)}
```

```
varstring <-
  c(
    "sex",
    "household.income",
    "p.edu",
    "race_ethnicity.factor",
    "ksads_factor",
    "pds_category",
    "adi.z",
    "coi.z",
    "nihtbx_flanker_agecorrected.z",
    "nihtbx_cryst_agecorrected.z",
    "nihtbx_totalcomp_agecorrected.z",
```

Table 1: F tests for relation of discrete DVs with FD after scrubbing, by threshold (pt. 1)

Threshold	DV	F	DF_1	DF_2	MODEL_P
0.1mm	sex	40.6	1	9.37e+03	1.99e-10
0.1mm	household.income	20	5	8.6e+03	6.39e-20
0.1mm	p.edu	17.9	4	9.36e+03	1.3e-14
0.1mm	race_ethnicity.factor	27.5	4	9.36e+03	9.06e-23
0.1mm	ksads_factor	4.61	2	9.37e+03	0.00996
0.1mm	pds_category	2.36	4	9.29e+03	0.0512
0.2mm	sex	88	1	9.58e+03	7.92e-21
0.2mm	household.income	27.5	5	8.79e+03	9.8e-28
0.2mm	p.edu	27.7	4	9.56e+03	6.43e-23
0.2mm	race_ethnicity.factor	32.9	4	9.57e+03	2.95e-27
0.2mm	ksads_factor	6.75	2	9.57e+03	0.00118
0.2mm	pds_category	1.42	4	9.49e+03	0.224
0.3mm	sex	116	1	9.59e+03	7.92e-27
0.3mm	household.income	29.5	5	8.8e+03	8.34e-30
0.3mm	p.edu	31.7	4	9.57e+03	2.97e-26
0.3mm	race_ethnicity.factor	31.5	4	9.58e+03	4.18e-26
0.3mm	ksads_factor	6.94	2	9.59e+03	0.000977
0.3mm	pds_category	0.843	4	9.5e+03	0.498

Table 2: F tests for relation of discrete DVs with FD after scrubbing, by threshold (pt. 2)

Threshold	DV	F	DF_1	DF_2	MODEL_P
0.4mm	sex	128	1	9.59e+03	2.21e-29
0.4mm	household.income	29.5	5	8.8e+03	8.32e-30
0.4mm	p.edu	32.7	4	9.58e+03	3.97e-27
0.4mm	race_ethnicity.factor	28.8	4	9.58e+03	8.03e-24
0.4mm	ksads_factor	7.06	2	9.59e+03	0.000859
0.4mm	pds_category	1.15	4	9.5e+03	0.331
0.5mm	sex	136	1	9.59e+03	3.02e-31
0.5mm	household.income	28.8	5	8.8e+03	4.6e-29
0.5mm	p.edu	33.5	4	9.58e+03	7.92e-28
0.5mm	race_ethnicity.factor	26.5	4	9.59e+03	6.25e-22
0.5mm	ksads_factor	7	2	9.59e+03	0.000914
0.5mm	pds_category	1.12	4	9.51e+03	0.347

```
      "pea_wiscv_tss.z",
      "pfactor.z",
      "INT.z",
      "EXT.z",
      "interview_age.z",
      "bmiAgeZ"
  )

bvmodel <- \(x,newline="\n") {
  cstrings <-
    c(
      "!ABCD_rsfmri_QC1",
      "!any_rest",
      "!gframes_0.5_375",
      "!gframes_0.4_375",
      "!gframes_0.3_375",
      "!ABCD_rsfmri_QC2",
      "!gframes_0.2_375",
      "!gframes_0.1_375"
    )
  model.list<-sapply(cstrings,\(y) glm(as.formula(paste0(y," ~ ",x)),data=df,family="binomial"),simplify=FALSE)
  sapply(model.list,
         \(x) {
           lf<-length(coef(x))
           if(lf==2) {
             table<-t(c(coef=coef(x)[2:lf],
                       confint(x,2:lf,level=.95),
                       p=coef(summary(x))[2:lf,4]))
           } else {
             table<-cbind(coef=coef(x)[2:lf],
                       confint(x,2:lf,level=.95),
                       p=coef(summary(x))[2:lf,4])
           }

           #Bonferonni correction
           if(names(coef(x))[2] %in% c("adi.z",
                                       "coi.z",
                                       "nihtbx_totalcomp_agecorrected.z",
                                       "pea_wiscv_tss.z")) {
             table[,4]<-table[,4]*2
           }
           m<-paste0(ORformat(table[,1]),
                               sdlabFunctions::starPs(table[,4]),
                               newline,
```

```
                                    ORformat(table[,2]),
                                    "-",
                                    ORformat(table[,3]),
                                    newline,
                                    pformat(table[,4]))
            names(m)<-names(coef(x)[2:lf])
            sub(".000","<.001",m)
    },
    simplify=TRUE)
}

bvtable<-reduce(sapply(varstring,bvmodel),rbind)

alllevels<-c(
  "Sex (Male)",
      "  $0-$25k",
      "  $25-$50k",
      "  $50-$75k",
      "  $75-$100k",
      "  >$200k",
      "  <High School",
      "  HS Grad.",
      "  Some College",
      "  Graduate",
      "  Black",
      "  Hispanic",
      "  Asian",
      "  Other",
      "  1 Trauma",
      "  >=2 Trauma",
      "  Early Puberty",
      "  Mid Puberty",
      "  Late Puberty",
      "  Post Pubertal",
      "Area Disadvantage",
      "Child Opportunity",
      "NIHTB Flanker",
      "NIHTB Crystalized",
      "NIHTB Total",
      "WISC V Matrix",
      "Psychopathology",
      "Internalizing",
      "Externalizing",
      "Age",
```

```r
    "BMI"
)


#1 column per level
bvhux <- as_hux(bvtable) %>%
  insert_column(alllevels) %>%
  insert_row(
    "Household Income (ref: $100-$200k)",
    after = 1,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Highest Parental Education (ref: College Degree)",
    after = 7,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Census Race/Ethnicity (ref: White)",
    after = 12,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "KSADS Trauma Count (ref: 0 Exposures)",
    after = 17,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Pubertal Status (ref: pre-pubertal)",
    after = 20,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(c("Variable", rep(c("OR\n[90% CI]\np"), 8)), after = 0) %>%
  insert_row(c("", "T", "C", ".5", ".4", ".3", "R", ".2", ".1"))

tables.xlsx<-as_Workbook(bvhux[1:20,],
                   Workbook=tables.xlsx,
                   sheet="bvmPage1")
tables.xlsx<-as_Workbook(bvhux[c(1:2,21:nrow(bvhux)),],
```

```
                Workbook=tables.xlsx,
                sheet="bvmPage2")
```

1.5 Condition (Adjusted) Models

```
#terms 22 and 23 (neighborhood factors), as well as term 26 and 27 (general cognition measures) need to be corrected.
pcorrect<-\(x) {
  x[22:23]<-x[22:23]*2
  x[26:27]<-x[26:27]*2
  x
}

adjmodelrightterms<-"~ sex + household.income + p.edu + race_ethnicity.factor +
ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z +
nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ"
model.QC1 <-
  glm(as.formula(paste("!ABCD_rsfmri_QC1 ", adjmodelrightterms)),
      data = df, family = "binomial")
model.QC1.results <-
  cbind(confint(model.QC1, level = .95), model.QC1$coefficients)[, c(1, 3, 2)]
model.QC1.results <-
  cbind(model.QC1.results, pcorrect(coef(summary(model.QC1))[, 4]))

model.ABCC <-
  glm(as.formula(paste("!any_rest ", adjmodelrightterms)), data = df, family =
        "binomial")
model.ABCC.results <-
  cbind(confint(model.ABCC, level = .95), model.ABCC$coefficients)[, c(1, 3, 2)]
model.ABCC.results <-
  cbind(model.ABCC.results, pcorrect(coef(summary(model.ABCC))[, 4]))

model.5mm <-
  glm(as.formula(paste("!gframes_0.5_375 ", adjmodelrightterms)), data = df, family =
        "binomial")
model.5mm.results <-
  cbind(confint(model.5mm, level = .95), model.5mm$coefficients)[, c(1, 3, 2)]
model.5mm.results <-
  cbind(model.5mm.results, pcorrect(coef(summary(model.5mm))[, 4]))

model.4mm <-
  glm(as.formula(paste("!gframes_0.4_375 ", adjmodelrightterms)), data = df, family =
```

```
      "binomial")
model.4mm.results <-
  cbind(confint(model.4mm, level = .95), model.4mm$coefficients)[, c(1, 3, 2)]
model.4mm.results <-
  cbind(model.4mm.results, pcorrect(coef(summary(model.4mm))[, 4]))

model.3mm <-
  glm(as.formula(paste("!gframes_0.3_375 ", adjmodelrightterms)), data = df, family =
      "binomial")
model.3mm.results <-
  cbind(confint(model.3mm, level = .95), model.3mm$coefficients)[, c(1, 3, 2)]
model.3mm.results <-
  cbind(model.3mm.results, pcorrect(coef(summary(model.3mm))[, 4]))

model.QC2 <-
  glm(as.formula(paste("!ABCD_rsfmri_QC2 ", adjmodelrightterms)), data = df, family =
      "binomial")
model.QC2.results <-
  cbind(confint(model.QC2, level = .95), model.QC2$coefficients)[, c(1, 3, 2)]
model.QC2.results <-
  cbind(model.QC2.results, pcorrect(coef(summary(model.QC2))[, 4]))

model.2mm <-
  glm(as.formula(paste("!gframes_0.2_375 ", adjmodelrightterms)), data = df, family =
      "binomial")
model.2mm.results <-
  cbind(confint(model.2mm, level = .95), model.2mm$coefficients)[, c(1, 3, 2)]
model.2mm.results <-
  cbind(model.2mm.results, pcorrect(coef(summary(model.2mm))[, 4]))


model.1mm <-
  glm(as.formula(paste("!gframes_0.1_375 ", adjmodelrightterms)), data = df, family =
      "binomial")
model.1mm.results <-
  cbind(confint(model.1mm, level = .95), model.1mm$coefficients)[, c(1, 3, 2)]
model.1mm.results <-
  cbind(model.1mm.results, pcorrect(coef(summary(model.1mm))[, 4]))
```

1.5.1 Adjusted Model Tables (Table 5)

```
Mformat<-\(m) {
  stars<-sdlabFunctions::starPs(m[,4])
```

```r
  paste0(ORformat(m[,2]),
       stars,
       "\n",ORformat(m[,1]),"-",ORformat(m[,3]),"",
       "\n",DescTools::Format(m[,4],digits=3,ldigits=0,sci=30))
}

h1modelstable<-cbind(Mformat(model.QC1.results),
                  Mformat(model.ABCC.results),
                  Mformat(model.5mm.results),
                  Mformat(model.4mm.results),
                  Mformat(model.3mm.results),
                  Mformat(model.QC2.results),
                  Mformat(model.2mm.results),
                  Mformat(model.1mm.results))

h1modelstable[h1modelstable==".000"]<-"<.001"

alllevels<-c("Intercept",
     "Sex (Male)",
     "  $0-$25k",
     "  $25-$50k",
     "  $50-$75k",
     "  $75-$100k",
     "  >$200k",
     "  <High School",
     "  HS Grad.",
     "  Some College",
     "  Graduate",
     "  Black",
     "  Hispanic",
     "  Asian",
     "  Other",
     "  1 Trauma",
     "  >=2 Trauma",
     "  Early Puberty",
     "  Mid Puberty",
     "  Late Puberty",
     "  Post Pubertal",
     "Area Disadvantage",
     "Child Opportunity",
     "NIHTB Flanker",
     "NIHTB Crystalized",
     "NIHTB Total",
     "WISC V Matrix",
```

```r
      "Psychopathology",
      "Internalizing",
      "Externalizing",
      "Age",
      "BMI"
      )

#1 column per level
h1hux <- as_hux(h1modelstable) %>%
  insert_column(alllevels) %>%
  insert_row(
    "Household Income (ref: $100-$200k)",
    after = 2,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Highest Parental Education (ref: College Degree)",
    after = 8,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Census Race/Ethnicity (ref: White)",
    after = 13,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "KSADS Trauma Count (ref: 0 Exposures)",
    after = 18,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Pubertal Status (ref: pre-pubertal)",
    after = 21,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(c("Variable", rep(c("OR\n[90% CI]\np"), 8)), after = 0) %>%
  insert_row(c("", "T", "C", ".5", ".4", ".3", "R", ".2", ".1"))

colnames(h1hux)[1]<-"h1"
```

```r
tables.xlsx<-as_Workbook(h1hux[1:20,],
                         Workbook=tables.xlsx,
                         sheet="AdjustedModels1")
tables.xlsx<-as_Workbook(h1hux[c(1:2,21:nrow(h1hux)),],
                         Workbook=tables.xlsx,
                         sheet="AdjustedModels2")
```

1.5.2 Site and MRI Effects – sensitivity analysis.

These are run only in some conditions: ABCC, QC2, 2mm

```r
model.ABCC.scanner<-glm(
  !any_rest ~ sex + household.income + p.edu + race_ethnicity.factor +
    ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
    + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
    pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
    mri_model,
  data=df,
  family="binomial")
model.ABCC.scanner.results<-cbind(confint(model.ABCC.scanner,level=.95),model.ABCC.scanner$coefficients)[,c(1,3,2)]
model.ABCC.scanner.results<-cbind(model.ABCC.scanner.results,p.adjust(coef(summary(model.ABCC.scanner))[,4],method="fdr"))

model.ABCC.scanner.site<-glm(
  !any_rest ~ sex + household.income + p.edu + race_ethnicity.factor +
    ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
    + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
    pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
    mri_model + site_id_l,
  data=df,
  family="binomial")
model.ABCC.scanner.site.results<-cbind(confint(model.ABCC.scanner.site,level=.95),model.ABCC.scanner.site$coefficients)[,c(1,3,2)]
model.ABCC.scanner.site.results<-cbind(model.ABCC.scanner.site.results,p.adjust(coef(summary(model.ABCC.scanner.site))[,4],method="fdr"))

model.2mm.scanner<-glm(
  !gframes_0.2_375 ~ sex + household.income + p.edu + race_ethnicity.factor +
    ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
    + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
    pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
    mri_model,
  data=df,
  family="binomial")
model.2mm.scanner.results<-cbind(confint(model.2mm.scanner,level=.95),model.2mm.scanner$coefficients)[,c(1,3,2)]
model.2mm.scanner.results<-cbind(model.2mm.scanner.results,p.adjust(coef(summary(model.2mm.scanner))[,4],method="fdr"))
```

```
model.2mm.scanner.site<-glm(
   !gframes_0.2_375 ~ sex + household.income + p.edu + race_ethnicity.factor +
     ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
     + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
     pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
     mri_model + site_id_l,
   data=df,
   family="binomial")
model.2mm.scanner.site.results<-cbind(confint(model.2mm.scanner.site,level=.95),model.2mm.scanner.site$coefficients)[,c(1,3,2)]
model.2mm.scanner.site.results<-cbind(model.2mm.scanner.site.results,p.adjust(coef(summary(model.2mm.scanner.site))[,4],method="fdr"))

model.QC2.scanner<-glm(
   !ABCD_rsfmri_QC2 ~ sex + household.income + p.edu + race_ethnicity.factor +
     ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
     + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
     pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
     mri_model,
   data=df,
   family="binomial")
model.QC2.scanner.results<-cbind(confint(model.QC2.scanner,level=.95),model.QC2.scanner$coefficients)[,c(1,3,2)]
model.QC2.scanner.results<-cbind(model.QC2.scanner.results,p.adjust(coef(summary(model.QC2.scanner))[,4],method="fdr"))

model.QC2.scanner.site<-glm(
   !ABCD_rsfmri_QC2 ~ sex + household.income + p.edu + race_ethnicity.factor +
     ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
     + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
     pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
     mri_model + site_id_l,
   data=df,
   family="binomial")
model.QC2.scanner.site.results<-cbind(confint(model.QC2.scanner.site,level=.95),model.QC2.scanner.site$coefficients)[,c(1,3,2)]
model.QC2.scanner.site.results<-cbind(model.QC2.scanner.site.results,p.adjust(coef(summary(model.QC2.scanner.site))[,4],method="fdr"))

model.ABCC.site<-glm(
   !any_rest ~ sex + household.income + p.edu + race_ethnicity.factor +
     ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
     + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
     pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
     site_id_l,
   data=df,
   family="binomial")
model.ABCC.site.results<-cbind(confint(model.ABCC.site,level=.95),model.ABCC.site$coefficients)[,c(1,3,2)]
model.ABCC.site.results<-cbind(model.ABCC.site.results,p.adjust(coef(summary(model.ABCC.site))[,4],method="fdr"))
```

```
model.2mm.site<-glm(
  !gframes_0.2_375 ~ sex + household.income + p.edu + race_ethnicity.factor +
    ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
    + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
    pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
    site_id_l,
  data=df,
  family="binomial")
model.2mm.site.results<-cbind(confint(model.2mm.site,level=.95),model.2mm.site$coefficients)[,c(1,3,2)]
model.2mm.site.results<-cbind(model.2mm.site.results,p.adjust(coef(summary(model.2mm.site))[,4],method="fdr"))

model.QC2.site<-glm(
  !ABCD_rsfmri_QC2 ~ sex + household.income + p.edu + race_ethnicity.factor +
    ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected.z
    + nihtbx_cryst_agecorrected.z + nihtbx_totalcomp_agecorrected.z +
    pea_wiscv_tss.z + pfactor.z + INT.z + EXT.z + interview_age.z + bmiAgeZ +
    site_id_l,
  data=df,
  family="binomial")
model.QC2.site.results<-cbind(confint(model.QC2.site,level=.95),model.QC2.site$coefficients)[,c(1,3,2)]
model.QC2.site.results<-cbind(model.QC2.site.results,p.adjust(coef(summary(model.QC2.site))[,4],method="fdr"))
```

1.5.2.1 Supplementary Tables This is a table of the model output for just the scanner terms, with and without site control, which we reference in the text:

```
MformatCI <- \(m) {
  stars <- sdlabFunctions::starPs(m[, 4])
  cbind(
    paste0(
      ORformat(m[, 2]),
      stars,
      " [",
      ORformat(m[, 1]),
      "-",
      ORformat(m[, 3]),
      "]"))
}

sitecoefs<-rbind(cbind(MformatCI(model.ABCC.scanner.results)[33:36,],
      MformatCI(model.QC2.scanner.results)[33:36,],
      MformatCI(model.2mm.scanner.results)[33:36,]),
      cbind(MformatCI(model.ABCC.scanner.site.results)[33:36,],
      MformatCI(model.QC2.scanner.site.results)[33:36,],
      MformatCI(model.2mm.scanner.site.results)[33:36,]))
```

```r
as_huxtable(sitecoefs) %>%
  insert_row(c("ABCC", "QC2", ".2mm")) %>%
  insert_row(c("Without Site Control", "", "")) %>%
  merge_cells(1, 1:3) %>%
  merge_cells(2, 1:3) %>%
  insert_row(c("With Site Control","",""),after = 6) %>%
  merge_cells(7,1:3) %>%
  insert_column(
    c(
      "",
      "Scanner (ref: Philips Achieva)",
      "GE Discovery MR750",
      "Philips Ingenia",
      "Siemens Prisma",
      "Siemens Prisma fit",
      "",
      "GE Discovery MR750",
      "Philips Ingenia",
      "Siemens Prisma",
      "Siemens Prisma fit"
    )
  ) %>%
  set_caption("OR of scanner terms, with and without site control [95\\% CI]. ***: p<.001") %>%
  set_wrap(TRUE) %>%
  set_width(1)
```

1.5.2.2 Supplementary Forest Plots

These forest plots show ORs with a significant (p<.05) model term in either the site-controlled or the site un-controlled adjusted model. They are here to visualize the effects of site control on the adjusted models.

```r
ABCC.compdf <- rbind(
  model.ABCC.site.results %>%
    as.data.frame() %>%
    rownames_to_column("term") %>%
    rename_with( ~ c("term", "lowCI", "OR", "highCI", "p")) %>%
    mutate(
      lowCI = replace_na(lowCI, min(lowCI, na.rm = TRUE)),
      highCI = replace_na(highCI, max(highCI, na.rm = TRUE)),
      sig = p < .05,
      model = "Site Control"
    ) %>%
    mutate(across(lowCI:highCI, exp)),
  model.ABCC.results %>%
    as.data.frame() %>%
```

Table 3: OR of scanner terms, with and without site control [95% CI]. ***: p<.001

Scanner (ref: Philips Achieva)	Without Site Control ABCC		
GE Discovery MR750	.86 [.72–1.04]	.56*** [.46–.69]	.70*** [.59–.84]
Philips Ingenia	.20*** [.13–.31]	.87 [.63–1.19]	.42*** [.30–.57]
Siemens Prisma	.15*** [.12–.20]	.27*** [.22–.33]	.20*** [.17–.25]
Siemens Prisma fit	.39*** [.32–.47]	.43*** [.36–.53]	.43*** [.36–.51]
With Site Control			
GE Discovery MR750	3894.69 [1.69e-14–NA]	44437.50 [.00–NA]	13726.89 [.00–NA]
Philips Ingenia	8698.06 [4.91e-09–NA]	7.24e+09 [5.44e-11–NA]	9.30e+09 [.01–NA]
Siemens Prisma	3109.47 [1.16e-14–NA]	8613.49 [.00–NA]	15297.36 [.00–NA]
Siemens Prisma fit	6030.51 [2.24e-14–NA]	8257.41 [.00–NA]	16617.52 [.00–NA]

```r
    rownames_to_column("term") %>%
    rename_with( ~ c("term", "lowCI", "OR", "highCI", "p")) %>%
    mutate(
      lowCI = replace_na(lowCI, min(lowCI, na.rm = TRUE)),
      highCI = replace_na(highCI, max(highCI, na.rm = TRUE)),
      sig = p < .05,
      model = "No Control"
    ) %>%
    mutate(across(lowCI:highCI, exp))
) %>%
  mutate(model = as.factor(model),
         alphaval=case_when(sig ~ 1,
                            !sig ~ .5))


#terms where one model coefficient was sig
sigterms <-
  (ABCC.compdf %>% group_by(term) %>% summarise(anysig = sum(sig) != 0) %>% filter(anysig))$term

ABCC.compdf<-ABCC.compdf %>% mutate(alphaval=factor(alphaval,levels=c(.5,1),labels=c("p>.05","p<.05")))
```

```r
test<-ggplot(ABCC.compdf %>% filter(term %in% sigterms),
       aes(x = term, color = model, alpha = alphaval)) +
  geom_point(aes(y = OR), position = position_dodge(.2)) +
  geom_linerange(aes(ymin = lowCI, ymax = highCI, x = term),
                 position = position_dodge(.2)) +
  ylim(-4, 4) +
  geom_hline(lty = 2, aes(yintercept = 1), colour = 'black') +
  coord_flip() + theme(legend.title = element_blank())


QC2.compdf <- rbind(
  model.QC2.site.results %>%
    as.data.frame() %>%
    rownames_to_column("term") %>%
    rename_with( ~ c("term", "lowCI", "OR", "highCI", "p")) %>%
    mutate(
      lowCI = replace_na(lowCI, min(lowCI, na.rm = TRUE)),
      highCI = replace_na(highCI, max(highCI, na.rm = TRUE)),
      sig = p < .05,
      model = "Site Control"
    ) %>%
    mutate(across(lowCI:highCI, exp)),
  model.QC2.results %>%
    as.data.frame() %>%
    rownames_to_column("term") %>%
    rename_with( ~ c("term", "lowCI", "OR", "highCI", "p")) %>%
    mutate(
      lowCI = replace_na(lowCI, min(lowCI, na.rm = TRUE)),
      highCI = replace_na(highCI, max(highCI, na.rm = TRUE)),
      sig = p < .05,
      model = "No Control"
    ) %>%
    mutate(across(lowCI:highCI, exp))
) %>%
  mutate(model = as.factor(model),
         alphaval=case_when(sig ~ 1,
                            !sig ~ .5))

#terms where one model coefficient was sig
sigterms<-(QC2.compdf %>% group_by(term) %>% summarise(anysig=sum(sig)!=0) %>% filter(anysig))$term

QC2.compdf<-QC2.compdf %>% mutate(alphaval=factor(alphaval,levels=c(.5,1),labels=c("p>.05","p<.05")))

ggplot(QC2.compdf %>% filter(term %in% sigterms),
       aes(x = term, color = model, alpha = alphaval)) +
```

```
geom_point(aes(y = OR), position = position_dodge(.2)) +
geom_linerange(aes(ymin = lowCI, ymax = highCI, x = term),
               position = position_dodge(.2)) +
ylim(-4, 4) +
geom_hline(lty = 2, aes(yintercept = 1), colour = 'black') +
coord_flip() + theme(legend.title = element_blank())
```
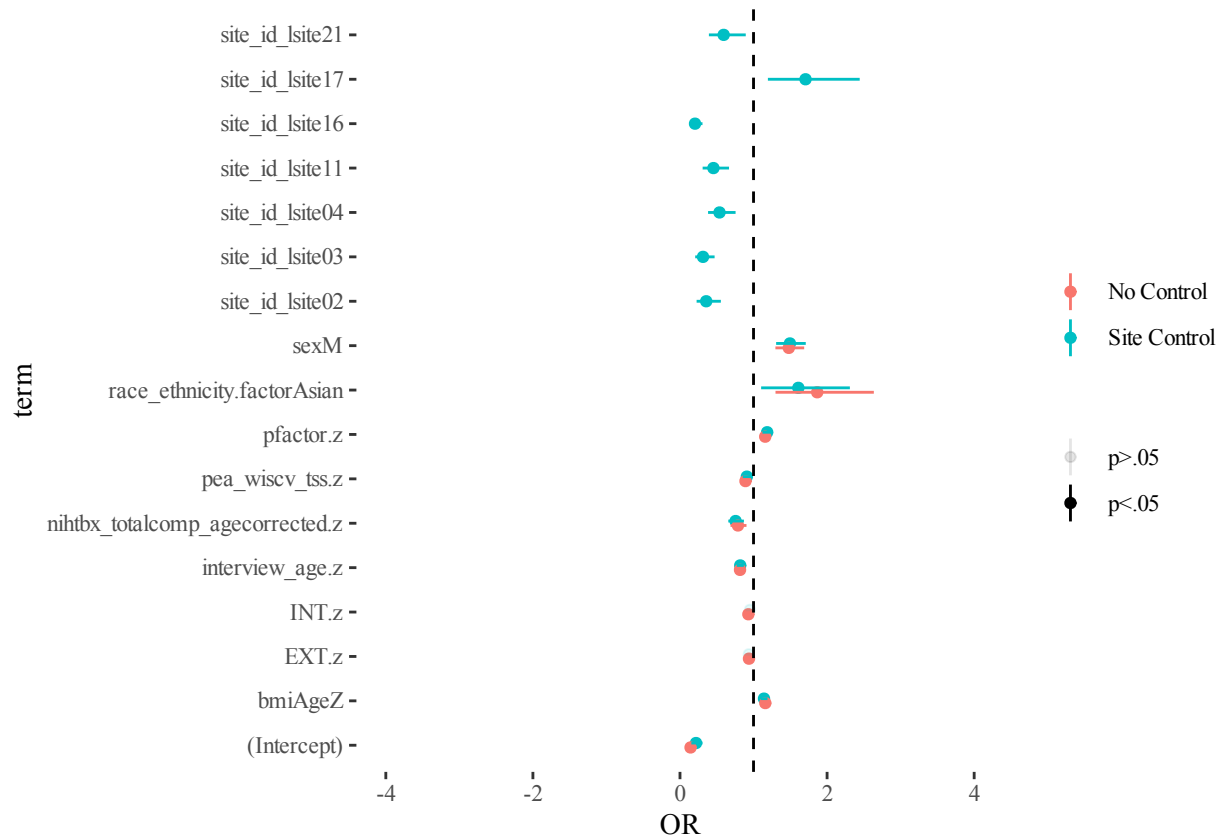

Figure 1: Significant ORs predicting DAIRC Recommended missingness with and without site control

```
twomm.compdf <- rbind(
  model.2mm.site.results %>%
    as.data.frame() %>%
    rownames_to_column("term") %>%
    rename_with( ~ c("term", "lowCI", "OR", "highCI", "p")) %>%
    mutate(
```

```r
      lowCI = replace_na(lowCI, min(lowCI, na.rm = TRUE)),
      highCI = replace_na(highCI, max(highCI, na.rm = TRUE)),
      sig = p < .05,
      model = "Site Control"
    ) %>%
    mutate(across(lowCI:highCI, exp)),
  model.2mm.results %>%
    as.data.frame() %>%
    rownames_to_column("term") %>%
    rename_with( ~ c("term", "lowCI", "OR", "highCI", "p")) %>%
    mutate(
      lowCI = replace_na(lowCI, min(lowCI, na.rm = TRUE)),
      highCI = replace_na(highCI, max(highCI, na.rm = TRUE)),
      sig = p < .05,
      model = "No Control"
    ) %>%
    mutate(across(lowCI:highCI, exp))
) %>%
  mutate(model = as.factor(model),
         alphaval=case_when(sig ~ 1,
                            !sig ~ .5))


#terms where one model coefficient was sig
sigterms <-
  (twomm.compdf %>% group_by(term) %>% summarise(anysig = sum(sig) != 0) %>% filter(anysig))$term

twomm.compdf<-twomm.compdf %>% mutate(alphaval=factor(alphaval,levels=c(.5,1),labels=c("p>.05","p<.05")))

ggplot(twomm.compdf %>% filter(term %in% sigterms),
       aes(x = term, color = model, alpha = alphaval)) +
  geom_point(aes(y = OR), position = position_dodge(.2)) +
  geom_linerange(aes(ymin = lowCI, ymax = highCI, x = term),
                 position = position_dodge(.2)) +
  ylim(-4, 4) +
  geom_hline(lty = 2, aes(yintercept = 1), colour = 'black') +
  coord_flip() + theme(legend.title = element_blank())
```

```r
sitesummary <- df %>% group_by(site_id_l) %>%
  summarise(
    "ExcludedPerc.QC2" = mean(!ABCD_rsfmri_QC2),
```
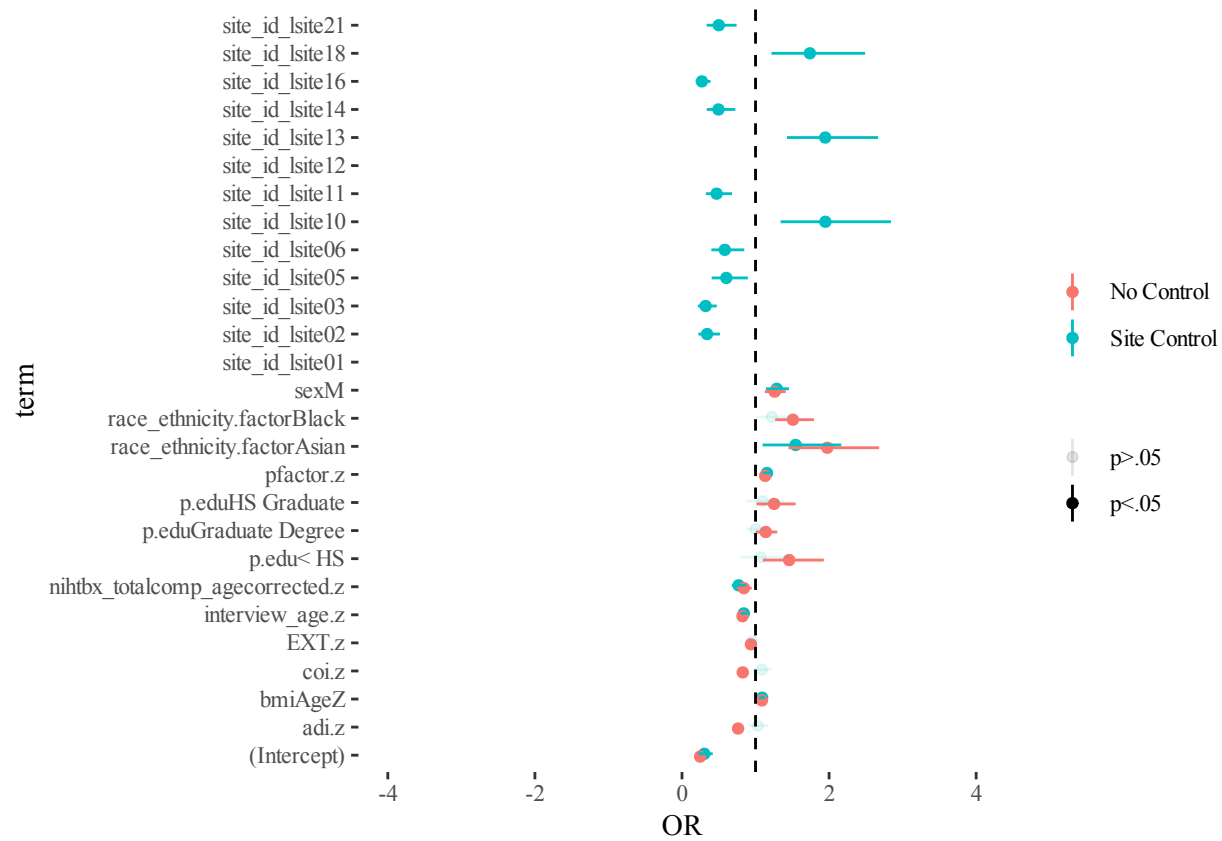
Figure 2: Significant ORs predicting .2mm missingness with and without site control

```r
    "ExcludedPerc.ABCC" = mean(!any_rest),
    "ExcludedPerc.2mm" = mean(!gframes_0.2_375),
    "ExcludedPerc.fastqc" = mean(!fastqcok)
  ) %>%
  arrange(site_id_l) %>%
  mutate(
    sitelabel = recode_factor(
      site_id_l,
      `site01` = "Los Angeles, CA (CHLA)",
      `site02` = "Boulder, CO",
      `site03` = "Miami, FL",
      `site04` = "Tulsa, OK",
      `site05` = "Columbia, SC",
      `site06` = "Portland, OR",
      `site07` = "Rochester, NY",
      `site08` = "Menlo Park, CA",
      `site09` = "Los Angeles, CA (UCLA)",
      `site10` = "San Diego, CA",
      `site11` = "Gainsville, FL",
      `site12` = "Baltimore, MD",
      `site13` = "Ann Arbor, MI",
      `site14` = "Minneapolis, MN",
      `site15` = "Pittsburgh, Pa",
      `site16` = "Salt Lake City, UT",
      `site17` = "Burlington, VT",
      `site18` = "Milwaukee, WI",
      `site19` = "Richmond, VA",
      `site20` = "St. Louis, MO",
      `site21` = "New Haven, CT",
      `site22` = "New York, NY"
    )
  )

sitesummary2 <- sitesummary %>%
  mutate(sitelabel = fct_reorder(sitelabel, ExcludedPerc.QC2)) %>%
  pivot_longer(
    cols = starts_with("ExcludedPerc"),
    names_to = "sample",
    names_prefix = "ExcludedPerc\\.",
    values_to = "Percent_Missing"
  ) %>%
  mutate(sample = factor(
    sample,
    levels = c("fastqc",
```

```
                "QC2",
                "ABCC",
                "2mm"),
    labels = c("Current FastQC",
                "DAIC inclusion",
                "ABCC",
                "ABCC <.2mm FD")
))

sensplot1<-ggplot(sitesummary2,
       aes(x=sitelabel,y=Percent_Missing,fill=sample)) +
  geom_col(position="dodge") +
  theme(axis.title.x=element_blank(),
        legend.title= element_blank()) +
  ylab("% Excluded") +
  scale_x_discrete(guide = guide_axis(angle = 45)) +
  scale_fill_viridis(discrete=TRUE) +
  ggtitle("% Missing by site, inclusion criteria")
```

1.5.2.3 Figure 4: Missingness by site, inclusion criteria

We reference the sites with the highest % exclusion, by site, in the text – they are given by this table.

```
sitesummary2 %>%
  group_by(sample) %>%
  filter(Percent_Missing == max(Percent_Missing)) %>%
  mutate(Percent_Missing = Percent_Missing*100) %>%
  select(sitelabel:Percent_Missing) %>%
  rename("Site"="sitelabel",
         "Condition"="sample",
         "% Excluded"="Percent_Missing") %>%
  as_hux() %>%
  set_caption("Sites with Highest \\% Excluded, by Condition")
```

```
sitesummary3 <- df %>% group_by(site_id_l,race_ethnicity.factor) %>%
  summarise(
    "ExcludedN" = sum(!any_rest),
    "TotalN" = n()) %>%
  mutate(
    sitelabel = recode_factor(
      site_id_l,
      `site01` = "Los Angeles, CA (CHLA)",
      `site02` = "Boulder, CO",
      `site03` = "Miami, FL",
```

Table 4: Sites with Highest % Excluded, by Condition

Site	Condition	% Excluded
Pittsburgh, Pa	DAIC inclusion	32.5
Pittsburgh, Pa	Current FastQC	30.3
Los Angeles, CA (CHLA)	ABCC <.2mm FD	75.6
Baltimore, MD	ABCC	72.7

```r
      `site04` = "Tulsa, OK",
      `site05` = "Columbia, SC",
      `site06` = "Portland, OR",
      `site07` = "Rochester, NY",
      `site08` = "Menlo Park, CA",
      `site09` = "Los Angeles, CA (UCLA)",
      `site10` = "San Diego, CA",
      `site11` = "Gainsville, FL",
      `site12` = "Baltimore, MD",
      `site13` = "Ann Arbor, MI",
      `site14` = "Minneapolis, MN",
      `site15` = "Pittsburgh, Pa",
      `site16` = "Salt Lake City, UT",
      `site17` = "Burlington, VT",
      `site18` = "Milwaukee, WI",
      `site19` = "Richmond, VA",
      `site20` = "St. Louis, MO",
      `site21` = "New Haven, CT",
      `site22` = "New York, NY"
  )) %>%
 mutate(sitelabel = factor(sitelabel,levels=levels(sitesummary2$sitelabel)))

sensplot2<-ggplot(sitesummary3,
      aes(x=sitelabel,y=ExcludedN,fill=race_ethnicity.factor)) +
 geom_col(position="stack") +
 theme(axis.title.x=element_blank(),
      legend.title=element_blank()) +
 ylab("N Excluded") +
 scale_x_discrete(guide = guide_axis(angle = 45)) +
 scale_fill_viridis(discrete=TRUE) +
 ggtitle("N Missing in ABCC by site, race/ethnicity")
```

```r
sensplot3<-ggplot(sitesummary3,
        aes(x=sitelabel,y=TotalN,fill=race_ethnicity.factor)) +
  geom_col(position="stack") +
  theme(axis.title.x=element_blank(),
        legend.title=element_blank()) +
  ylab("N") +
  scale_x_discrete(guide = guide_axis(angle = 45)) +
  scale_fill_viridis(discrete=TRUE) +
  ggtitle("N in ABCC by site, race/ethnicity")
```

```r
ggsave("Manuscript/Figures/sensplot.svg",(sensplot1 / sensplot2 / sensplot3),height=9,width=7.5)
```

1.6 H2 Model

This model specifically looks at changes associated with motion threshold choice in the ABCC data. In the pre-registration, we originally proposed a mixed effects logistic model where each participant had a case for each motion threshold, and the percentage of data missing from each threshold was used as a within-subject variable. In simulation, we discovered a number of problems with this approach:

1. The proposed mixed effects logistic regression created very poor predictions and model fits in simulation, and often did not converge. A linear model of probability proved more stable and accurate and we determined to use that approach (a logistic model is also presented below, but it is similarly flawed).
2. We proposed that an interaction between a variable and the 'pmiss' variable would indicate that bias was worsening as more data was excluded. However, in simulating this approach we discovered that in the case that there was a bias but it did not worsen as more data was excluded, as, for example, if 2 males were excluded for every female, this would surface as a linear interaction between pmiss and sex in the model. Instead, a worsening bias would be indicated by an interaction between a variable and a polynomial of pmiss, which would indicate a curvilinear trajectory.

(notes from the simulation are available on osf)

An important remaining question from the simulation study was whether to include an intercept and main effects in the model. As, at pmiss=0, the probability of exclusion ought to be 0 for everyone, one could argue the intercept should be omitted. We will try both methods and compare model fit.

```r
#Wide > Long using all the thresholds.
pmisssum <- df %>% summarise(across(c(any_rest,
                                      ends_with("_375")),
                                    ~mean(!.))) %>%
  unlist()

dflong <- df %>%
  pivot_longer(cols=c(any_rest,gframes_0.01_375:gframes_0.5_375),
                         names_to="threshold",
                         values_to="missing") %>%
  mutate(missing=!missing)

dflong$pmiss<-pmisssum[dflong$threshold]
```

```r
# h2 model with intercepts.
m.h2 <- lmer(
  missing ~ poly(pmiss, 2) +
    sex + sex:poly(pmiss, 2) +
    household.income + household.income:poly(pmiss, 2) +
    p.edu + p.edu:poly(pmiss, 2) +
    race_ethnicity.factor + race_ethnicity.factor:poly(pmiss, 2) +
    ksads_factor + ksads_factor:poly(pmiss, 2) +
    pds_category + pds_category:poly(pmiss, 2) +
    adi.z + adi.z:poly(pmiss, 2) +
    coi.z + coi.z:poly(pmiss, 2) +
    nihtbx_flanker_agecorrected.z + nihtbx_flanker_agecorrected.z:poly(pmiss, 2) +
    nihtbx_cryst_agecorrected.z + nihtbx_cryst_agecorrected.z:poly(pmiss, 2) +
    nihtbx_totalcomp_agecorrected.z + nihtbx_totalcomp_agecorrected.z:poly(pmiss, 2) +
    pea_wiscv_tss.z + pea_wiscv_tss.z:poly(pmiss, 2) +
    pfactor.z + pfactor.z:poly(pmiss, 2) +
    INT.z + INT.z:poly(pmiss, 2) +
    EXT.z + EXT.z:poly(pmiss, 2) +
    interview_age.z + interview_age.z:poly(pmiss, 2) +
    bmiAgeZ + bmiAgeZ:poly(pmiss, 2) +
    (pmiss | subjectkey),
  data = dflong
)

# h2 model with intercepts removed
m.h2.stripped <- lmer(
  missing ~ pmiss + poly(pmiss, 2) +
    sex:poly(pmiss,2) +
    household.income:poly(pmiss,2) +
    p.edu:poly(pmiss,2) +
    race_ethnicity.factor:poly(pmiss,2) +
    ksads_factor:poly(pmiss,2) +
    pds_category:poly(pmiss,2) +
    adi.z:poly(pmiss,2) +
    coi.z:poly(pmiss,2) +
    nihtbx_flanker_agecorrected.z:poly(pmiss,2) +
    nihtbx_cryst_agecorrected.z:poly(pmiss,2) +
    nihtbx_totalcomp_agecorrected.z:poly(pmiss,2) +
    pea_wiscv_tss.z:poly(pmiss,2) +
    pfactor.z:poly(pmiss,2) +
    INT.z:poly(pmiss,2) +
    EXT.z:poly(pmiss,2) +
    interview_age.z:poly(pmiss,2) +
    bmiAgeZ:poly(pmiss,2) +
```

```
     + 0 +
     (pmiss + 0 | subjectkey),
   data = dflong
)

# h2 model as a glm - this does not converge
# m.h2.glm<- glmer(
#   missing ~ poly(pmiss, 2) +
#     sex + sex:poly(pmiss, 2) +
#     household.income + household.income:poly(pmiss, 2) +
#     p.edu + p.edu:poly(pmiss, 2) +
#     race_ethnicity.factor + race_ethnicity.factor:poly(pmiss, 2) +
#     ksads_factor + ksads_factor:poly(pmiss, 2) +
#     pds_category + pds_category:poly(pmiss, 2) +
#     adi.z + adi.z:poly(pmiss, 2) +
#     coi.z + coi.z:poly(pmiss, 2) +
#     nihtbx_flanker_agecorrected.z + nihtbx_flanker_agecorrected.z:poly(pmiss, 2) +
#     nihtbx_cryst_agecorrected.z + nihtbx_cryst_agecorrected.z:poly(pmiss, 2) +
#     nihtbx_totalcomp_agecorrected.z + nihtbx_totalcomp_agecorrected.z:poly(pmiss, 2) +
#     pea_wiscv_tss.z + pea_wiscv_tss.z:poly(pmiss, 2) +
#     pfactor.z + pfactor.z:poly(pmiss, 2) +
#     INT.z + INT.z:poly(pmiss, 2) +
#     EXT.z + EXT.z:poly(pmiss, 2) +
#     interview_age.z + interview_age.z:poly(pmiss, 2) +
#     bmiAgeZ + bmiAgeZ:poly(pmiss, 2) +
#     (pmiss | subjectkey),
#   family="binomial",
#   data = dflong,
#   nAGQ=0,
#   control=glmerControl(optimizer="nloptwrap")
# )

performance::print_md(performance::compare_performance(m.h2,m.h2.stripped))
```

Table 5: Comparison of Model Performance Indices

Name	Model	AIC (weights)	AICc (weights)	BIC (weights)	R2 (cond.)	R2 (marg.)	ICC	RMSE	Sigma
m.h2	lmerModLmerTest	-1.8e+05 (>.999)	-1.8e+05 (>.999)	-1.8e+05 (>.999)	0.91	0.19	0.89	0.18	0.19
m.h2.stripped	lmerModLmerTest	3.5e+05 (<.001)	3.5e+05 (<.001)	3.5e+05 (<.001)	0.51	0.28	0.31	0.34	0.34

The 'full' model has superior fit characteristics across indices. The GLM model, which is closer to what we originally proposed, performs poorly.

```
h2sum<-summary(m.h2)
temp = capture.output(h2sum)
End  = grep("Fixed effects", temp)-1
print(unname(temp[0:End]))
```

```
##  [1] "Linear mixed model fit by REML. t-tests use Satterthwaite's method ["
##  [2] "lmerModLmerTest]"
##  [3] "Formula: "
##  [4] "missing ~ poly(pmiss, 2) + sex + sex:poly(pmiss, 2) + household.income +  "
##  [5] "    household.income:poly(pmiss, 2) + p.edu + p.edu:poly(pmiss,  "
##  [6] "    2) + race_ethnicity.factor + race_ethnicity.factor:poly(pmiss,  "
##  [7] "    2) + ksads_factor + ksads_factor:poly(pmiss, 2) + pds_category +  "
##  [8] "    pds_category:poly(pmiss, 2) + adi.z + adi.z:poly(pmiss, 2) +  "
##  [9] "    coi.z + coi.z:poly(pmiss, 2) + nihtbx_flanker_agecorrected.z +  "
## [10] "    nihtbx_flanker_agecorrected.z:poly(pmiss, 2) + nihtbx_cryst_agecorrected.z +  "
## [11] "    nihtbx_cryst_agecorrected.z:poly(pmiss, 2) + nihtbx_totalcomp_agecorrected.z +  "
## [12] "    nihtbx_totalcomp_agecorrected.z:poly(pmiss, 2) + pea_wiscv_tss.z +  "
## [13] "    pea_wiscv_tss.z:poly(pmiss, 2) + pfactor.z + pfactor.z:poly(pmiss,  "
## [14] "    2) + INT.z + INT.z:poly(pmiss, 2) + EXT.z + EXT.z:poly(pmiss,  "
## [15] "    2) + interview_age.z + interview_age.z:poly(pmiss, 2) + bmiAgeZ +  "
## [16] "    bmiAgeZ:poly(pmiss, 2) + (pmiss | subjectkey)"
## [17] "   Data: dflong"
## [18] ""
## [19] "REML criterion at convergence: -176817.8"
## [20] ""
## [21] "Scaled residuals: "
## [22] "    Min      1Q  Median      3Q     Max "
## [23] "-5.5260 -0.3032  0.0429  0.2260  5.0294 "
## [24] ""
## [25] "Random effects:"
## [26] " Groups     Name        Variance Std.Dev. Corr "
## [27] " subjectkey (Intercept) 0.26699  0.5167        "
## [28] "            pmiss       0.29719  0.5452   -0.94"
## [29] " Residual               0.03466  0.1862        "
## [30] "Number of obs: 463488, groups:  subjectkey, 9088"
## [31] ""
```

Table for the fixed effects:

```
pnames<-rownames(h2sum$coefficients)

h2table<-h2sum$coefficients %>% as.data.frame() %>%
  rownames_to_column(var = "Parameter") %>%
  left_join(parameters::ci(m.h2)) %>%
```

```
  mutate(star = sdlabFunctions::starPs(`Pr(>|t|)`)) %>%
  select(c(Parameter, Estimate, CI_low, CI_high,star)) %>%
  mutate(level=c(1:3,rep(1,31),rep(2:3,31)),
         basep=c(rep("",3),pnames[4:34],rep(pnames[4:34],each=2))) %>%
  pivot_wider(id_cols=basep,names_from=level,values_from = Estimate:star) %>%
  mutate(across(where(is.numeric),ORformat))

h2hux<-h2table[,c(1,2,5,8,11,3,6,9,12,4,7,10,13)] %>%
  as_hux(add_colnames=FALSE) %>%
  insert_row("","Intercept","","","","pmiss","","","","pmiss^2","","","") %>%
  merge_cells(1, 2:5) %>% merge_cells(1, 6:9) %>% merge_cells(1, 10:12) %>%
  insert_row("","Main Eff.","","","","Interaction","","","","Q.Interaction","","","",after=2) %>%
  merge_cells(3, 2:5) %>% merge_cells(3, 6:9) %>% merge_cells(3, 10:12) %>%
  insert_row("",rep(c("B","CI.l","CI.u","*"),3))

h2hux[,1] <- c(rep("",4),alllevels[-1])

h2hux <- h2hux %>% insert_row(
    "Household Income (ref: $100-$200k)",
    after = 6,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Highest Parental Education (ref: College Degree)",
    after = 11,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Census Race/Ethnicity (ref: White)",
    after = 13+3,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "KSADS Trauma Count (ref: 0 Exposures)",
    after = 18+3,
    colspan = 9,
    fill = ""
  ) %>%
  insert_row(
    "Pubertal Status (ref: pre-pubertal)",
    after = 21+3,
```

```
      colspan = 9,
      fill = ""
  )


h2hux[c(1:4,5:24),] %>% set_caption("H2 / Omnibus model parameters, pt. 1")



h2hux[c(1:4,25:nrow(h2hux)),] %>% set_caption("H2 / Omnibus model parameters, pt. 2")
```

1.7   H2 Marginal Means plotting (Figure 5)

```
# What ticks do we want for our plot?
xticks.mm<-c(.5,.45,.4,.35,.3,.25,.2,.15,.1,.05)
xticks.pmiss<-pmisssum[c("any_rest",paste0("gframes_",xticks.mm,"_375"))]
xticks.perc<-round(xticks.pmiss*100,0)
xticks.labela<-c("C",paste0(xticks.mm,"mm"))
xticks.label<-paste0(xticks.labela,"\n(",xticks.perc,"%)")
xticks.label[c(2:7)]<-""

# These functions all draw a plot with predicted probability of missingness
# (by group) - probability of missingness in the whole sample.
# 1 This one includes both model generated marginal means and simple mean
# inclusion.
detrend_catemplot_model<-function(x,label,model=NULL) {
  quosure<-quo(!!sym(x))
  if(!is.null(model)) {
  lmean <- emmeans(model,
                   as.formula(paste("~",x,"| pmiss")),
                     "pmiss",
                    at = list(pmiss=xticks.pmiss,
                              z=levels(df %>%
                                         group_by(!!quosure) %>%
                                         filter(n() >= 50) %>%
                                         ungroup() %>%
                                         select(!!quosure) %>%
                                         unlist())),
                 rg.limit = 400000)

  plot.dat<-emmip(lmean,
                  as.formula(paste(x,"~ pmiss")),
                  CIs=TRUE,
                  plotit = FALSE)
```

Table 6: H2 / Omnibus model parameters, pt. 1

	B	CI.l	CI.u	*	B	CI.l	CI.u	*	B	CI.l	CI.u	*
	Intercept				pmiss				pmiss^2			
	1.36	1.33	1.39	***	1.40e+87	9.68e+83	2.04e+90	***	4.66e+08	1.18e+08	1.84e+09	***
	Main Eff.				Interaction				Q.Interaction			
Sex (Male)	1.03	1.01	1.05	***	1.73	.01	329.14		.00	6.56e-08	.00	***
$0-$25k	1.02	.99	1.05		.00	.00	72.33		.69	.11	4.32	
Household Income (ref: $100-$200k)												
$25-$50k	.98	.96	1.01		105.65	.02	507780.08		44.64	9.04	220.55	***
$50-$75k	.99	.97	1.02		1165.69	.56	2426424.45		.01	.00	.06	***
$75-$100k	.99	.97	1.01		1.98	.00	2541.64		25.95	6.74	99.95	***
>$200k	.99	.97	1.02		.64	.00	1587.92		12.81	2.94	55.89	***
Highest Parental Education (ref: College Degree)												
<High School	1.07	1.02	1.12	**	8.22e-09	6.06e-15	.01	**	.00	.00	.05	***
HS Grad.	1.04	1.00	1.07	*	.00	2.26e-08	18.33		.00	.00	.01	***
Some College	1.00	.98	1.03		1.64	.00	1525.01		.06	.02	.22	***
Graduate	1.02	1.00	1.04	*	.08	.00	29.44		.05	.02	.15	***
Census Race/Ethnicity (ref: White)												
Black	1.07	1.05	1.10	***	3.79e-09	7.89e-13	.00	***	.00	.00	.01	***
Hispanic	1.02	1.00	1.04	*	.00	.00	1.51		.41	.11	1.46	
Asian	1.12	1.06	1.17	***	8.26e-12	.00	.00	**	.00	.00	.00	***
Other	1.03	1.01	1.05	*	.00	5.80e-09	.02	**	18.54	4.43	77.57	***
KSADS Trauma Count (ref: 0 Exposures)												
1 Trauma	1.00	.99	1.02		.23	.00	43.30		5.69	2.12	15.29	***
>=2 Trauma	1.02	.99	1.04		.07	.00	205.22		.00	.00	.00	***

Table 7: H2 / Omnibus model parameters, pt. 2

	B	CI.l	CI.u	*	B	CI.l	CI.u	*	B	CI.l	CI.u	*
	Intercept				pmiss				pmiss^2			
	1.36	1.33	1.39	***	1.40e+87	9.68e+83	2.04e+90	***	4.66e+08	1.18e+08	1.84e+09	***
	Main Eff.				Interaction				Q.Interaction			
Pubertal Status (ref: pre-pubertal)												
Early Puberty	1.00	.99	1.02		1.19	.00	359.61		.28	.10	.83	*
Mid Puberty	1.01	.99	1.03		2.05	.00	2067.75		.03	.01	.10	***
Late Puberty	1.06	1.00	1.13	*	1.91e-12	.00	.00	**	3926.89	109.53	140784.45	***
Post Pubertal	.94	.75	1.19		5.18e+15	.00	2.07e+48		1.14e-13	.00	.00	***
Area Disadvantage	.95	.94	.97	***	597703.26	13593.00	2.63e+07	***	2.14	1.05	4.36	*
Child Opportunity	.97	.95	.98	***	4424.30	64.41	303924.31	***	11.79	5.31	26.16	***
NIHTB Flanker	1.00	.99	1.01		1.25	.07	23.95		.68	.39	1.19	
NIHTB Crystalized	1.01	1.00	1.03	*	7.21	.07	784.75		.00	.00	.00	***
NIHTB Total	.97	.96	.99	**	1.21	.00	301.16		32303.00	11419.24	91379.42	***
WISC V Matrix	1.00	.99	1.00		.68	.05	8.64		20.63	12.78	33.30	***
Psychopathology	1.02	1.01	1.03	***	.04	.00	.41	**	.01	.01	.02	***
Internalizing	.99	.98	1.00	*	.85	.08	9.31		104.30	66.47	163.66	***
Externalizing	.99	.98	1.00	**	5.66	.53	59.92		6.72	4.31	10.49	***
Age	.97	.96	.98	***	23.98	2.39	240.40	**	3247.86	2103.56	5014.63	***
BMI	1.01	1.01	1.02	***	18.73	3.73	94.21	***	.00	.00	.00	***

```r
}
#Add detrended prediction
plot.dat$ydt<-plot.dat$yvar-plot.dat$pmiss
#Add the probability (average) without any cross control.

mdf<-dflong %>%
   select(c(!!quosure,pmiss,missing)) %>%
   group_by(!!quosure,pmiss) %>%
   filter(n() >= 50) %>% # Remove cases with n<50. Otherwise they are very noisy.
   summarise(avg=mean(missing)) %>%
   mutate(ydt=avg-pmiss,
          yvar=avg)
names(mdf)[1]<-"tvar"

ggplot(plot.dat,aes(x=pmiss,y=ydt,color=tvar,fill=tvar)) +
     geom_smooth(method="loess") +
     geom_point(data=mdf) +
     scale_color_viridis(name=x,discrete = TRUE,na.value="gray") +
     scale_fill_viridis(name=x,discrete= TRUE,na.value="gray") +
     labs(title=paste("De-trended Mean Missingness by",label)) +
   scale_x_continuous(breaks=xticks.pmiss,labels = xticks.label) +
   theme(axis.title.x=element_blank(),
         axis.title.y=element_blank(),
         legend.title=element_blank(),
         legend.text=element_text(size=10))# +
   #geom_function(fun = \(x) (-6.349e+00-3.078e-01)*x - (8.929e-01+1.996e+01)*x^2 + 3.078e-01)
}


#2 As above, but without specifying a model.

detrend_catemplot <- function(x, label) {
  quosure <- quo(!!sym(x))
  #The probability (average) without any cross control.
  mdf <- dflong %>%
    select(c(!!quosure, pmiss, missing)) %>%
    group_by(!!quosure, pmiss) %>%
    filter(n() >= 50) %>% # Remove cases with n<50. Otherwise they are very noisy.
    summarise(avg = mean(missing)) %>%
    mutate(ydt = avg - pmiss,
           yvar = avg)
  names(mdf)[1] <- "tvar"
  ggplot(mdf, aes(x = pmiss,y = ydt,color = tvar,fill = tvar)) +
    geom_point() +
    scale_color_viridis(name = x,
```

```r
                        discrete = TRUE,
                         na.value = "gray") +
      scale_fill_viridis(name = x,
                         discrete = TRUE,
                         na.value = "gray") +
      labs(title = paste("De-trended Mean Missingness by", label)) +
      scale_x_continuous(breaks = xticks.pmiss, labels = xticks.label) +
      theme(
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        legend.title = element_blank(),
        legend.text = element_text(size = 10)
      )
}

# 3 As above, but for continuous variables.
detrend_contemplot<-function(x,label) {
  quosure<-quo(!!sym(x))
  #Add the probability (average) without any cross control.
  mdf<-dflong %>%
    select(c(!!quosure,pmiss,missing)) %>%
    mutate(quantile=ntile(!!quosure,5)) %>%
    group_by(quantile,pmiss) %>%
    summarise(avg=mean(missing)) %>%
    mutate(ydt=avg-pmiss,
           Quantile=as.factor(quantile))
  ggplot(mdf,aes(x=pmiss,y=ydt,color=Quantile)) +
      geom_point() +
      scale_color_viridis(discrete=TRUE) +
      labs(title=paste("De-Trended Mean Missingness by",label)) +
    scale_x_continuous(breaks=xticks.pmiss,labels = xticks.label) +
    theme(axis.title.x=element_blank(),
          axis.title.y=element_blank(),
          legend.text=element_text(size=10))
}
```

If we generate a plot with both model predictions and the average inclusion within each factor level and motion threshold (here, lines are the model marginal means and points are the averages from the data), we can see that they do not match, at least for this variable:

```r
detrend_catemplot_model("race_ethnicity.factor","Census Race/Ethnicity",m.h2)
```

De-trended Mean Missingness by Census Race/Ethnicity
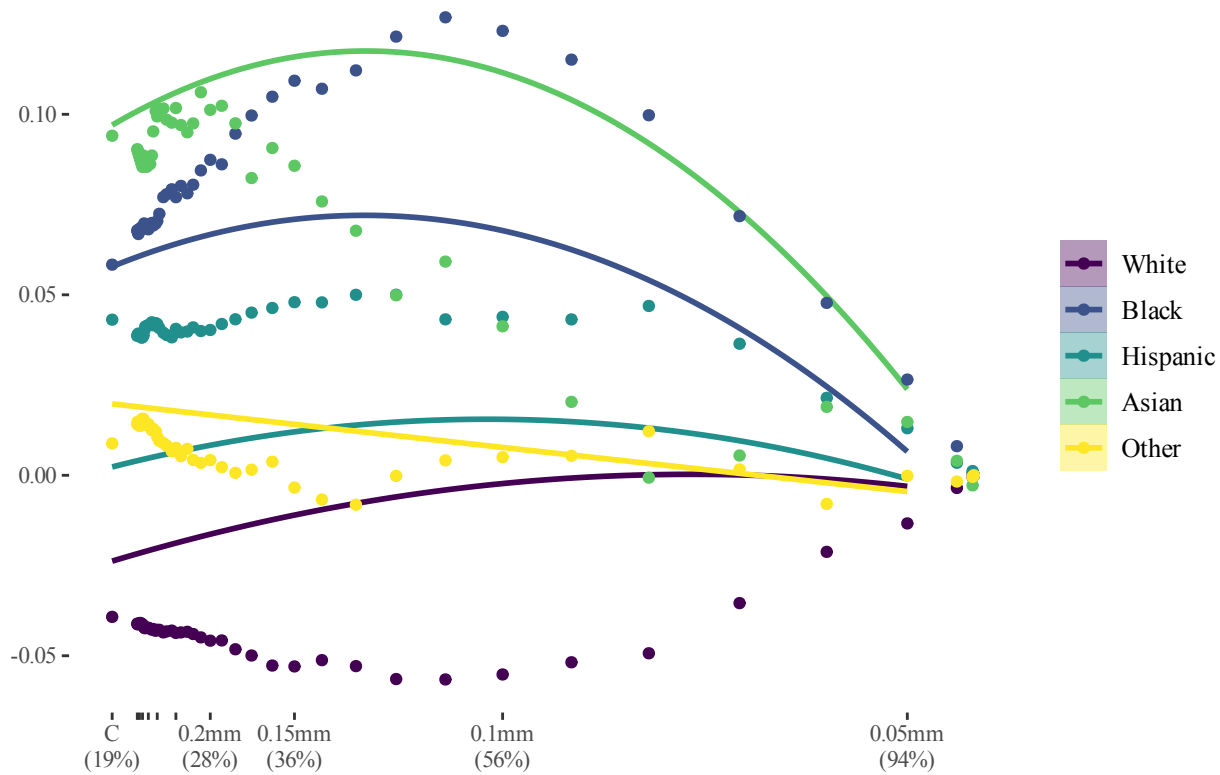
If we use a model with only the race/ethnicity factor:

```
# One variable, mixed-effect model.
m.h2.raceethnicity <- lmer(
  missing ~ poly(pmiss, 2) +
    race_ethnicity.factor + race_ethnicity.factor:poly(pmiss, 2) +
    (pmiss | subjectkey),
  data = dflong
)

detrend_catemplot_model("race_ethnicity.factor",
                        "Census Race/Ethnicity",
                        m.h2.raceethnicity)
```

De-trended Mean Missingness by Census Race/Ethnicity
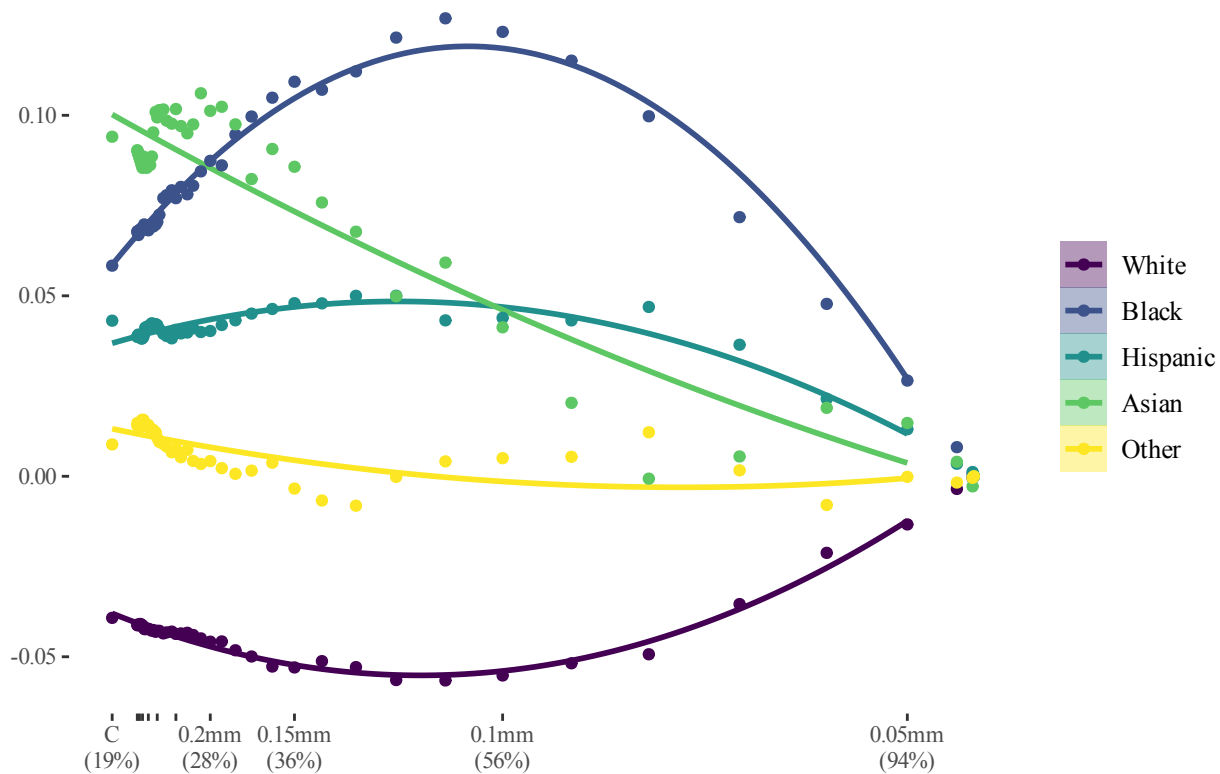
That is much improved, so the distortions in the first plot are most likely due to interference from other variables. Regardless H2 is generally not supported by these results. The descriptive data is probably more helpful (and will be presented).

```r
# Sub-Plots:
# Categorical:
h2p1<-detrend_catemplot("sex","Sex")
h2p2<-detrend_catemplot("ksads_factor","Trauma Count")
h2p3<-detrend_catemplot("race_ethnicity.factor","Census Race/Ethnicity")
h2p4<-detrend_catemplot("household.income","Household Income")
h2p5<-detrend_catemplot("p.edu","Highest Parental Education")
h2p6<-detrend_catemplot("pds_category","Pubertal Status")

# Continuous:
h2plot.adi<-detrend_contemplot("adi.z","ADI")
h2plot.coi<-detrend_contemplot("coi.z","COI")
h2plot.flank<-detrend_contemplot("nihtbx_flanker_agecorrected.z","NIHTB Flanker Score")
h2plot.cryst<-detrend_contemplot("nihtbx_cryst_agecorrected.z","NIHTB Crystalized")
```

```r
h2plot.ntot<-detrend_contemplot("nihtbx_totalcomp_agecorrected.z","NIHTB Total")
h2plot.wisc<-detrend_contemplot("pea_wiscv_tss.z","Wisc V Matrix")
h2plot.pfac<-detrend_contemplot("pfactor.z","P-factor")
h2plot.int<-detrend_contemplot("INT.z","Internalizing")
h2plot.ext<-detrend_contemplot("EXT.z","Externalizing")
h2plot.age<-detrend_contemplot("interview_age.z","Age")
h2plot.bmi<-detrend_contemplot("bmiAgeZ","BMI")

# A few outstanding effects are saved to a plot for the manuscript:
svg(filename="Manuscript/Figures/h2catfig.svg",height=9,width=7.5,family="Times New Roman")
h2p4 / h2p5 / h2plot.ntot / h2plot.bmi
grid::grid.draw(grid::textGrob("Proportion of Participants Excluded - Total Proportion Excluded",x=.01,rot=90,gp = grid::gpar(fontfamily = "Times
grid::grid.draw(grid::textGrob("Motion Threshold (Proportion of Sample Missing)",y=.01,gp = grid::gpar(fontfamily = "Times New Roman")))
dev.off()
```

```
## cairo_pdf
##         2
```

For this supplement, we will also generate plots for the other variables in groups of four:

```r
h2p1 / h2p2 / h2p3 / h2p6
grid::grid.draw(grid::textGrob("Proportion of Participants Excluded - Sample Mean",x=.01,rot=90,gp = grid::gpar(fontfamily = "Times New Roman")))
grid::grid.draw(grid::textGrob("Motion Threshold / Proportion of Sample Missing",y=.01,gp = grid::gpar(fontfamily = "Times New Roman")))
```

```r
h2plot.adi / h2plot.coi / h2plot.flank / h2plot.cryst
grid::grid.draw(grid::textGrob("Proportion of Participants Excluded - Sample Mean",x=.01,rot=90,gp = grid::gpar(fontfamily = "Times New Roman")))
grid::grid.draw(grid::textGrob("Motion Threshold / Proportion of Sample Missing",y=.01,gp = grid::gpar(fontfamily = "Times New Roman")))
```

```r
h2plot.wisc / h2plot.pfac / h2plot.int / h2plot.ext / h2plot.age
grid::grid.draw(grid::textGrob("Proportion of Participants Excluded - Sample Mean",x=.01,rot=90,gp = grid::gpar(fontfamily = "Times New Roman")))
grid::grid.draw(grid::textGrob("Motion Threshold / Proportion of Sample Missing",y=.01,gp = grid::gpar(fontfamily = "Times New Roman")))
```

1.8   QCFC Figures (Figure 6)

```r
# Get the Glasser Distances
files=c("ABCC_download/GlasserDistances/glasserdistances.txt",
        list.files("ABCC_download/",pattern="QCFC.*csv",full.names = TRUE))

#A list of vectors:
```
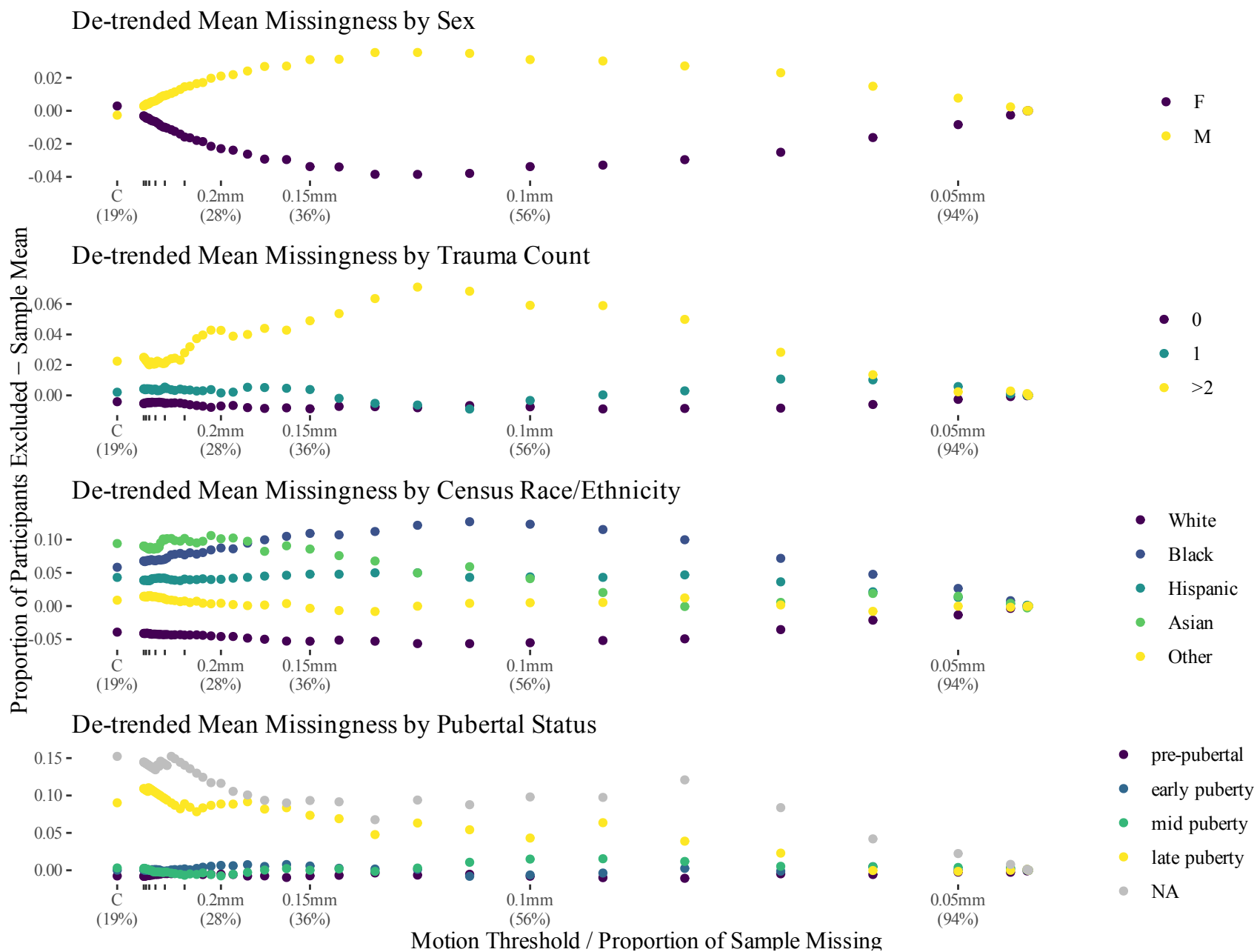
Figure 3: Mean Inclusion/Exclusion by Supplemental Variables and condition (part 1)
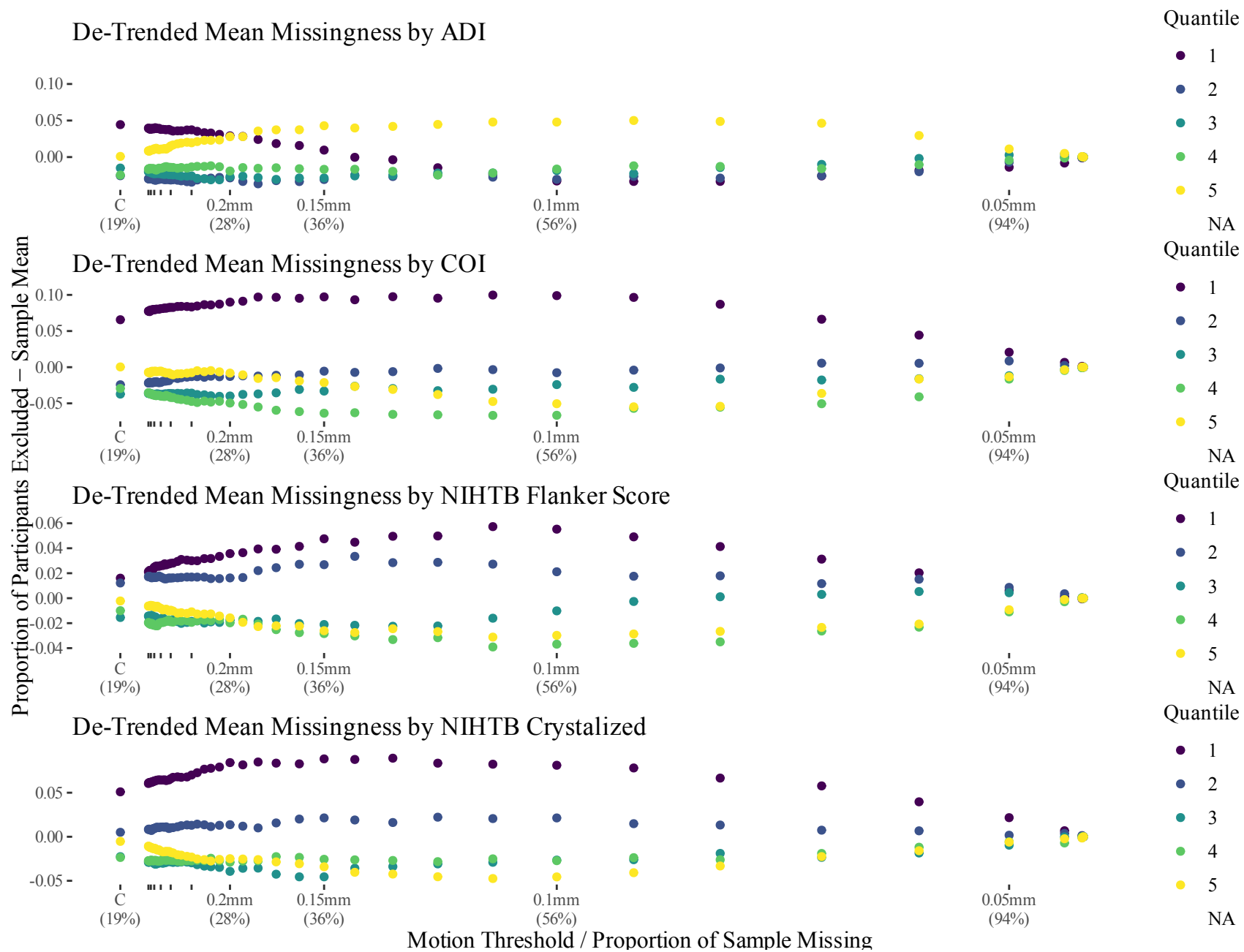
Figure 4: Mean Inclusion/Exclusion by Supplemental Variables and Condition (part 2)

Figure 5: Mean Inclusion/Exclusion by Supplemental Variables and Condition (part 3)

```r
QCFCvals<-lapply(files,scan,sep=",")

# The distance matrix only has cortical (as opposed to subcortical distances, so
# we trim the QC-FC values to those dimensions. All of these matrices are
# symmetrical across the diagonal, so we can also omit the diagonal and lower
# triangle.

tdim=sqrt(length(QCFCvals[[1]]))
QCFCtrimvals <- lapply(QCFCvals,\(x) {
  m=matrix(x,ncol=sqrt(length(x)),byrow=TRUE)
  m2=m[1:tdim,1:tdim]
  m2[lower.tri(m2,diag = TRUE)] <- NA
  na.omit(c(m2))
  })

QCFCdf<-as.data.frame(reduce(QCFCtrimvals,cbind))
names(QCFCdf)<-c("distance","<.1mm","<.2mm","<.3mm","<.4mm","<.5mm","No Scrubbing")
QCFCdflong <- QCFCdf %>% pivot_longer(cols=-distance,
                                      names_to="threshold",
                                      values_to="QCFC")

# Calculate Slopes
QCFCdfz <- QCFCdf %>%
  mutate(across(.fns=scale))

annotext<-function(x) {
  m<-lm(QCFCdfz$distance ~ QCFCdfz[,x])
  ci<-confint(m,level=.80)
  spec="%.2f"
  paste0(" = ",
         sprintf(spec,m$coefficients[2]),
         " 99% CI [",
         sprintf(spec,ci[2,1]),
         " ",
         sprintf(spec,ci[2,2]),
         "]")
}

annot<-data.frame(threshold=unique(QCFCdflong$threshold),
                  text=sapply(unique(QCFCdflong$threshold),annotext))
annot$text<-gsub("0\\.", ".", annot$text)
annot$distance=mean(range(QCFCdflong$distance))
annot$QCFC=.1
```

```
ggsave("Manuscript/Figures/qcfc.svg",
ggplot(QCFCdflong,aes(x=distance,y=QCFC)) +
  geom_hex(aes(color=..count..),bins=40) +
  labs(x="ROI Distance (mm)",
       y="QC-FC") +
  theme(legend.position="none") +
  scale_fill_gradient(low="white",high="black") +
  scale_colour_gradient(low="white",high="black") +
  geom_smooth(method="lm",color="red") +
  facet_wrap(~threshold) +
  geom_text(data=annot,mapping=aes(label=text,family="Times New Roman"),parse=FALSE),
height=5.5,width=7.5,)
```

1.9   Multiply at Risk Cell Counts

```
subdf1 <- df %>%
  filter(race_ethnicity.factor!="White",
         pfactor.z>=1.5)

subtable1<-data.frame(
  "F"=nrow(subdf1),
  "T"=sum(subdf1$ABCD_rsfmri_QC1),
  "C"=sum(subdf1$any_rest),
  ".5"=sum(subdf1$gframes_0.5_375),
  ".4"=sum(subdf1$gframes_0.4_375),
  ".3"=sum(subdf1$gframes_0.3_375),
  "R"=sum(subdf1$ABCD_rsfmri_QC2),
  ".2"=sum(subdf1$gframes_0.2_375),
  ".1"=sum(subdf1$gframes_0.1_375)
  ) %>%
  as_hux() %>%
  set_header_rows(1,TRUE) %>%
  set_caption("Non-white (census) with psychopathology at z >= 1.5")


subtable1[1,]<-gsub("X\\.","\\.",subtable1[1,])

subdf2 <- df %>%
  filter(nihtbx_totalcomp_agecorrected.z<=-1.5,
         sex=="M")

subtable2<-data.frame(
```

```
  "F"=nrow(subdf2),
  "T"=sum(subdf2$ABCD_rsfmri_QC1),
  "C"=sum(subdf2$any_rest),
  ".5"=sum(subdf2$gframes_0.5_375),
  ".4"=sum(subdf2$gframes_0.4_375),
  ".3"=sum(subdf2$gframes_0.3_375),
  "R"=sum(subdf2$ABCD_rsfmri_QC2),
  ".2"=sum(subdf2$gframes_0.2_375),
  ".1"=sum(subdf2$gframes_0.1_375)
) %>%
as_hux() %>%
set_header_rows(1,TRUE) %>%
set_caption("Male participants with NIH toolbox total scores of z <= -1.5")

subtable2[1,]<-gsub("X\\.","\\.",subtable2[1,])
```

subtable1

Table 8: Non-white (census) with psychopathology at z >= 1.5

F	T	C	.5	.4	.3	R	.2	.1
552	510	407	388	385	373	399	339	180

Vs. 0.1797951

subtable2

Table 9: Male participants with NIH toolbox total scores of z <= -1.5

F	T	C	.5	.4	.3	R	.2	.1
382	343	275	254	248	235	232	207	94

Vs. 0.180442

2 Supplementary Results

2.1 Categorical Variable Table by Condition

```r
catvars<-c("sex",
           "household.income",
           "p.edu",
           "race_ethnicity.factor",
           "ksads_factor",
           "pds_category")

catTableAtThresh <- function(dataorig,
                             boolean,
                             base = NULL,
                             vars = catvars) {
  n <- sum(boolean)
  tl <- list(cbind(n, n * 100 / 11876))
  pdivisor <- n
  data <- dataorig[boolean, ]
  row.names(tl[[1]]) <- "Total"
  for (i in vars) {
    st <-
      cbind(table(data[, i], useNA = "always"), (table(data[, i], useNA = "always") *
                                                   100) / pdivisor)
    tl[[length(tl) + 1]] <- st
  }
  t <- reduce(tl, rbind)
  if (is.null(base)) {
    t
  } else {
    cbind(t, t[, 2] - base[, 2])
  }
}

CatVar375Table <- cbind(
  catTableAtThresh(df, rep(TRUE,nrow(df))),
  catTableAtThresh(df, df$ABCD_rsfmri_QC1),
  catTableAtThresh(df, df$any_rest),
  catTableAtThresh(df, df$gframes_0.5_375),
  catTableAtThresh(df, df$gframes_0.4_375),
  catTableAtThresh(df, df$gframes_0.3_375),
  catTableAtThresh(df, df$ABCD_rsfmri_QC2),
  catTableAtThresh(df, df$gframes_0.2_375),
```

```r
  catTableAtThresh(df, df$gframes_0.1_375)
)

#Try to draw as a huxtable
rownames(CatVar375Table)<-gsub("\\$","\\\\$",rownames(CatVar375Table))
CatVar375Hux<-as_hux(CatVar375Table) %>%
  theme_basic() %>%
  insert_row(value=c("n","\\%",rep(c("n","\\%"),8))) %>%
  insert_row("Full","",
             "QC1","",
             "ABCC","",
             ".5","",
             ".4","",
             ".3mm","",
             "QC2","",
             ".2mm","",
             ".1mm","") %>%
  merge_cells(1,1:2) %>%
  merge_cells(1,3:4) %>%
  merge_cells(1,5:6) %>%
  merge_cells(1,7:8) %>%
  merge_cells(1,9:10) %>%
  merge_cells(1,11:12) %>%
  merge_cells(1,13:14) %>%
  merge_cells(1,15:16) %>%
  merge_cells(1,17:18) %>%
  insert_column(value=c("","",rownames(CatVar375Table) %>% replace_na("Missing"))) %>%
  insert_column(value=c("","","Total","Sex","","",
                        "Income","","","","","","",
                        "Parent Ed.","","","","","","",
                        "Race/Ethnicity","","","","","",
                        "Trauma Count","","","",
                        "Puberty","","","","","")) %>%
  set_caption("Categorical Values at Each Level of Stringency (Subjects excluded with < 375 Frames), Pt 1") %>%
  set_escape_contents(FALSE)  %>%
  set_font_size(8) %>%
  set_number_format(3:nrow(.),3+0:7*2,value=fmt_pretty()) %>%
  set_number_format(3:nrow(.),4+0:7*2,value="%.2g") %>%
  set_all_padding(1)

CatVar375Hux
```

Table 10: Categorical Values at Each Level of Stringency (Subjects excluded with < 375 Frames), Pt 1

		Full		QC1		ABCC		0.5		0.4		0.3mm		QC2		0.2mm		0.1mm	
		n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Total	Total	11,876	1e+02	11,355	96	9,600	81	9,320	78	9,262	78	9,098	77	9,627	81	8,507	72	5.25e+03	44.2
Sex	F	5,680	48	5,444	48	4,575	48	4,475	48	4,458	48	4,408	48	4,792	50	4,199	49	2.7e+03	51.5
	M	6,196	52	5,911	52	5,025	52	4,845	52	4,804	52	4,690	52	4,835	50	4,308	51	2.55e+03	48.5
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Income	$100k to $200k	3,314	28	3,192	28	2,735	28	2,673	29	2,660	29	2,623	29	2,785	29	2,470	29	1.62e+03	30.9
	$0 to $25k	1,635	14	1,526	13	1,206	13	1,139	12	1,126	12	1,106	12	1,188	12	988	12	538	10.2
	$25k to $50k	1,588	13	1,524	13	1,291	13	1,250	13	1,239	13	1,212	13	1,277	13	1,135	13	676	12.9
	$50k to $75k	1,499	13	1,441	13	1,254	13	1,215	13	1,208	13	1,189	13	1,204	13	1,106	13	630	12
	$75k to $100k	1,572	13	1,508	13	1,316	14	1,281	14	1,276	14	1,250	14	1,315	14	1,188	14	773	14.7
	Over $200k	1,250	11	1,206	11	1,014	11	995	11	992	11	978	11	1,056	11	930	11	626	11.9
	Missing	1,018	8.6	958	8.4	784	8.2	767	8.2	761	8.2	740	8.1	802	8.3	690	8.1	385	7.33
Parent Ed.	College Degree	3,015	25	2,877	25	2,504	26	2,441	26	2,429	26	2,393	26	2,486	26	2,268	27	1.44e+03	27.4
	< HS	593	5	551	4.9	388	4	370	4	365	3.9	359	3.9	439	4.6	324	3.8	173	3.3
	HS Graduate	1,132	9.5	1,068	9.4	855	8.9	815	8.7	809	8.7	786	8.6	846	8.8	721	8.5	373	7.1
	Some College	3,079	26	2,957	26	2,528	26	2,432	26	2,410	26	2,361	26	2,451	25	2,179	26	1.29e+03	24.6
	Graduate Degree	4,043	34	3,889	34	3,313	35	3,250	35	3,238	35	3,188	35	3,396	35	3,007	35	1.97e+03	37.5
	Missing	14	0.12	13	0.11	12	0.12	12	0.13	11	0.12	11	0.12	9	0.093	8	0.094	3	0.0571
Race/Ethnicity	White	6,180	52	5,945	52	5,238	55	5,104	55	5,075	55	4,999	55	5,146	53	4,710	55	3.07e+03	58.5
	Black	1,784	15	1,672	15	1,338	14	1,279	14	1,269	14	1,241	14	1,313	14	1,122	13	569	10.8
	Hispanic	2,411	20	2,312	20	1,845	19	1,799	19	1,788	19	1,746	19	1,966	20	1,630	19	960	18.3
	Asian	252	2.1	236	2.1	180	1.9	175	1.9	175	1.9	168	1.8	196	2	155	1.8	101	1.92
	Other	1,247	11	1,188	10	997	10	961	10	953	10	942	10	1,004	10	888	10	545	10.4
	Missing	2	0.017	2	0.018	2	0.021	2	0.021	2	0.022	2	0.022	2	0.021	2	0.024	2	0.0381
Trauma Count	0	7,723	65	7,383	65	6,275	65	6,102	65	6,059	65	5,957	65	6,295	65	5,586	66	3.47e+03	66.1
	1	3,004	25	2,882	25	2,422	25	2,345	25	2,330	25	2,285	25	2,437	25	2,147	25	1.34e+03	25.5
	>2	1,149	9.7	1,090	9.6	903	9.4	873	9.4	873	9.4	856	9.4	895	9.3	774	9.1	440	8.38
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Puberty	pre-pubertal	5,938	50	5,690	50	4,845	50	4,707	51	4,671	50	4,575	50	4,818	50	4,284	50	2.67e+03	50.9
	early puberty	2,815	24	2,680	24	2,275	24	2,207	24	2,195	24	2,160	24	2,268	24	1,998	23	1.26e+03	24
	mid puberty	2,798	24	2,681	24	2,253	23	2,190	23	2,181	24	2,150	24	2,289	24	2,025	24	1.2e+03	22.8
	late puberty	188	1.6	180	1.6	135	1.4	127	1.4	126	1.4	126	1.4	152	1.6	118	1.4	75	1.43
	post pubertal	12	0.1	10	0.088	10	0.1	9	0.097	9	0.097	9	0.099	7	0.073	7	0.082	3	0.0571
	Missing	125	1.1	114	1	82	0.85	80	0.86	80	0.86	78	0.86	93	0.97	75	0.88	43	0.819

2.2 Alternative Continuous Variable Plots (psychopathology)

Here are some alternative ways of plotting continuous variables (using psychopathology as an example).

```
df %>% filter(ABCD_rsfmri_QC2==TRUE) %>%
ggplot(aes(x=pfactor.z)) +
  geom_density(aes(color = "F"), alpha = .2, data = df %>% filter(ABCD_rsfmri_QC1==TRUE)) +
  geom_density(aes(color = "C"), alpha = .2, data = df %>% filter(any_rest==TRUE)) +
  geom_density(aes(color = "R"), alpha = .2, data = df %>% filter(ABCD_rsfmri_QC2==TRUE)) +
  geom_density(aes(color = "1"), alpha = .2, data = df %>% filter(gframes_0.1_375==TRUE))
```
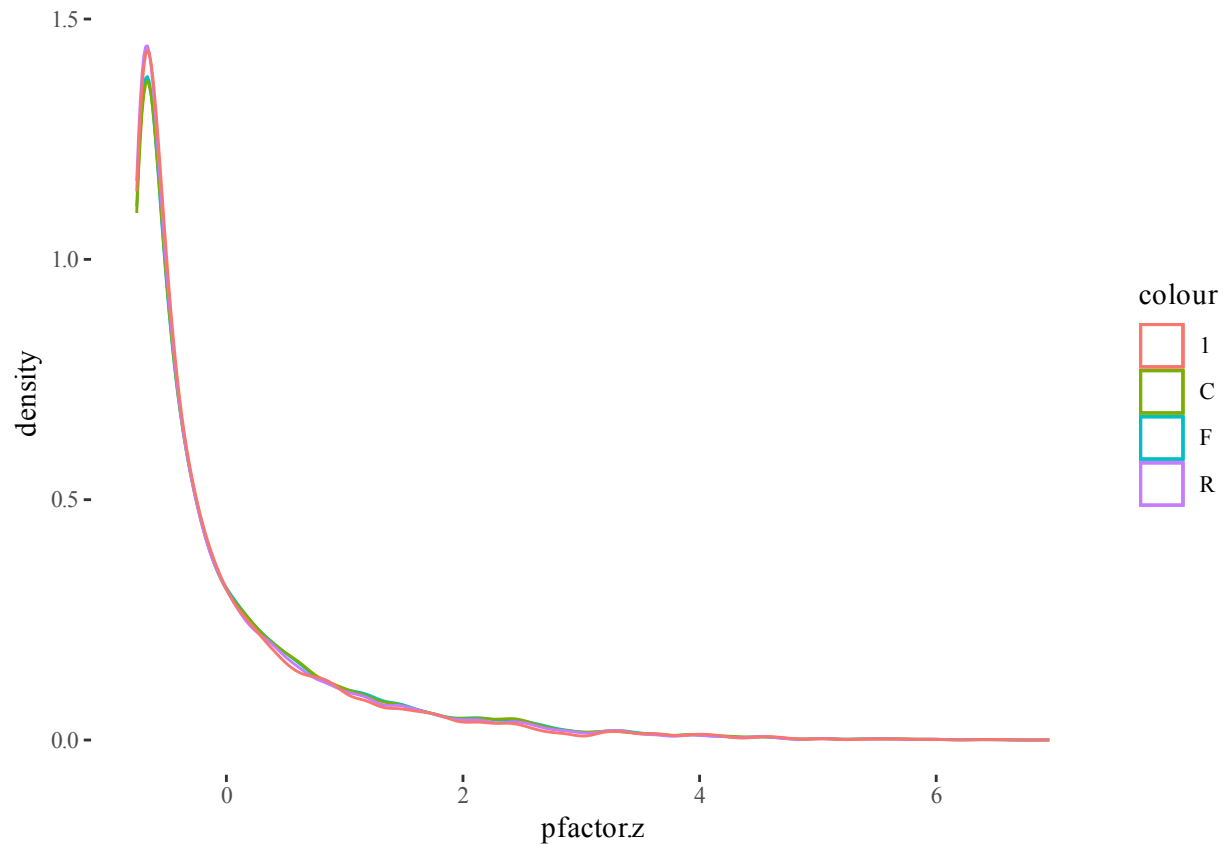


```
psubset<-df %>% filter(ABCD_rsfmri_QC2==TRUE) %>% select(pfactor)
psubset2<-df %>% filter(ABCD_rsfmri_QC1==TRUE) %>% select(pfactor)

test<-as.data.frame(rbind(quantile(psubset2$pfactor,probs=c(.05,.20,.35,.5,.65,.80,.95),na.rm=TRUE),
          quantile(psubset$pfactor,probs=c(.05,.20,.35,.5,.65,.80,.95),na.rm=TRUE)))
```

```
test$label=as.factor(c("T","R"))

ggplot(test,aes(x=label)) +
  geom_segment(aes(xend=label,y=`5%`,yend=`95%`),color="red",size=2) +
  geom_segment(aes(xend=label,y=`20%`,yend=`80%`),color="blue",size=2) +
  geom_segment(aes(xend=label,y=`35%`,yend=`65%`),color="yellow",size=2) +
  geom_point(aes(y=`50%`))
```
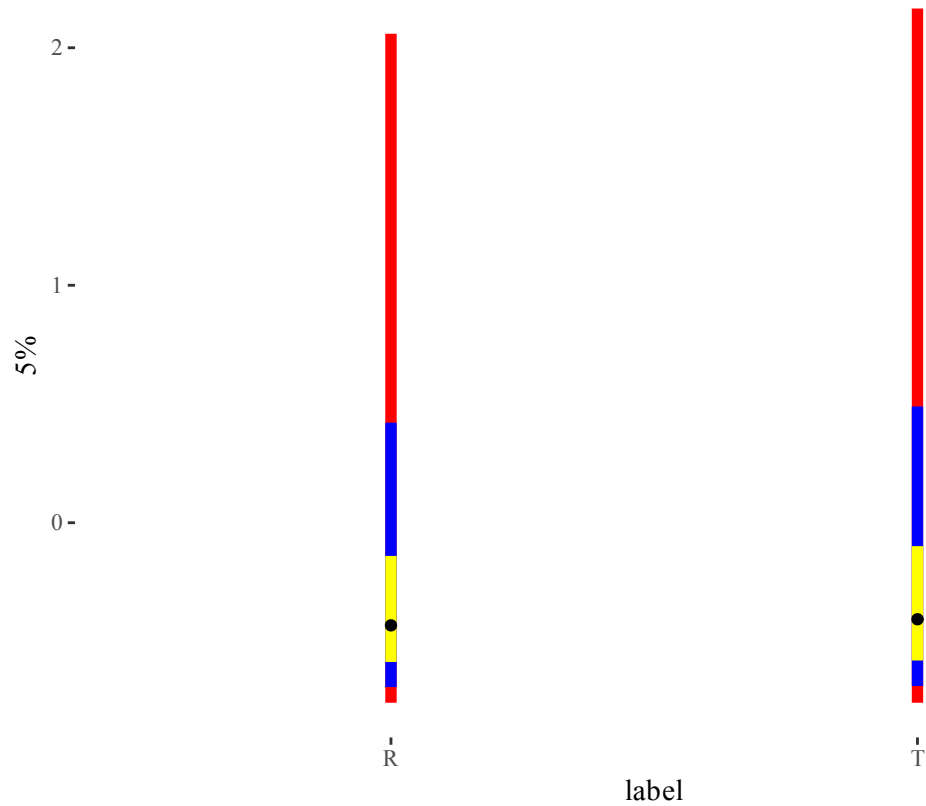


```
ggplot(df,aes(x=filtered_fd_0.2,y=pfactor)) +
  geom_bin2d()
```
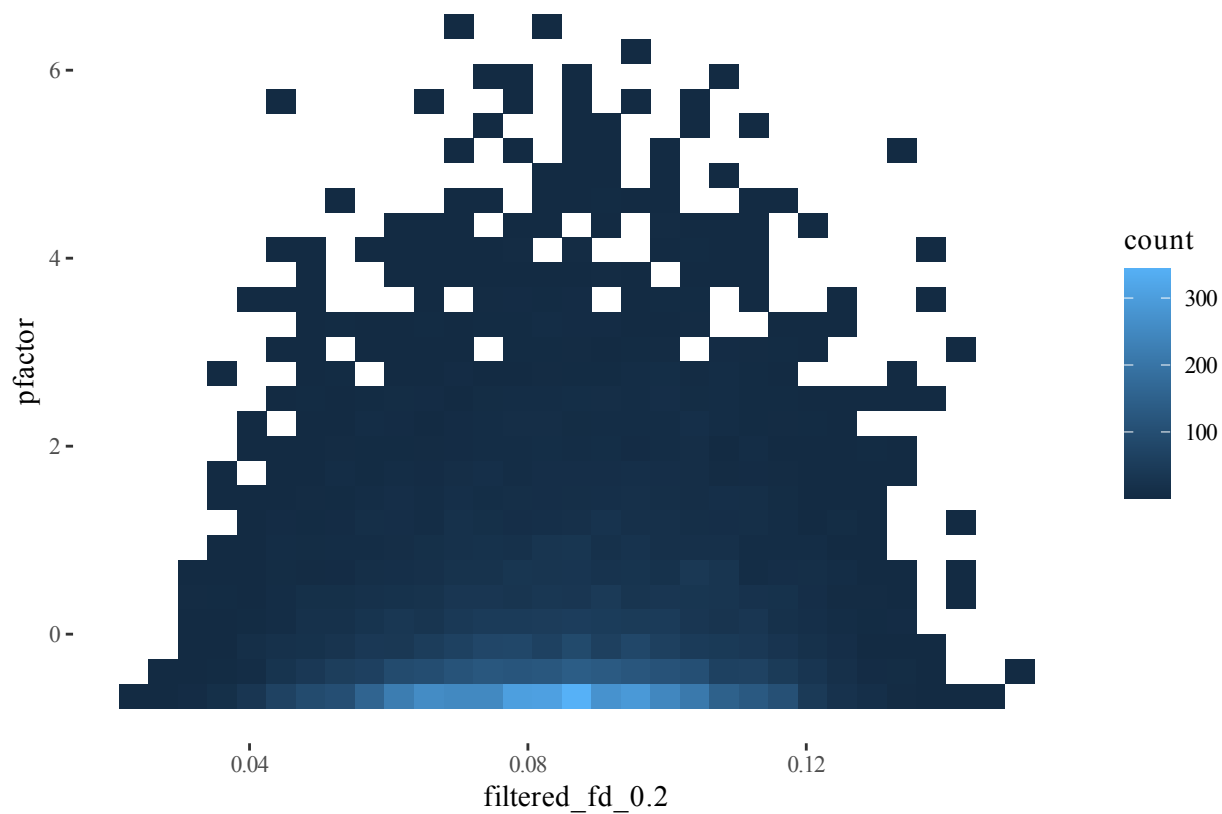
2.3 Missingness in Behavioral Data

This table shows missingness counts in non-imaging variables with more than 50 missing cases.

```
df %>%
  select(c(interview_age:household.income,p.edu,race_ethnicity.factor:site_id_l)) %>%
  rename(c("Child Opportunity Index"="coi","Household Income"="household.income","Area Disadvantage Index"="adi","NIH Toolbox Total"="nihtbx_tota
  summarise_all(~sum(is.na(.))) %>%
  pivot_longer(everything()) %>%
  filter(value>50) %>%
  mutate(`% missing`=(value/nrow(df))*100) %>%
  arrange(desc(value)) %>%
  rename(c("Variable"="name","n Missing"="value")) %>% as_hux() %>%
  set_header_rows(1,TRUE) %>%
  set_caption("Behavioral and Demographic variables with >50 missing values")
```

Table 11: Behavioral and Demographic variables with
>50 missing values

Variable	n Missing	% missing
Child Opportunity Index	1093	9.2
Household Income	1018	8.57
Area Disadvantage Index	863	7.27
NIH Toolbox Total	397	3.34
NIH Toolbox Crystalized	338	2.85
WISC V	249	2.1
ethn.iden.hisp	153	1.29
NIH Toolbox Flanker	153	1.29
Pubertal Status	125	1.05

2.4 Adjusted Models – Forest Plot

A graphical depiction of effects from the adjusted/condition models.

```
MFformat<-\(m) {
  stars<-sdlabFunctions::starPs(m[,4])
  paste0(ORformat(m[,2]),
         stars,
         "\n[",ORformat(m[,1]),"-",ORformat(m[,3]),"]")
}

allmodelterms<-c(
    "(Intercept)",
    "sexM",
    "household.incomeheader",
    "household.income$0 to $25k",
    "household.income$25k to $50k",
    "household.income$50k to $75k",
    "household.income$75k to $100k",
    "household.incomeOver $200k",
    "p.eduheader",
    "p.edu< HS",
```

```
    "p.eduGraduate Degree",
    "p.eduHS Graduate",
    "p.eduSome College",
    "race_ethnicityheader",
    "race_ethnicity.factorAsian",
    "race_ethnicity.factorBlack",
    "race_ethnicity.factorHispanic",
    "race_ethnicity.factorOther",
    "ksadsheader",
    "ksads_factor>2",
    "ksads_factor1",
    "pdsheader",
    "pds_categoryearly puberty",
    "pds_categorylate puberty",
    "pds_categorymid puberty",
    "pds_categorypost pubertal",
    "adi.z",
    "coi.z",
    "nihtbx_flanker_agecorrected.z",
    "nihtbx_cryst_agecorrected.z",
    "nihtbx_totalcomp_agecorrected.z",
    "pea_wiscv_tss.z",
    "pfactor.z",
    "INT.z",
    "EXT.z",
    "interview_age.z",
    "bmiAgeZ")

alllevelsggplot<-c("Intercept",
    "Sex (Male)",
    "**Household Income**",
    "  $0-$25k",
    "  $25-$50k",
    "  $50-$75k",
    "  $75-$100k",
    "  &gt;$200k",
    "**Highest Parental Education**",
    "  &lt;High School",
    "  HS Grad.",
    "  Some College",
    "  Graduate",
    "**Census Race/Ethnicity**",
    "  Black",
    "  Hispanic",
```

```
               "   Asian",
               "   Other",
               "**Trauma Count**",
               "   1 Trauma",
               "   &gt;=2 Trauma",
               "**Pubertal Status**",
               "   Early Puberty",
               "   Mid Puberty",
               "   Late Puberty",
               "   Post Pubertal",
               "Area Disadvantage",
               "Child Opportunity",
               "NIHTB Flanker",
               "NIHTB Crystalized",
               "NIHTB Total",
               "WISC V Matrix",
               "Psychopathology",
               "Internalizing",
               "Externalizing",
               "Age",
               "BMI"
               )

modeltermsA<-allmodelterms[1:18]
modeltermLabelsA<-alllevelsggplot[1:18]
modeltermsB<-allmodelterms[19:length(allmodelterms)]
modeltermLabelsB<-alllevelsggplot[19:length(allmodelterms)]

forestdf <- rbind(
  model.QC1.results %>% data.frame() %>% rownames_to_column(),
  model.ABCC.results %>% data.frame() %>% rownames_to_column(),
  model.5mm.results %>% data.frame() %>% rownames_to_column(),
  model.4mm.results %>% data.frame() %>% rownames_to_column(),
  model.3mm.results %>% data.frame() %>% rownames_to_column(),
  model.QC2.results %>% data.frame() %>% rownames_to_column(),
  model.2mm.results %>% data.frame() %>% rownames_to_column(),
  model.1mm.results %>% data.frame() %>% rownames_to_column()
) %>% mutate(threshold=factor(rep(c("T","C","5","4","3","R","2","1"),each=nrow(model.QC1.results)))) %>%
  rename_with(~ c("term","lowCI","OR","highCI","p","threshold")) %>%
  mutate(across(lowCI:highCI,exp)) %>%
  add_row(term="household.incomeheader") %>%
  add_row(term="p.eduheader") %>%
  add_row(term="ksadsheader") %>%
  add_row(term="pdsheader") %>%
```

```r
  add_row(term="race_ethnicityheader") %>%
  mutate(threshold=factor(threshold))

maxsigCI<-max(forestdf[forestdf$lowCI>1,"highCI"],na.rm=TRUE)
forestdf[forestdf$highCI>maxsigCI & !is.na(forestdf$highCI),"highCI"]<-maxsigCI


forestdfA<- forestdf %>% filter(term %in% modeltermsA) %>%
  mutate(term=factor(term,
                     levels=modeltermsA,
                     labels=modeltermLabelsA))
forestdfB<- forestdf %>% filter(term %in% modeltermsB) %>%
  mutate(term=factor(term,
                     levels=modeltermsB,
                     labels=modeltermLabelsB))

dodge<-.8


ggplot(forestdfA,aes(y=term,color=threshold)) +
  geom_vline(lty=2,aes(xintercept=1),colour='black') +
  geom_point(aes(x=OR),position=ggstance::position_dodgev(height=dodge)) +
  geom_linerange(aes(xmin=lowCI,xmax=highCI,y=term),position=ggstance::position_dodgev(height=dodge)) +
  xlim(0,maxsigCI) +
  geom_vline(lty=2,aes(xintercept=1),colour='black') +
  scale_y_discrete(drop=FALSE,limits=rev) +
  scale_color_discrete(limits=rev,drop=TRUE,na.translate = F) +
  theme(axis.text.y=element_markdown(),
        axis.title.y=element_blank()) +
  xlab("Odds Ratio of Missingness + 95% CI")


ggplot(forestdfB,aes(y=term,color=threshold)) +
  geom_vline(lty=2,aes(xintercept=1),colour='black') +
  geom_point(aes(x=OR),position=ggstance::position_dodgev(height=dodge)) +
  geom_linerange(aes(xmin=lowCI,xmax=highCI,y=term),position=ggstance::position_dodgev(height=dodge)) +
  xlim(0,maxsigCI) +
  scale_y_discrete(drop=FALSE,limits=rev) +
  scale_color_discrete(limits=rev,drop=TRUE,na.translate = F) +
  theme(axis.text.y=element_markdown(),
        axis.title.y=element_blank()) +
  xlab("Odds Ratio of Missingness + 95% CI")
```
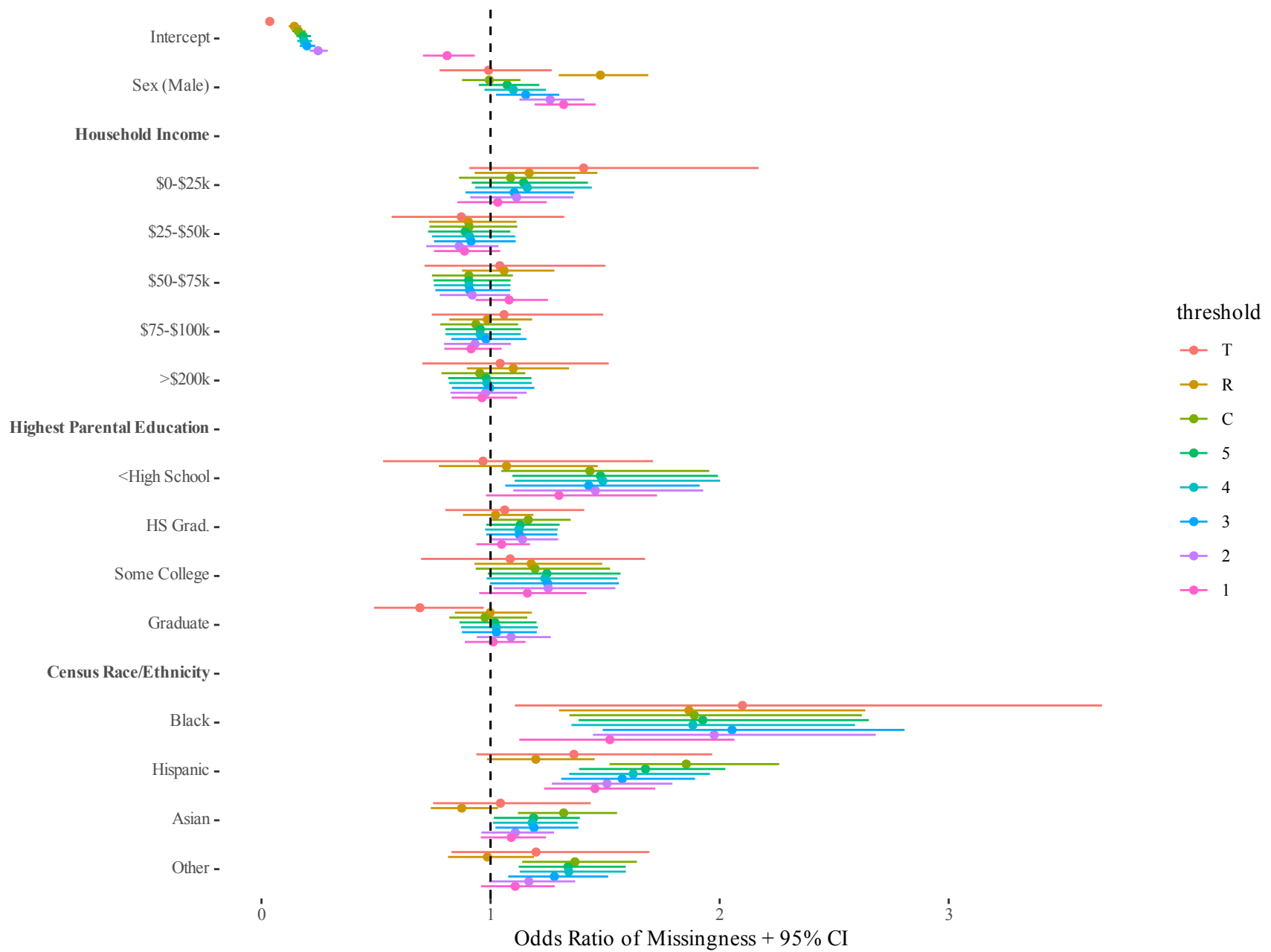
Figure 6: Odds ratios for missingness by threshold – Adjusted Models (part 1)
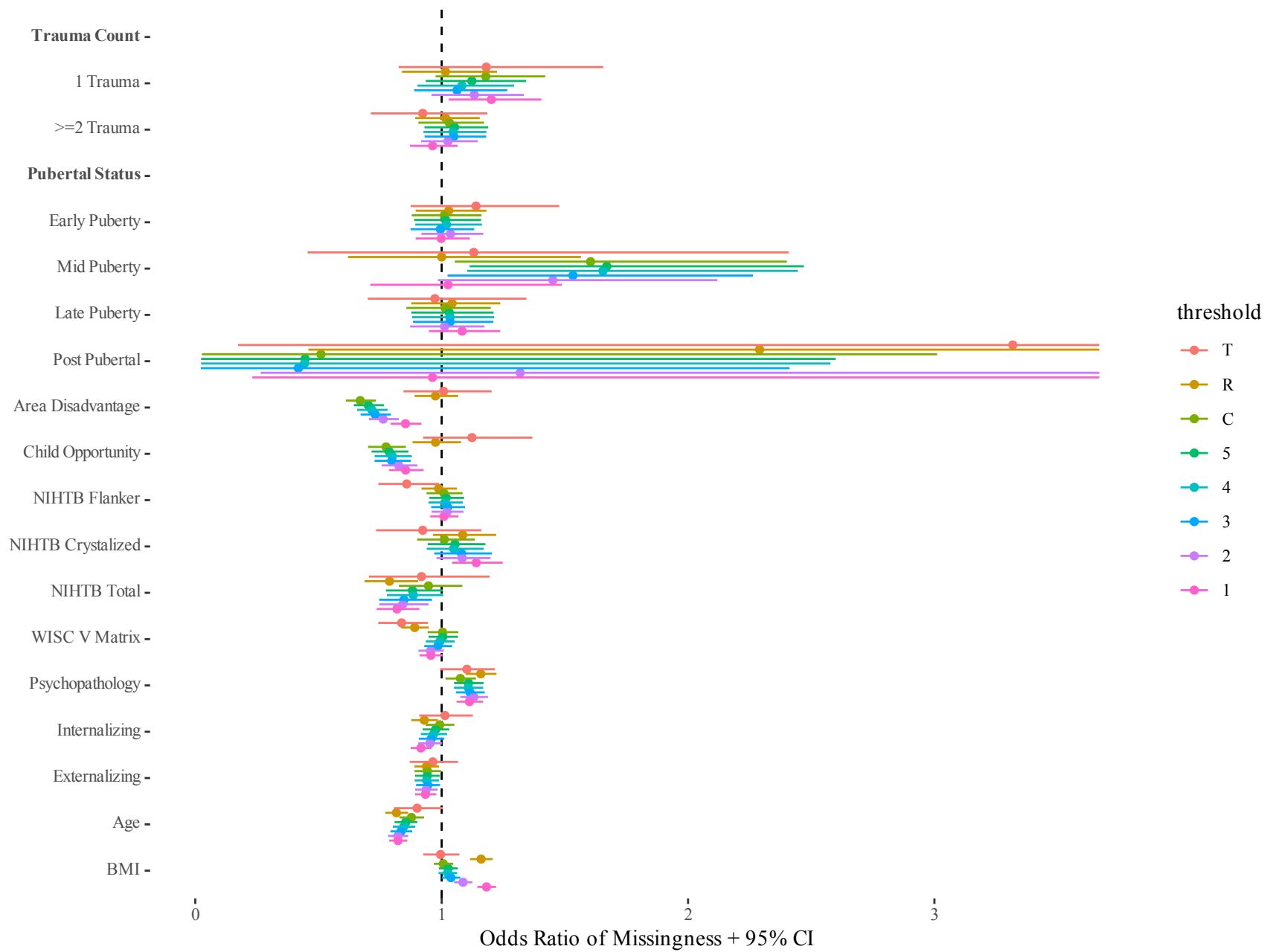
Figure 7: Odds ratios for missingness by threshold – Adjusted Models (part 2)

2.5   Current Fast Track QC vs DAIRC recommendations

We state in the text that the currently published fastqc is non-overlapping with DAIRC inclusion recommendations. This table illustrates that non-overlap.

The variable 'fastqcok' is TRUE for participants who have at least one T1, rs-fMRI, and field map image marked useable in the current (Last modified 09/09/2019) abcd_fastqc01.csv file. These recommendations are non overlapping with the tabular data list as well as DAIRC recommendations:

```
table(df$fastqcok,df$ABCD_rsfmri_QC1,dnn=c("FastQC","Tabulated"))
```

```
##        Tabulated
## FastQC  FALSE TRUE
##   FALSE   410 1378
##   TRUE    111 9977
```

```
table(df$fastqcok,df$ABCD_rsfmri_QC2,dnn=c("FastQC","DAIRC"))
```

```
##        DAIRC
## FastQC  FALSE TRUE
##   FALSE   822  966
##   TRUE   1427 8661
```

2.6   Association of QC missingness with Propensity weighting.

Gard (2020) notes that analyses using propensity weighting can be biased when missing data is correlated with population weights. They present data that association between missingness and population weights are small when considering structural and task-based brain data from ABCD. Here, we repeat these analyses considering rs-fMRI data in each condition.

Gard et al. appear to have used Welch's two-sample t-tests to examine differences in propensity weight between included and excluded groups, and we will do the same here.

```
gard_weighting_test<-function(boolvar) {
  formulastring=paste("acs_raked_propensity ~",boolvar)
  res<-broom::tidy(t.test(as.formula(formulastring), data=df, na.action="na.fail", conf.level=0.95))
  names(res)<-c("delta","Mexcluded","Mincluded","t","p","df","ci.low","ci.high","method","alternative")
  res %>% select(delta:ci.high)
}

weight_ttests<-reduce(lapply(bools[-1],gard_weighting_test),rbind)
weight_ttests$d<-weight_ttests$delta/sd(df$acs_raked_propensity)
as_hux(weight_ttests) %>%
  insert_column(c("",levelsdf$Level[-1]),after=0) %>%
  set_caption("Propensity weights of included vs. excluded participants by condition; t-test parameters")
```

Table 12: Propensity weights of included vs. excluded participants by condition; t-test parameters

	delta	Mexcluded	Mincluded	t	p	df	ci.low	ci.high	d
ABCD 4 Tabulated (T)	-0.487	691	691	-0.0309	0.975	568	-31.5	30.5	-0.00139
ABCC (C)	-3.68	688	692	-0.472	0.637	3.64e+03	-19	11.6	-0.0105
ABCC < 0.5mm	-4.63	688	692	-0.617	0.537	4.33e+03	-19.3	10.1	-0.0132
ABCC < 0.4mm	-3.49	689	692	-0.467	0.64	4.46e+03	-18.1	11.2	-0.00994
ABCC < 0.3mm	-2.93	689	692	-0.4	0.689	4.89e+03	-17.3	11.4	-0.00835
ABCD 4 Recommended (R)	22.4	709	687	2.72	0.00652	3.38e+03	6.25	38.5	0.0637
ABCD < 0.2mm	8.1	697	689	1.16	0.248	6.44e+03	-5.63	21.8	0.0231
ABCD < 0.1mm	41.6	710	668	6.42	1.45e-10	1.12e+04	28.9	54.3	0.119

2.7 Association of Exclusion with Expanded Demographic Variables

Available information on household income and (especially) race/ethnicity was greatly simplified in the main analysis for the purpose of brevity and comparibility with existing literature. For the sake of completeness and inclusivity, we present descriptive tables and bivariate models for the more granular variables here:

2.7.1 Household Income

```
#Reorder the factor in increasing order and relabel for descriptives.
df$household.income.full<-factor(df$household.income.full,
                    levels=c("Less than $5,000",
                             "$5,000 through $11,999",
                             "$12,000 through $15,999",
                             "$16,000 through $24,999",
                             "$25,000 through $34,999",
                             "$35,000 through $49,999",
                             "$50,000 through $74,999",
                             "$75,000 through $99,999",
                             "$100,000 through $199,999",
                             "$200,000 and greater"),
                    labels=c("Less than \\$5k",
                             "\\$5k through \\$11.9k",
                             "\\$12k through \\$15.9k",
                             "\\$16k through \\$24.9k",
                             "\\$25k through \\$34.9k",
```

```r
                                "\\$35k through \\$49.9k",
                                "\\$50k through \\$74.9k",
                                "\\$75k through \\$99.9k",
                                "\\$100k through \\$199.9k",
                                "\\$200k and greater"))

FullIncomeTableTable <- cbind(
  catTableAtThresh(df, rep(TRUE,nrow(df)),vars="household.income.full"),
  catTableAtThresh(df, df$ABCD_rsfmri_QC1,vars="household.income.full"),
  catTableAtThresh(df, df$any_rest,vars="household.income.full"),
  catTableAtThresh(df, df$gframes_0.5_375,vars="household.income.full"),
  catTableAtThresh(df, df$gframes_0.4_375,vars="household.income.full"),
  catTableAtThresh(df, df$gframes_0.3_375,vars="household.income.full"),
  catTableAtThresh(df, df$ABCD_rsfmri_QC2,vars="household.income.full"),
  catTableAtThresh(df, df$gframes_0.2_375,vars="household.income.full"),
  catTableAtThresh(df, df$gframes_0.1_375,vars="household.income.full")
)


#Try to draw as a huxtable
#rownames(FullIncomeTableTable)<-gsub("\\$","\\\\$",rownames(FullIncomeTableTable))
FullIncomeTableHux<-as_hux(FullIncomeTableTable) %>%
  theme_basic() %>%
  insert_row(value=c("n","\\%",rep(c("n","\\%"),8))) %>%
  insert_row("Full","",
             "QC1","",
             "ABCC","",
             ".5","",
             ".4","",
             ".3mm","",
             "QC2","",
             ".2mm","",
             ".1mm","") %>%
  merge_cells(1,1:2) %>%
  merge_cells(1,3:4) %>%
  merge_cells(1,5:6) %>%
  merge_cells(1,7:8) %>%
  merge_cells(1,9:10) %>%
  merge_cells(1,11:12) %>%
  merge_cells(1,13:14) %>%
  merge_cells(1,15:16) %>%
  merge_cells(1,17:18) %>%
  insert_column(value=c("","",rownames(FullIncomeTableTable) %>% replace_na("Missing"))) %>%
  set_caption("Granular Household Income at Each Level of Stringency (Subjects excluded with < 375 Frames)") %>%
  set_escape_contents(FALSE)  %>%
```

```
set_font_size(8) %>%
set_number_format(3:nrow(.),2+0:7*2,value=fmt_pretty()) %>%
set_number_format(3:nrow(.),3+0:7*2,value="%.2g") %>%
set_all_padding(1)

FullIncomeTableHux
```

Table 13: Granular Household Income at Each Level of Stringency (Subjects excluded with < 375 Frames)

	Full		QC1		ABCC		0.5		0.4		0.3mm		QC2		0.2mm		0.1mm	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Total	11,876	1e+02	11,355	96	9,600	81	9,320	78	9,262	78	9,098	77	9,627	81	8,507	72	5.25e+03	44.2
Less than $5k	417	3.5	379	3.3	289	3	279	3	273	2.9	266	2.9	280	2.9	228	2.7	125	2.38
$5k through $11.9k	421	3.5	386	3.4	311	3.2	287	3.1	286	3.1	283	3.1	293	3	252	3	130	2.48
$12k through $15.9k	273	2.3	266	2.3	213	2.2	200	2.1	199	2.1	196	2.2	220	2.3	184	2.2	102	1.94
$16k through $24.9k	524	4.4	495	4.4	393	4.1	373	4	368	4	361	4	395	4.1	324	3.8	181	3.45
$25k through $34.9k	654	5.5	628	5.5	521	5.4	503	5.4	499	5.4	491	5.4	537	5.6	459	5.4	269	5.12
$35k through $49.9k	934	7.9	896	7.9	770	8	747	8	740	8	721	7.9	740	7.7	676	7.9	407	7.75
$50k through $74.9k	1,499	13	1,441	13	1,254	13	1,215	13	1,208	13	1,189	13	1,204	13	1,106	13	630	12
$75k through $99.9k	1,572	13	1,508	13	1,316	14	1,281	14	1,276	14	1,250	14	1,315	14	1,188	14	773	14.7
$100k through $200k	3,314	28	3,192	28	2,735	28	2,673	29	2,660	29	2,623	29	2,785	29	2,470	29	1.62e+03	30.9
$200k and greater	1,250	11	1,206	11	1,014	11	995	11	992	11	978	11	1,056	11	930	11	626	11.9
Missing	1,018	8.6	958	8.4	784	8.2	767	8.2	761	8.2	740	8.1	802	8.3	690	8.1	385	7.33

```
#For models, place the modal level first and reformat labels.
df$household.income.full.ref<-relevel(df$household.income.full,9)
df$household.income.full.ref <- factor(df$household.income.full.ref,
                              labels=c(
                 "\\$100k through \\newline \\$199.9k",
                 "Less than \\newline \\$5k",
                 "\\$5k through \\newline \\$11.9k",
                 "\\$12k through \\newline \\$15.9k",
                 "\\$16k through \\newline \\$24.9k",
                 "\\$25k through \\newline \\$34.9k",
                 "\\$35k through \\newline \\$49.9k",
                 "\\$50k through \\newline \\$74.9k",
                 "\\$75k through \\newline \\$99.9k",
                 "\\$200k and greater"
                 ))
```

```r
m.bvincome<-bvmodel("household.income.full")
#Reformat the table:
# Replace the first \n with " ["
m.bvincome.f <- sub("\n", " [", m.bvincome, fixed = TRUE)
# Replace the second \n with "]"
m.bvincome.f <- sub("\n", "]", m.bvincome.f, fixed = TRUE)
# Remove the p value
m.bvincome.f <- gsub("].*","]", m.bvincome.f)


as_hux(m.bvincome.f) %>%
  insert_column(levels(df$household.income.full)[-1]) %>%
  insert_row(
    "Household Income (ref: \\$100-\\$199.9k)", rep(c("OR [95\\% CI]"), 8)) %>%
  insert_row(c("", "T", "C", ".5", ".4", ".3", "R", ".2", ".1")) %>%
  set_caption("Bivariate Model of Granular Household Income and Missingness by Condition. Ref: \\$100k-\\$199k") %>%
  set_escape_contents(value=FALSE) %>%
  set_wrap(TRUE) %>%
  set_width(1)
```

2.7.2 Detailed Race/Ethnicity

Census Race/Ethnicity Categories were used in the manuscript for the sake of brevity and because they are conventional and easy to compare to other work (which is of primary importance in a work commenting on methods). However, we acknowledge that the census categories are reductive both in the number of identifications present and in the lack of nuance they present (i.e., inability to pick multiple categories). ABCD collects information on participant race/ethnicity in considerably more detail, specifically as a series of yes/no questions about particular identities. We here present a comparison tables between the two measures, descriptives of the more granular race/ethnicity data, and bivariate models using those variables.

2.7.2.1 Descriptives This table presents a comparison of the census race/ethnicity measure included in ABCD and the extended race/ethnicity coding.

```r
retable1<-df %>% select(race_ethnicity.factor:ethn.iden.hisp) %>%
  group_by(race_ethnicity.factor) %>%
  summarise(across(
    race.iden.white:ethn.iden.hisp,
    ~ mean(. == TRUE, na.rm = TRUE),
    .names = "percentage_{.col}"
  )) %>% as_hux() %>% set_number_format(2:nrow(.),2:ncol(.),value=fmt_percent()) %>%
  t() %>% set_caption("Percentage Endorsing Granular Race/Ethnicity Variables within Census Race/Ethnicity Groups")

retable1[,1] <- c("Census Race/Ethnicity",
                  "White",
                  "Black",
```

Table 14: Bivariate Model of Granular Household Income and Missingness by Condition. Ref: $100k-$199k

	T	C	0.5	0.4	0.3	R	0.2	0.1
Household Income (ref: $100-$200k)	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]
$5k through $11.9k	.90 [.56-1.46]	.80 [.59-1.08]	.94 [.71-1.26]	.89 [.67-1.19]	.86 [.65-1.14]	.89 [.67-1.19]	.81 [.61-1.06]	.96 [.71-1.29]
$12k through $15.9k	.26** [.11-.56]	.64* [.44-.90]	.74 [.53-1.03]	.70* [.50-.98]	.69* [.50-.96]	.49*** [.34-.70]	.58*** [.42-.80]	.72* [.52-.99]
$16k through $24.9k	.58* [.35-.96]	.75 [.56-1.00]	.82 [.62-1.08]	.80 [.61-1.06]	.80 [.61-1.04]	.67** [.50-.89]	.74* [.57-.97]	.81 [.61-1.07]
$25k through $34.9k	.41*** [.24-.69]	.58*** [.43-.76]	.61*** [.46-.80]	.59*** [.45-.77]	.58*** [.45-.76]	.45*** [.33-.59]	.51*** [.40-.66]	.61*** [.47-.79]
$35k through $49.9k	.42*** [.27-.67]	.48*** [.37-.63]	.51*** [.39-.66]	.50*** [.38-.64]	.52*** [.40-.67]	.54*** [.41-.69]	.46*** [.36-.59]	.55*** [.43-.71]
$50k through $74.9k	.40*** [.26-.62]	.44*** [.34-.57]	.47*** [.37-.60]	.46*** [.36-.58]	.46*** [.36-.58]	.50*** [.39-.64]	.43*** [.34-.54]	.59*** [.47-.74]
$75k through $99.9k	.42*** [.28-.65]	.44*** [.34-.56]	.46*** [.36-.58]	.44*** [.35-.56]	.45*** [.36-.57]	.40*** [.31-.51]	.39*** [.31-.49]	.44*** [.35-.56]
$100k through $200k	.38*** [.26-.56]	.48*** [.38-.60]	.48*** [.39-.61]	.47*** [.37-.58]	.46*** [.37-.58]	.39*** [.31-.49]	.41*** [.33-.51]	.45*** [.36-.56]
$200k and greater	.36*** [.23-.57]	.53*** [.41-.68]	.52*** [.41-.66]	.49*** [.39-.63]	.49*** [.39-.62]	.38*** [.29-.49]	.42*** [.33-.52]	.43*** [.34-.54]

```
                "American Indian",
                "Alaska Native",
                "Native Hawaiian",
                "Guamanian",
                "Samoan",
                "Pacific Islander",
                "Asian Indian",
                "Filipino",
                "Chinese",
                "Japanese",
                "Korean",
                "Vietnamese",
                "Other Asian",
                "Other",
                "Refuse",
                "Don't Know",
                "Hispanic")

retable1[,1:6]
```

Next, here are the count and percentages of each of the granular variables, by QC condition:

```
boolTable<-\(data=df,bool=TRUE) {
  df[bool,] %>% summarise(across(race.iden.white:ethn.iden.hisp,
             list(n= ~ sum(.,na.rm = TRUE),
                  Percent = ~ round(mean(. == 1, na.rm = TRUE)*100,2)
                  ))) %>%
    pivot_longer(cols=everything(),
                 names_to=c(".value","measure"),
                 names_sep="_") %>% t()
}

retable2 <- cbind(
  boolTable(bool = TRUE),
  boolTable(bool = df$ABCD_rsfmri_QC1),
  boolTable(bool = df$any_rest),
  boolTable(bool = df$gframes_0.5_375),
  boolTable(bool = df$gframes_0.4_375),
  boolTable(bool = df$gframes_0.3_375),
  boolTable(bool = df$ABCD_rsfmri_QC2),
  boolTable(bool = df$gframes_0.2_375),
  boolTable(bool = df$gframes_0.1_375))[-1,]

retable2 %>%  as_hux() %>%
```

Table 15: Percentage Endorsing Granular Race/Ethnicity Variables within Census Race/Ethnicity Groups

Census Race/Ethnicity	White	Black	Hispanic	Asian	Other
White	99.9%	0.0%	68.0%	0.0%	79.3%
Black	0.0%	99.8%	8.9%	0.0%	41.9%
American Indian	0.0%	0.0%	4.7%	0.0%	23.4%
Alaska Native	0.0%	0.0%	0.1%	0.0%	0.2%
Native Hawaiian	0.0%	0.0%	0.2%	0.0%	1.4%
Guamanian	0.0%	0.0%	0.0%	0.0%	0.2%
Samoan	0.0%	0.0%	0.0%	0.0%	0.9%
Pacific Islander	0.0%	0.0%	0.3%	0.0%	2.6%
Asian Indian	0.0%	0.0%	0.2%	21.0%	4.4%
Filipino	0.0%	0.0%	0.7%	34.1%	8.1%
Chinese	0.0%	0.0%	1.5%	17.1%	7.1%
Japanese	0.0%	0.0%	0.2%	6.0%	5.1%
Korean	0.0%	0.0%	0.3%	10.3%	5.3%
Vietnamese	0.0%	0.0%	0.2%	10.3%	2.6%
Other Asian	0.0%	0.0%	0.4%	13.5%	3.7%
Other	0.0%	0.0%	27.2%	0.0%	11.5%
Refuse	0.0%	0.0%	1.2%	0.0%	2.5%
Don't Know	0.0%	0.0%	3.7%	0.0%	1.2%
Hispanic	0.0%	0.0%	100.0%	0.0%	0.0%

```r
  set_number_format(1:nrow(.),seq(1,ncol(.),2),value=fmt_pretty()) %>%
  set_number_format(1:nrow(.),seq(2,ncol(.),2),value=fmt_pretty(digits=2,format="f")) %>%
  insert_column(c("White",
                  "Black",
                  "American Indian",
                  "Alaska Native",
                  "Native Hawaiian",
                  "Guamanian",
                  "Samoan",
                  "Pacific Islander",
                  "Asian Indian",
                  "Filipino",
                  "Chinese",
                  "Japanese",
                  "Korean",
                  "Vietnamese",
                  "Other Asian",
                  "Other",
                  "Refuse",
                  "Don't Know",
                  "Hispanic")) %>%
  insert_row(c("",rep(c("n","%"),length(bools)))) %>%
  insert_row(c("",
               "F","",
               "QC1","",
               "C","",
               ".5","",
               ".4","",
               ".3","",
               "QC2","",
               ".2","",
               ".1","")) %>%
  set_caption("Granular Race/Ethnicity Data by Condition")
```

2.7.2.2 Bivariate Models Finally, we present bivariate models of the granular binary race/ethnicity coding variables:

```r
remodels<-lapply(names(df %>% select(race.iden.white:ethn.iden.hisp)),bvmodel,newline="\\newline ")

remodeltable<-as_hux(reduce(remodels,rbind)) %>%
  insert_column(c("White",
                  "Black",
                  "American Indian",
                  "Alaska Native",
```

Table 16: Granular Race/Ethnicity Data by Condition

	F		QC1		C		0.5		0.4		0.3		QC2		0.2		0.1	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
White	8,804	74	8,461	75	7,320	76	7,127	76	7,083	76	6,977	77	7,300	76	6,557	77	4,211	80
Black	2,518	21	2,375	21	1,918	20	1,840	20	1,825	20	1,785	20	1,903	20	1,631	19	870	17
American Indian	406	3.4	387	3.4	326	3.4	314	3.4	310	3.4	303	3.3	320	3.3	289	3.4	182	3.5
Alaska Native	5	0.04	3	0.03	2	0.02	1	0.01	1	0.01	1	0.01	2	0.02	1	0.01	1	0.02
Native Hawaiian	23	0.19	22	0.19	20	0.21	20	0.21	20	0.22	20	0.22	19	0.2	18	0.21	6	0.11
Guamanian	2	0.02	2	0.02	2	0.02	2	0.02	2	0.02	2	0.02	0	0	2	0.02	1	0.02
Samoan	12	0.1	12	0.11	11	0.11	11	0.12	11	0.12	11	0.12	12	0.12	11	0.13	7	0.13
Pacific Islander	40	0.34	36	0.32	31	0.32	30	0.32	30	0.32	30	0.33	30	0.31	27	0.32	17	0.32
Asian Indian	114	0.96	108	0.95	92	0.96	88	0.94	88	0.95	86	0.95	90	0.93	79	0.93	50	0.95
Filipino	203	1.7	191	1.7	152	1.6	150	1.6	149	1.6	146	1.6	158	1.6	135	1.6	95	1.8
Chinese	167	1.4	157	1.4	124	1.3	121	1.3	121	1.3	119	1.3	135	1.4	116	1.4	63	1.2
Japanese	84	0.71	82	0.72	60	0.62	59	0.63	59	0.64	57	0.63	72	0.75	54	0.63	41	0.78
Korean	100	0.84	98	0.86	80	0.83	78	0.84	78	0.84	76	0.84	85	0.88	70	0.82	48	0.91
Vietnamese	63	0.53	60	0.53	47	0.49	45	0.48	44	0.48	43	0.47	50	0.52	42	0.49	28	0.53
Other Asian	90	0.76	86	0.76	74	0.77	69	0.74	69	0.74	67	0.74	69	0.72	62	0.73	39	0.74
Other	800	6.7	761	6.7	581	6	567	6.1	561	6.1	549	6	645	6.7	513	6	292	5.6
Refuse	59	0.5	56	0.49	41	0.43	39	0.42	39	0.42	37	0.41	44	0.46	33	0.39	16	0.3
Don't Know	104	0.88	97	0.85	77	0.8	75	0.8	75	0.81	74	0.81	84	0.87	72	0.85	38	0.72
Hispanic	2,411	21	2,312	21	1,845	19	1,799	20	1,788	20	1,746	19	1,966	21	1,630	19	960	18

```
                "Native Hawaiian",
                "Guamanian",
                "Samoan",
                "Pacific Islander",
                "Asian Indian",
                "Filipino",
                "Chinese",
                "Japanese",
                "Korean",
                "Vietnamese",
                "Other Asian",
                "Other",
                "Refuse",
                "Don't Know",
                "Hispanic")) %>%
  insert_row(
    "", rep(c("OR\\newline [90\\% CI]\\newline p"), 8)) %>%
  insert_row(c("", "T", "C", ".5", ".4", ".3", "R", ".2", ".1")) %>%
  set_escape_contents(value=FALSE)  %>%
  set_wrap(TRUE) %>%
  set_width(1)


remodeltable[c(1:2,3:8),] %>% set_caption("Granular Race/Ethnicity Variables: Bivariate Models, Pt. 1") %>% set_label("gran.race.ethn.1")


remodeltable[c(1:2,9:14),] %>% set_caption("Granular Race/Ethnicity Variables: Bivariate Models, Pt. 2") %>% set_label("gran.race.ethn.2")


remodeltable[c(1:2,15:19),] %>% set_caption("Granular Race/Ethnicity Variables: Bivariate Models, Pt. 3") %>% set_label("gran.race.ethn.3")
```

3   Changes since pre-registration

The following changes were made to the analysis plan post-registration:

- BMI was added as a study variable, to enhance comparibility with Cosgrove et. al. (2020) and reflect the known relation between BMI and in-scanner motion.
- The Behavioral Inhibition scale was removed. It is not available in the ABCD baseline data and was included in error.
- More data was excluded prior to motion filtering than expected and there were more differences in inclusion criteria between ABCD versions than expected. Consequently, three 'QC' conditions were added in addition to the motion thresholds, to provide additional detail to inform study design.
- The originally planned approach to evaluate H2 was found to be infeasible (See 'H2 Model,' above). In practice, a visual inspection of the data did not support H2 (see H2 Marginal Means Plotting). Specifically, because so much data was missing, biases in the missing data were self-correcting as more data was excluded. I.e., in the event that males were more likely to be excluded than females (which appears to be true in this dataset), the bias is strongest as data is first excluded

Table 17: Granular Race/Ethnicity Variables: Bivariate Models, Pt. 1

	T	C	0.5	0.4	0.3	R	0.2	0.1
	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p
White	.66*** .55-.80 <.001	.58*** .53-.64 <.001	.59*** .53-.65 <.001	.59*** .54-.65 <.001	.58*** .53-.64 <.001	.64*** .58-.71 <.001	.60*** .55-.65 <.001	.56*** .51-.61 <.001
Black	1.43*** 1.17-1.74 <.001	1.43*** 1.29-1.59 <.001	1.47*** 1.33-1.62 <.001	1.47*** 1.33-1.63 <.001	1.47*** 1.33-1.62 <.001	1.53*** 1.37-1.70 <.001	1.51*** 1.37-1.65 <.001	1.67*** 1.52-1.83 <.001
American Indian	1.07 .65-1.67 .769	1.04 .80-1.32 .779	1.07 .84-1.35 .570	1.10 .87-1.38 .419	1.12 .89-1.40 .338	1.16 .90-1.47 .240	1.02 .82-1.27 .838	.97 .80-1.19 .798
Alaska Native	14.58** 1.92-88.19 .003	6.33* 1.05-48.11 .043	14.61* 2.16-285.80 .016	14.19* 2.10-277.71 .018	13.12* 1.94-256.66 .021	6.43* 1.06-48.83 .042	10.11* 1.50-197.83 .039	3.17 .47-62.03 .302
Native Hawaiian	.99 .06-4.73 .993	.63 .15-1.85 .459	.55 .13-1.60 .329	.53 .13-1.55 .307	.49 .12-1.43 .250	.90 .26-2.40 .850	.70 .23-1.76 .483	2.25 .93-6.24 .088
Guamanian	.00 NA-3.46e+11 .967	.00 NA-125672.14 .942	.00 NA-108642.02 .941	.00 NA-105570.45 .941	.00 NA-97579.29 .941	451815.50 .00-NA .926	.00 NA-75234.92 .939	.79 .03-20.04 .869

Table 18: Granular Race/Ethnicity Variables: Bivariate Models, Pt. 2

	T	C	0.5	0.4	0.3	R	0.2	0.1
	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p
Samoan	.00 NA-835.53 .964	.38 .02-1.97 .358	.33 .02-1.70 .290	.32 .02-1.66 .278	.30 .02-1.53 .246	.00 NA-.34 .938	.23 .01-1.18 .158	.57 .17-1.77 .331
Pacific Islander	2.43 .73-6.10 .093	1.23 .55-2.47 .592	1.22 .56-2.41 .592	1.18 .55-2.34 .648	1.09 .51-2.16 .810	1.43 .66-2.83 .329	1.22 .61-2.32 .562	1.07 .57-2.04 .828
Asian Indian	1.21 .47-2.54 .647	1.01 .62-1.58 .971	1.08 .68-1.65 .737	1.05 .66-1.60 .837	1.07 .68-1.62 .767	1.14 .71-1.77 .563	1.12 .74-1.66 .579	1.01 .70-1.48 .940
Filipino	1.38 .72-2.38 .287	1.42* 1.02-1.95 .030	1.29 .94-1.76 .110	1.29 .93-1.75 .112	1.28 .94-1.74 .112	1.22 .87-1.69 .237	1.28 .95-1.71 .103	.90 .68-1.19 .454
Chinese	1.40 .69-2.53 .311	1.47* 1.03-2.07 .030	1.39 .98-1.95 .058	1.35 .95-1.89 .083	1.33 .94-1.85 .101	1.01 .68-1.48 .941	1.11 .79-1.54 .531	1.31 .96-1.81 .090
Japanese	.53 .09-1.68 .376	1.69* 1.03-2.69 .030	1.55 .95-2.45 .067	1.51 .93-2.38 .087	1.56 .97-2.44 .059	.71 .37-1.26 .277	1.41 .89-2.18 .136	.83 .54-1.28 .395

Table 19: Granular Race/Ethnicity Variables: Bivariate Models, Pt. 3

	T	C	0.5	0.4	0.3	R	0.2	0.1
	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p	OR [90% CI] p
Korean	.44 .07-1.40 .255	1.05 .63-1.69 .831	1.03 .62-1.62 .907	1.00 .61-1.58 .998	1.03 .64-1.61 .885	.75 .42-1.27 .314	1.08 .70-1.65 .716	.86 .58-1.27 .444
Vietnamese	1.09 .27-2.95 .884	1.44 .79-2.49 .210	1.46 .82-2.48 .175	1.53 .87-2.59 .120	1.53 .88-2.56 .119	1.11 .58-1.99 .730	1.26 .73-2.11 .382	.99 .60-1.64 .970
Other Asian	1.01 .31-2.44 .979	.91 .51-1.52 .737	1.11 .66-1.78 .675	1.08 .65-1.73 .761	1.13 .69-1.78 .627	1.31 .78-2.09 .287	1.14 .72-1.77 .563	1.04 .68-1.58 .867
Other	1.13 .79-1.55 .486	1.65*** 1.40-1.94 <.001	1.55*** 1.32-1.81 <.001	1.56*** 1.33-1.83 <.001	1.55*** 1.32-1.81 <.001	1.03 .86-1.23 .744	1.45*** 1.25-1.69 <.001	1.41*** 1.22-1.64 <.001
Refuse	1.17 .28-3.18 .793	1.86* 1.04-3.19 .029	1.88* 1.07-3.18 .023	1.82* 1.04-3.09 .029	1.95* 1.13-3.29 .013	1.46 .79-2.57 .205	2.00** 1.18-3.34 .009	2.14** 1.23-3.91 .010

from the dataset. Once a significantly larger proportion of males have been excluded than females, the over-representation of females in the sample results in their being more likely to be excluded. At the extreme, when all of the data is excluded, there is no bias. This is reflected in our marginal means plots).

Here, we present the analyses originally proposed as written, excepting H2. Specifically, we present results from adjusted models without the BMI variable in the motion scrubbing conditions.

```
model.h1prereg.5mm<-glm(!gframes_0.5_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z +
model.h1prereg.5mm.results<-cbind(confint(model.h1prereg.5mm,level=.95),model.h1prereg.5mm$coefficients)[,c(1,3,2)]
model.h1prereg.5mm.results<-cbind(model.h1prereg.5mm.results,pcorrect(coef(summary(model.h1prereg.5mm))[,4]))

model.h1prereg.4mm<-glm(
  !gframes_0.4_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecor
  data=df,
  family="binomial")
model.h1prereg.4mm.results<-cbind(confint(model.h1prereg.4mm,level=.95),model.h1prereg.4mm$coefficients)[,c(1,3,2)]
model.h1prereg.4mm.results<-cbind(model.h1prereg.4mm.results,pcorrect(coef(summary(model.h1prereg.4mm))[,4]))

model.h1prereg.3mm<-glm(
  !gframes_0.3_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecor
  data=df,
  family="binomial")
model.h1prereg.3mm.results<-cbind(confint(model.h1prereg.3mm,level=.95),model.h1prereg.3mm$coefficients)[,c(1,3,2)]
model.h1prereg.3mm.results<-cbind(model.h1prereg.3mm.results,pcorrect(coef(summary(model.h1prereg.3mm))[,4]))

model.h1prereg.2mm<-glm(
  !gframes_0.2_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecor
  data=df,
  family="binomial")
model.h1prereg.2mm.results<-cbind(confint(model.h1prereg.2mm,level=.95),model.h1prereg.2mm$coefficients)[,c(1,3,2)]
model.h1prereg.2mm.results<-cbind(model.h1prereg.2mm.results,pcorrect(coef(summary(model.h1prereg.2mm))[,4]))

model.h1prereg.1mm<-glm(
  !gframes_0.1_375 ~ sex + household.income + p.edu  + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_ageco
  data=df,
  family="binomial")
model.h1prereg.1mm.results<-cbind(confint(model.h1prereg.1mm,level=.95),model.h1prereg.1mm$coefficients)[,c(1,3,2)]
model.h1prereg.1mm.results<-cbind(model.h1prereg.1mm.results,pcorrect(coef(summary(model.h1prereg.1mm))[,4]))
```

3.0.1 Pre-registered Condition (Adjusted) Model Tables

```
Mformat2<-\(m) {
  stars<-sdlabFunctions::starPs(m[,4])
  paste0(ORformat(m[,2]),
```

```
        stars,
        " [",ORformat(m[,1]),"-",ORformat(m[,3]),"]")
}


h1modelspreregtable<-cbind(
                    Mformat2(model.h1prereg.5mm.results),
                    Mformat2(model.h1prereg.4mm.results),
                    Mformat2(model.h1prereg.3mm.results),
                    Mformat2(model.h1prereg.2mm.results),
                    Mformat2(model.h1prereg.1mm.results))

#h1modelspreregtable <- gsub("\n", "\\newline", h1modelspreregtable)

h1modelspreregtable[h1modelspreregtable==".000"]<-"<.001"

alllevels<-c("Intercept",
      "Sex (Male)",
      "  $0-$25k",
      "  $25-$50k",
      "  $50-$75k",
      "  $75-$100k",
      "  >$200k",
      "  <High School",
      "  HS Grad.",
      "  Some College",
      "  Graduate",
      "  Black",
      "  Hispanic",
      "  Asian",
      "  Other",
      "  1 Trauma",
      "  >=2 Trauma",
      "  Early Puberty",
      "  Mid Puberty",
      "  Late Puberty",
      "  Post Pubertal",
      "Area Disadvantage",
      "Child Opportunity",
      "NIHTB Flanker",
      "NIHTB Crystalized",
      "NIHTB Total",
      "WISC V Matrix",
      "Psychopathology",
      "Internalizing",
```

```
      "Externalizing",
      "Age"
      )


h1prereghux<-as_hux(h1modelspreregtable) %>%
  insert_column(alllevels) %>%
  insert_row("Household Income (ref: $100-$200k)",after=2,colspan=6,fill="") %>%
  insert_row("Highest Parental Education (ref: College Degree)",after=8,colspan=6,fill="") %>%
  insert_row("Census Race/Ethnicity (ref: White)",after=13,colspan=6,fill="") %>%
  insert_row("KSADS Trauma Count (ref: 0 Exposures)",after=18,colspan=6,fill="") %>%
  insert_row("Pubertal Status (ref: pre-pubertal)",after=21,colspan=6,fill="") %>%
  insert_row(c("Variable",rep(c("OR [95% CI]"),5)),after=0) %>%
  insert_row(c("",".5",".4",".3",".2",".1")) %>%
  set_wrap(TRUE) %>%
  set_width(1) %>%
  set_font_size(value=8)


colnames(h1prereghux)[1]<-"h1"


h1prereghux[c(1:2,3:20),] %>%
  set_caption("Adjusted Models Output, as Pre-registered (Pt. 1). ***: p<.001; **: p<.01; *:p<.05") %>% set_label("PreRegModel1")



h1prereghux[c(1:2,21:38),] %>%
  set_caption("Adjusted Models Output, as Pre-registered (Pt. 2). ***: p<.001; **: p<.01; *:p<.05") %>% set_label("PreRegModel2")
```

4   Save Tables

```
openxlsx::saveWorkbook(tables.xlsx,"Manuscript/Tables/tables.xlsx",overwrite=TRUE)
```

5   SessionInfo

R version 4.3.3 (2024-02-29)

Platform: x86_64-pc-linux-gnu (64-bit)

locale: *LC_CTYPE=en_US.UTF-8,   LC_NUMERIC=C,   LC_TIME=en_US.UTF-8,   LC_COLLATE=en_US.UTF-8,   LC_MONETARY=en_US.UTF-8,*
LC_MESSAGES=en_US.UTF-8, LC_PAPER=en_US.UTF-8, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8
and *LC_IDENTIFICATION=C*

attached base packages: *stats*, *graphics*, *grDevices*, *utils*, *datasets*, *methods* and *base*

Table 20: Adjusted Models Output, as Pre-registered (Pt. 1). ***: p<.001; **: p<.01; *:p<.05

	0.5	0.4	0.3	0.2	0.1
Variable	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]
Intercept	.18*** [.15-.22]	.18*** [.16-.22]	.20*** [.17-.23]	.25*** [.21-.29]	.80** [.70-.92]
Sex (Male)	1.09 [.96-1.23]	1.11 [.99-1.26]	1.17** [1.04-1.32]	1.31*** [1.17-1.46]	1.42*** [1.29-1.57]
Household Income (ref: $100-$200k)					
$0-$25k	1.15 [.93-1.44]	1.17 [.94-1.45]	1.11 [.90-1.38]	1.13 [.93-1.38]	1.06 [.88-1.28]
$25-$50k	.89 [.73-1.09]	.91 [.75-1.11]	.92 [.76-1.12]	.87 [.73-1.04]	.89 [.76-1.05]
$50-$75k	.92 [.76-1.10]	.92 [.76-1.10]	.92 [.77-1.10]	.93 [.79-1.10]	1.09 [.94-1.26]
$75-$100k	.96 [.80-1.14]	.96 [.81-1.13]	.98 [.83-1.16]	.94 [.80-1.09]	.92 [.80-1.05]
>$200k	.98 [.81-1.18]	.98 [.82-1.18]	1.00 [.83-1.19]	.98 [.82-1.15]	.96 [.82-1.11]
Highest Parental Education (ref: College Degree)					
<High School	1.49** [1.10-2.00]	1.50** [1.11-2.01]	1.44* [1.07-1.93]	1.49** [1.12-1.97]	1.35* [1.02-1.79]
HS Grad.	1.25 [.99-1.58]	1.24 [.99-1.56]	1.26* [1.01-1.58]	1.29* [1.04-1.59]	1.23* [1.01-1.50]
Some College	1.02 [.86-1.20]	1.03 [.87-1.21]	1.03 [.88-1.20]	1.10 [.95-1.27]	1.03 [.91-1.18]
Graduate	1.13 [.98-1.30]	1.12 [.98-1.29]	1.12 [.98-1.29]	1.14 [1.00-1.29]	1.04 [.93-1.17]
Census Race/Ethnicity (ref: White)					
Black	1.69*** [1.40-2.04]	1.63*** [1.35-1.97]	1.59*** [1.32-1.91]	1.53*** [1.29-1.82]	1.48*** [1.26-1.75]
Hispanic	1.20* [1.02-1.40]	1.19* [1.02-1.39]	1.21* [1.04-1.40]	1.14 [.99-1.32]	1.16* [1.02-1.32]
Asian	1.92*** [1.38-2.64]	1.88*** [1.35-2.58]	2.05*** [1.48-2.80]	1.96*** [1.44-2.66]	1.49** [1.10-2.01]
Other	1.34*** [1.13-1.59]	1.34*** [1.13-1.59]	1.28** [1.08-1.52]	1.18* [1.00-1.38]	1.12 [.97-1.29]

Table 21: Adjusted Models Output, as Pre-registered (Pt. 2). ***: p<.001; **: p<.01; *:p<.05

	0.5	0.4	0.3	0.2	0.1
Variable	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]	OR [95% CI]
KSADS Trauma Count (ref: 0 Exposures)					
1 Trauma	1.06 [.93-1.19]	1.05 [.93-1.19]	1.05 [.93-1.18]	1.03 [.92-1.15]	.96 [.87-1.06]
>=2 Trauma	1.12 [.93-1.34]	1.08 [.90-1.29]	1.06 [.89-1.26]	1.13 [.96-1.33]	1.20* [1.03-1.40]
Pubertal Status (ref: pre-pubertal)					
Early Puberty	1.02 [.90-1.17]	1.03 [.90-1.17]	1.01 [.89-1.15]	1.07 [.95-1.20]	1.07 [.96-1.19]
Mid Puberty	1.05 [.90-1.24]	1.06 [.90-1.24]	1.07 [.91-1.24]	1.08 [.94-1.25]	1.25*** [1.10-1.42]
Late Puberty	1.73** [1.16-2.55]	1.71** [1.15-2.52]	1.61* [1.08-2.38]	1.63* [1.11-2.37]	1.31 [.91-1.88]
Post Pubertal	.47 [.02-2.73]	.47 [.02-2.70]	.45 [.02-2.59]	1.55 [.31-6.50]	1.37 [.33-6.82]
Area Disadvantage	.70*** [.64-.77]	.72*** [.66-.78]	.73*** [.67-.79]	.76*** [.71-.83]	.86*** [.80-.93]
Child Opportunity	.78*** [.71-.86]	.80*** [.73-.88]	.80*** [.73-.87]	.82*** [.75-.90]	.84*** [.78-.92]
NIHTB Flanker	1.02 [.95-1.09]	1.01 [.95-1.09]	1.03 [.96-1.10]	1.02 [.96-1.09]	1.01 [.96-1.07]
NIHTB Crystalized	1.06 [.94-1.18]	1.05 [.94-1.17]	1.08 [.97-1.21]	1.09 [.99-1.21]	1.16** [1.06-1.27]
NIHTB Total	.88 [.77-1.00]	.88 [.78-1.00]	.84* [.74-.96]	.84** [.74-.94]	.81*** [.73-.90]
WISC V Matrix	1.00 [.95-1.06]	.99 [.93-1.05]	.98 [.93-1.04]	.95 [.90-1.00]	.95* [.90-.99]
Psychopathology	1.11*** [1.05-1.17]	1.11*** [1.05-1.17]	1.11*** [1.06-1.17]	1.13*** [1.07-1.18]	1.11*** [1.06-1.16]
Internalizing	.98 [.92-1.03]	.97 [.92-1.02]	.96 [.91-1.01]	.95 [.91-1.00]	.92*** [.88-.96]
Externalizing	.94* [.89-.99]	.94* [.89-.99]	.94* [.90-.99]	.94* [.89-.99]	.94** [.90-.98]
Age	.85*** [.81-.90]	.84*** [.80-.89]	.83*** [.79-.88]	.82*** [.78-.86]	.81*** [.78-.85]

other attached packages: *ggtext(v.0.1.2), emmeans(v.1.8.9), lmerTest(v.3.1-3), lme4(v.1.1-35.1), Matrix(v.1.6-5), ggupset(v.0.3.0), openxlsx(v.4.2.5.2), patchwork(v.1.1.3), viridis(v.0.6.4), viridisLite(v.0.4.2), huxtable(v.5.5.2), hexbin(v.1.28.3), pander(v.0.6.5), ggExtra(v.0.10.1), ggthemes(v.4.2.4), lubridate(v.1.9.3), forcats(v.1.0.0), stringr(v.1.5.0), dplyr(v.1.1.3), purrr(v.1.0.2), readr(v.2.1.4), tidyr(v.1.3.0), tibble(v.3.2.1), ggplot2(v.3.4.4), tidyverse(v.2.0.0)* and *rmarkdown(v.2.25)*

loaded via a namespace (and not attached): *gridExtra(v.2.3), gld(v.2.6.6), readxl(v.1.4.3), rlang(v.1.1.1), magrittr(v.2.0.3), e1071(v.1.7-13), compiler(v.4.3.3), vctrs(v.0.6.4), pkgconfig(v.2.0.3), crayon(v.1.5.2), fastmap(v.1.1.1), ellipsis(v.0.3.2), labeling(v.0.4.3), utf8(v.1.2.3), promises(v.1.2.1), tzdb(v.0.4.0), nloptr(v.2.0.3), xfun(v.0.43), later(v.1.3.1), DescTools(v.0.99.51), R6(v.2.5.1), stringi(v.1.7.12), boot(v.1.3-28), extrafontdb(v.1.0), cellranger(v.1.1.0), numDeriv(v.2016.8-1.1), estimability(v.1.4.1), Rcpp(v.1.0.11), assertthat(v.0.2.1), knitr(v.1.44), parameters(v.0.21.2), extrafont(v.0.19), httpuv(v.1.6.12), splines(v.4.3.3), timechange(v.0.2.0), tidyselect(v.1.2.0), rstudioapi(v.0.15.0), yaml(v.2.3.7), codetools(v.0.2-19), miniUI(v.0.1.1.1), lattice(v.0.22-4), shiny(v.1.7.5.1), withr(v.2.5.1), bayestestR(v.0.13.1), coda(v.0.19-4), evaluate(v.0.22), proxy(v.0.4-27), zip(v.2.3.0), xml2(v.1.3.5), pillar(v.1.9.0), insight(v.0.19.6), generics(v.0.1.3), sdlabFunctions(v.0.2.0), hms(v.1.1.3), munsell(v.0.5.0), scales(v.1.2.1), rootSolve(v.1.8.2.4), minqa(v.1.2.6), xtable(v.1.8-4), class(v.7.3-22), glue(v.1.6.2), lmom(v.3.0), tools(v.4.3.3), data.table(v.1.14.8), Exact(v.3.2), mvtnorm(v.1.2-3), grid(v.4.3.3), Rttf2pt1(v.1.3.12), datawizard(v.0.9.0), colorspace(v.2.1-0), nlme(v.3.1-163), performance(v.0.10.5), cli(v.3.6.1), fansi(v.1.0.5), expm(v.0.999-7), gtable(v.0.3.4), digest(v.0.6.33), farver(v.2.1.1), htmltools(v.0.5.6.1), lifecycle(v.1.0.3), httr(v.1.4.7), mime(v.0.12), gridtext(v.0.1.5)* and *MASS(v.7.3-60.0.1)*