

Supplement

Matthew Peverill

2023-11-17

Contents

1 Supplemental Methods

1.1	Coding and Releveling	
1.2	Descriptives	
1.2.1	Figure 1 Table 1: Inclusion counts and inconsistencies	
1.2.2	Categorical Variable Figure.	
1.2.3	Continuous Variables Figure	
1.3	Bivariate Models	
1.4	Condition (Adjusted) Models	
1.4.1	Adjusted Model Tables	
1.4.2	Site and MRI Effects – sensitivity analysis.	
1.5	Table: Associations of behavioral data with framewise displacement	
1.6	QCFC Figures	
1.7	Multiply at Risk Cell Counts	

2 Supplementary Results

2.1	Categorical Variable Table by Condition	
2.2	Alternative Continuous Variable Plots (psychopathology)	
2.3	Missingness in Behavioral Data	
2.4	Adjusted Models – Forest Plot	
2.5	Current Fast Track QC vs DAIRC recommendations	
2.6	Association of QC missingness with Propensity weighting.	

3

Changes since pre-registration

3.0.1

Pre-registered Condition (Adjusted) Model Tables

4

Save Tables

5

SessionInfo

List of Tables

1

OR of scanner terms, with and without site control [95% CI]

2

Non-white (census) with psychopathology at $z \geq 1.5$

3

Male participants with NIH toolbox total scores of $z \leq -1.5$

4

Categorical Values at Each Level of Stringency (Subjects excluded with < 375 Frames)

5

Categorical Values at Each Level of Stringency (Subjects excluded with < 375 frames) - omitted columns

6

Behavioral and Demographic variables with >50 missing values

7

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

8

Adjusted Models Output, as Pre-registered

1 Supplemental Methods

This section includes R code used to generate tables, figures, and statistics used in the main manuscript.

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

t1<-readRDS("RDS/tabulateddata.Rds")
t2<-readRDS("ABCC_download/ABCCsubjectdata.RDS")
df<-left_join(readRDS("RDS/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)) %>%
# 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("gframe"),\ (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))

#These Pt's have neuroimaging data but not tabular data:
# NDARINV*****
# NDARINV*****
# NDARINV*****
# NDARINV*****
# NDARINV*****
# NDARINV*****
# NDARINV*****
# NDARINV*****
# Possibly they have removed consent?

# Seed for random number generation
set.seed(42)
knitr::opts_chunk$set(cache.extra = knitr::rand_seed)

```

1.1 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"))

# Relevel factors 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

Upset figure:

```

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

upsetfig<-ggplot(setsdf,aes(x=sets)) +
  theme(axis.title.x = element_blank()) +
  ggtitle("B. Inclusion Inconsistencies ") +
  ylab("n") +
  geom_bar() +
  geom_text(stat='count',
           aes(label=after_stat(count)),

```

```

      vjust=-1,
      size=10 / .pt,
      family="Times New Roman") +
expand_limits(y=c(0,8600)) +
scale_x_upset()

```

Figure 1 panel a:

```

# 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*",
    "\"",
    "\"",
    "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$gframe5_375),
    sum(df$gframe4_375),
    sum(df$gframe3_375),
    sum(df$ABCD_rsfmri_QC2),
    sum(df$gframe2_375),
    sum(df$gframe1_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 < .5mm",
  "ABCC < .4mm",
  "ABCC < .3mm",
  "ABCD 4 Recommended (**R**)",
  "ABCD < .2mm",
  "ABCD < .1mm"
))

library(ggtext)
samplecounts<-ggplot(levelsdf,aes(x=Levelf,y=Count,label=Count)) +
  labs(title="A. Sample Key and Counts") +
  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=-1400,family="Times New Roman")

```

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")

```

```

ggsave("Manuscript/Figures/inclusioncounts.svg",
       (samplecounts+upsetfig),
       width=7.5,height=4)

```

1.2.2 Categorical Variable Figure.

```

bools<-c("full",
         "ABCD_rsfmri_QC1",
         "any_rest",
         "gframe5_375",
         "gframe4_375",
         "gframe3_375",
         "ABCD_rsfmri_QC2",
         "gframe2_375",
         "gframe1_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)
}

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:frame1_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))

cpl<-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)) +
  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(),  
          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())
```

```
svg(filename="Manuscript/Figures/CatVarFig.svg",height=9,width=7.5)
```

```
((cp1 / cp4 / cp3) |  
 (cp2 / cp6 / cp5))  
grid::grid.draw(grid::textGrob("Proportion at Threshold",x=.01,rot=90))  
grid::grid.draw(grid::textGrob("QC Threshold",y=.01))
```

```
dev.off()
```

```
## pdf
```

```
## 2
```



```
"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)
```

1.3 Bivariate Models

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

bvmodel <- \"(x) {
  cstrings <-
    c(
      "!ABCD_rsfmri_QC1",
      "!any_rest",
      "!gframe5_375",
      "!gframe4_375",
      "!gframe3_375",
      "!ABCD_rsfmri_QC2",
      "!gframe2_375",
      "!gframe1_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]),
              "\n",
              ORformat(table[,2]),
              "_",
              ORformat(table[,3]),
              "\n",
              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 Crystallized",
    "NIHTB Total",
    "WISC V Matrix",
    "Psychopathology",
    "Internalizing",
    "Externalizing",
    "Age",
    "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]\n"), 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")

```

1.4 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("!gframe5_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("!gframe4_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("!gframe3_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("!gframe2_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("!gframe1_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]))

```

1.4.1 Adjusted Model Tables

```

Mformat<-\ (m) {
  stars<-sdlabFunctions::starPs(m[,4])
  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 Crystallized",
"NIHTB Total",
"WISC V Matrix",
"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]\n"), 8)), after = 0) %>%
    insert_row(c("", "T", "C", ".5", ".4", ".3", "R", ".2", ".1"))

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

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")

```

1.4.2 Site and MRI Effects – sensitivity analysis.

Let's run some models that incorporate site and MRI model control. We can just run these at ABCC, QC2, 2mm

```

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,
  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,
  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(
  !gframe2_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected,
  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(
  !gframe2_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected,
  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,
  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,
  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,
  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(
  !gframe2_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrected,
  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,
  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"))

```

```

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"))

```

1.4.2.1 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]),
      "]" ,
      pformat(m[,4]))
  )
}

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,]))

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]")

```

Table 1: OR of scanner terms, with and without site control [95% CI]

	Without Site Control		
Scanner (ref: Philips Achieva)	ABCC		
GE Discovery MR750	.86 [.72–1.04].296	.56*** [.46–.69].000	.70*** [.59–.84].000
Philips Ingenia	.20*** [.13–.31].000	.87 [.63–1.19].620	.42*** [.30–.57].000
Siemens Prisma	.15*** [.12–.20].000	.27*** [.22–.33].000	.20*** [.17–.25].000
Siemens Prisma fit	.39*** [.32–.47].000	.43*** [.36–.53].000	.43*** [.36–.51].000
	With Site Control		
GE Discovery MR750	3894.69 [1.69e-14–NA].988	44437.50 [.00–NA].978	13726.89 [.00–NA].994
Philips Ingenia	8698.06 [9.76e-12–NA].988	7.24e+09 [.00–NA].978	9.30e+09 [4.22e+200–1.16e+207].994
Siemens Prisma	3109.47 [1.16e-14–NA].988	8613.49 [.00–NA].978	15297.36 [.00–NA].994
Siemens Prisma fit	6030.51 [2.24e-14–NA].988	8257.41 [.00–NA].978	16617.52 [.00–NA].994

1.4.2.2 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.

```

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() %>%
  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.compdb %>% group_by(term) %>% summarise(anysig = sum(sig) != 0) %>% filter(anysig))$term

ggplot(ABCC.compdb %>% filter(term %in% sigterms),
       aes(y = term, color = model, alpha = alphaval)) +
  geom_point(aes(x = OR), position = ggstance::position_dodgev(height =
                                                                .2)) +
  geom_linerange(aes(xmin = lowCI, xmax = highCI, y = term),
                position = ggstance::position_dodgev(height = .2)) +
  xlim(-4, 4) +
  geom_vline(lty = 2, aes(xintercept = 1), colour = 'black')

```

```

QC2.compdb <- 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"
    ) %>%

```

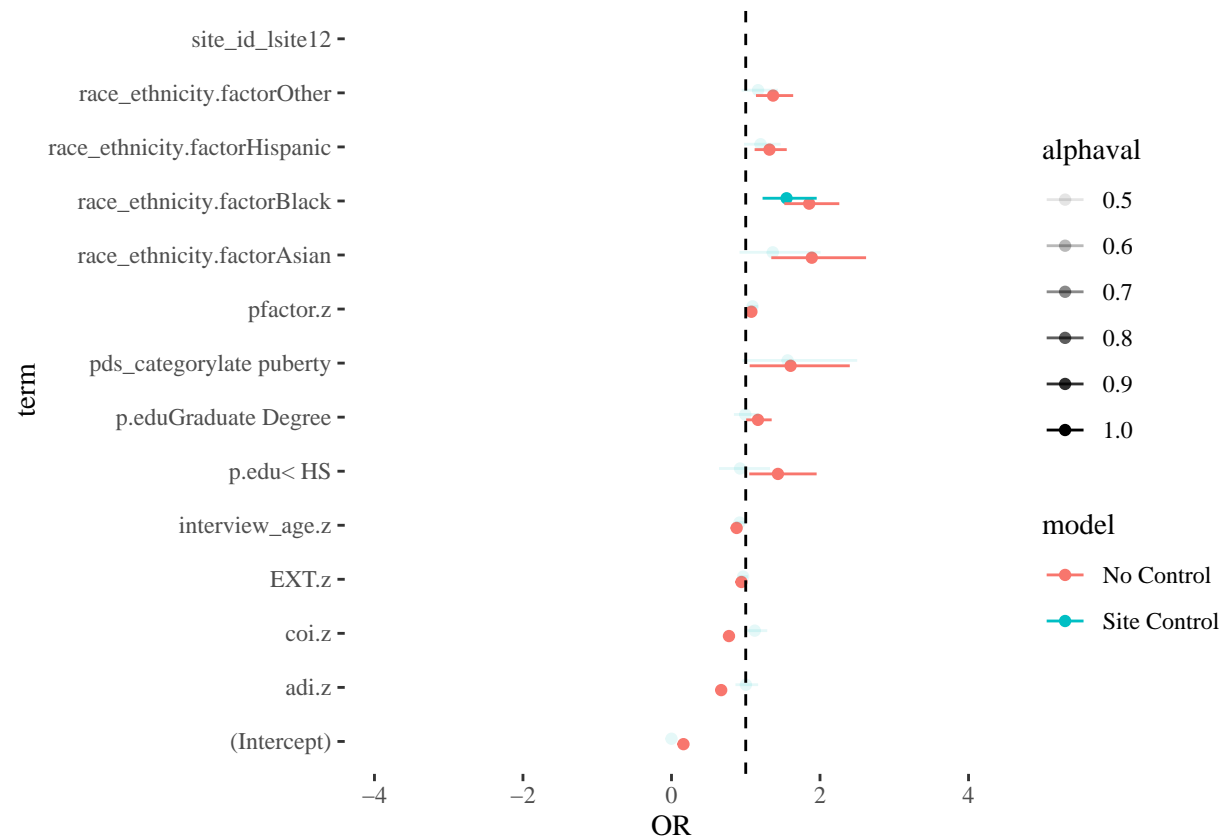



Figure 1: Significant ORs predicting ABCC missingness with and without site control

```

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

ggplot(QC2.compdf %>% filter(term %in% sigterms),aes(y=term,color=model,alpha=alphaval)) +
  geom_point(aes(x=OR),position=ggstance::position_dodgev(height=.2)) +
  geom_segment(aes(x=lowCI,xend=highCI,y=term,yend=term),position=ggstance::position_dodgev(height=.2)) +
  xlim(-3,3) +
  geom_vline(lty=2,aes(xintercept=1),colour='black')

```

```

twomm.compdf <- rbind(
  model.2mm.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.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.comppdf %>% group_by(term) %>% summarise(anysig = sum(sig) != 0) %>% filter(anysig))$term

ggplot(twomm.comppdf %>% filter(term %in% sigterms),
       aes(y = term, color = model, alpha = alphaval)) +
  geom_point(aes(x = OR), position = ggstance::position_dodgev(height =
                                                                .2)) +
  geom_segment(
    aes(
      x = lowCI,
      xend = highCI,
      y = term,
      yend = term
    ),
    position = ggstance::position_dodgev(height = .2)
  ) +
  xlim(-3, 3) +
  geom_vline(lty = 2, aes(xintercept = 1), colour = 'black')

```

```

sitesummary <- df %>% group_by(site_id_1) %>%
  summarise(
    "ExcludedPerc.QC2" = mean(!ABCD_rsfmri_QC2),
    "ExcludedPerc.ABCC" = mean(!any_rest),
    "ExcludedPerc.2mm" = mean(!gframe2_375),
    "ExcludedPerc.fastqc" = mean(!fastqcok)
  ) %>%
  arrange(site_id_1) %>%
  mutate(
    sitelabel = recode_factor(

```

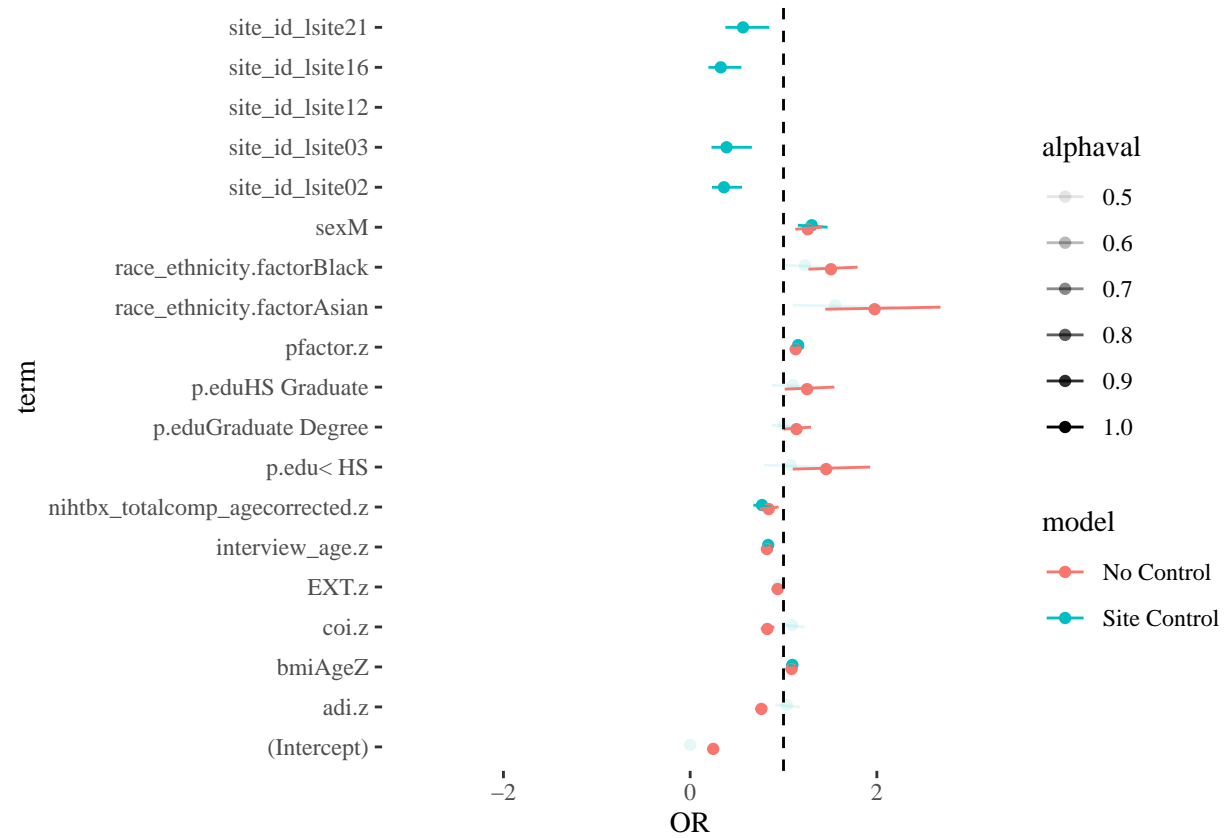


Figure 3: Significant ORs predicting .2mm missingness with and without site control

```

    site_id_1,
    `site01` = "Los Angeles, CA",
    `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",
    `site10` = "San Diego, CA",
    `site11` = "Gainesville, 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")

```

```

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",
      `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",
      `site10` = "San Diego, CA",
      `site11` = "Gainesville, 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")
```

```
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")
```

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

1.4.2.3 Figure: Missingness by site, inclusion criteria

1.5 Table: Associations of behavioral data with framewise displacement

```
sdcorr<-function(data,digits=2) {
  if (any(sapply(data, is.numeric) == FALSE)) {
    stop("Data must be numeric")
  }
  rawcorr <- Hmisc::rcorr(as.matrix(data))
  cors <- round(rawcorr[["r"]], digits)
  ps <- rawcorr[["P"]]
  stars <- sdlabFunctions::starPs(ps)
  dim(stars) <- dim(ps)
  #cors[upper.tri(cors, diag = TRUE)] <- ""
  #stars[upper.tri(stars, diag = TRUE)] <- ""
  cortable <- matrix(rbind(cors, stars), nrow = nrow(cors))
```



```

rownames(cortable) <- paste(1:ncol(cors), names(data), sep = ". ")
colnames(cortable) <- c(sapply(1:ncol(cors), function(x) c(x,
  paste(x, ".p")))))
cortable[, 1:(ncol(cortable) - 2)]
}

ccor<-df %>% select("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",
  filtered_fd_0.1:filtered_fd_0.5) %>%
  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
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 manuscript.
write.csv(contOut,"Manuscript/Tables/FDcorr_continuousDVs.csv",row.names=FALSE)
write.csv(discOut,"Manuscript/Tables/FDcorr_discreteDVs.csv",row.names=FALSE)

```

1.6 QCFC Figures

```

# 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:
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))
})

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

# Calculate Slopes
QCFCcdfz <- QCFCcdf %>%
  mutate(across(.fns=scale))

```

```

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

annot<-data.frame(threshold=unique(QCFCdflong$threshold),
                  text=apply(unique(QCFCdflong$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)),
height=5.5,width=7.5,)

```

1.7 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$gframe5_375),
".4"=sum(subdf1$gframe4_375),
".3"=sum(subdf1$gframe3_375),
"R"=sum(subdf1$ABCD_rsfmri_QC2),
".2"=sum(subdf1$gframe2_375),
".1"=sum(subdf1$gframe1_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$gframe5_375),
  ".4"=sum(subdf2$gframe4_375),
  ".3"=sum(subdf2$gframe3_375),
  "R"=sum(subdf2$ABCD_rsfmri_QC2),
  ".2"=sum(subdf2$gframe2_375),
  ".1"=sum(subdf2$gframe1_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,])

```

Table 2: 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

Vs. 0.180442

Table 3: Male participants with NIH toolbox total scores of $z \leq -1.5$

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

2 Supplementary Results

2.1 Categorical Variable Table by Condition

```
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])
  }
}

header<-catTableAtThresh(df, rep(TRUE,nrow(df)))

CatVar375Table <- cbind(
  header,
  catTableAtThresh(df, df$mri_ok,header),
  catTableAtThresh(df, df$any_rest,header),
```

```

catTableAtThresh(df, df$gframe5_375,header),
catTableAtThresh(df, df$gframe4_375,header),
catTableAtThresh(df, df$gframe3_375,header),
catTableAtThresh(df, df$gframe2_375,header),
catTableAtThresh(df, df$gframe1_375,header)
)

#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", "\\%", "$\\Sigma{} \\Delta{}\\%$", 7))) %>%
  insert_row("Full", "",
    "QC1", "", "",
    ".5mm Scrub", "", "",
    ".4mm Scrub", "", "",
    ".3mm Scrub", "", "",
    "QC2", "", "",
    ".2mm Scrub", "", "",
    ".1mm Scrub", "", "") %>%
  merge_cells(1,1:2) %>%
  merge_cells(1,3:5) %>%
  merge_cells(1,6:8) %>%
  merge_cells(1,9:11) %>%
  merge_cells(1,12:14) %>%
  merge_cells(1,15:17) %>%
  merge_cells(1,18:20) %>%
  merge_cells(1,21:23) %>%
  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)") %>%
  set_escape_contents(FALSE) %>%
  set_font_size(8) %>%
  set_number_format(3:nrow(.), c(3,5,8,11,14,17,20,23), value=fmt_pretty()) %>%
  set_number_format(3:nrow(.), c(4,6,7,9,10,12,13,15,16,18,19,21,22,24,25), value="%.2g") %>%
  set_all_padding(1)

CatVar375Hux[, c(1:13, 20:ncol(CatVar375Hux))]

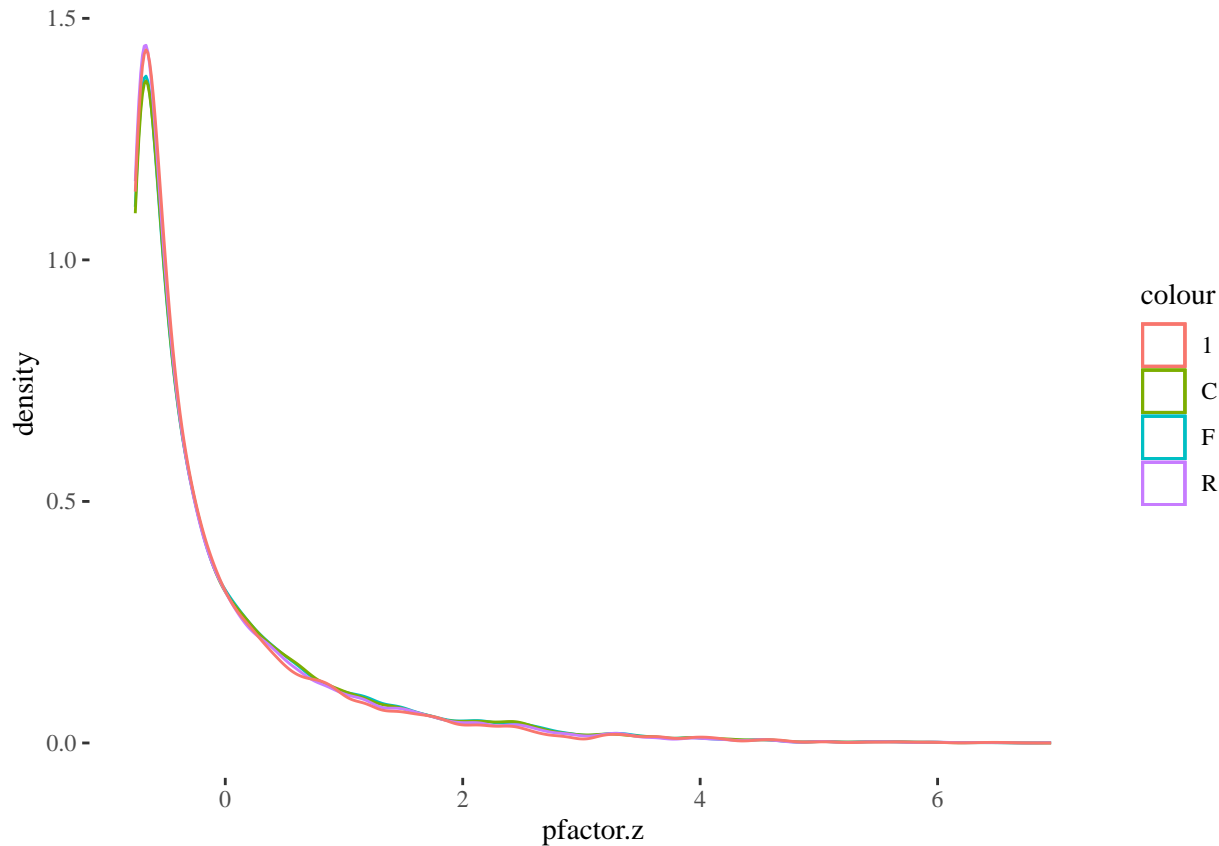
```

```
CatVar375Hux[,c(1:2,14:19)] %>%
  set_caption("Categorical Values at Each Level of Stringency (Subjects excluded with < 375 frames) - omitted columns")
```

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(gframe1_375==TRUE))
```



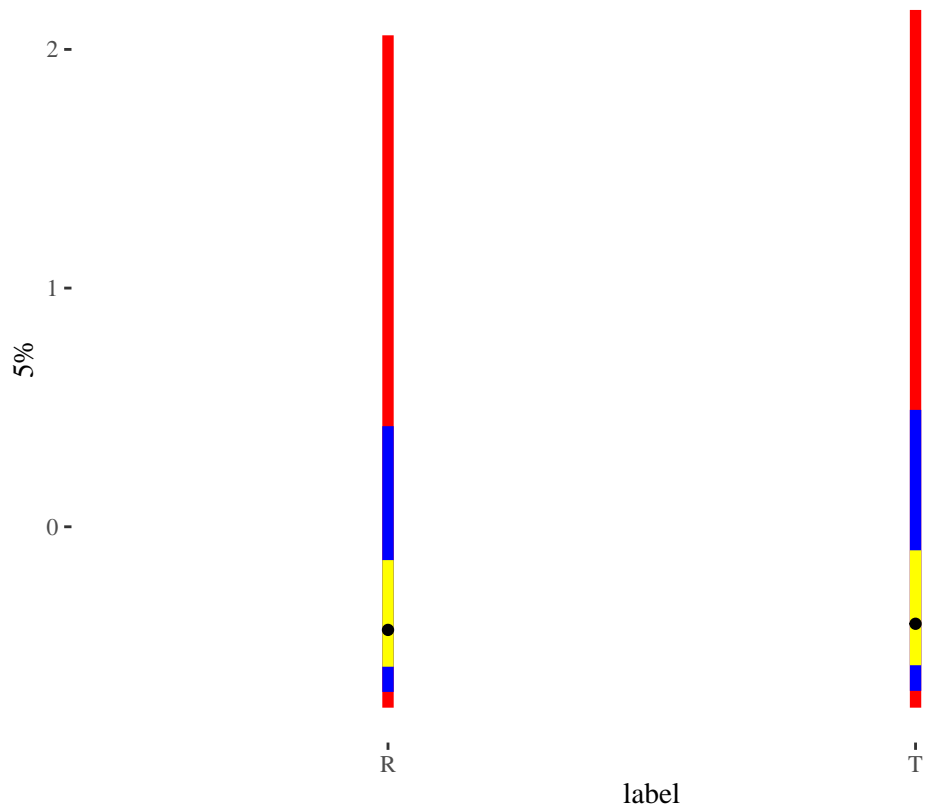

```

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%`))

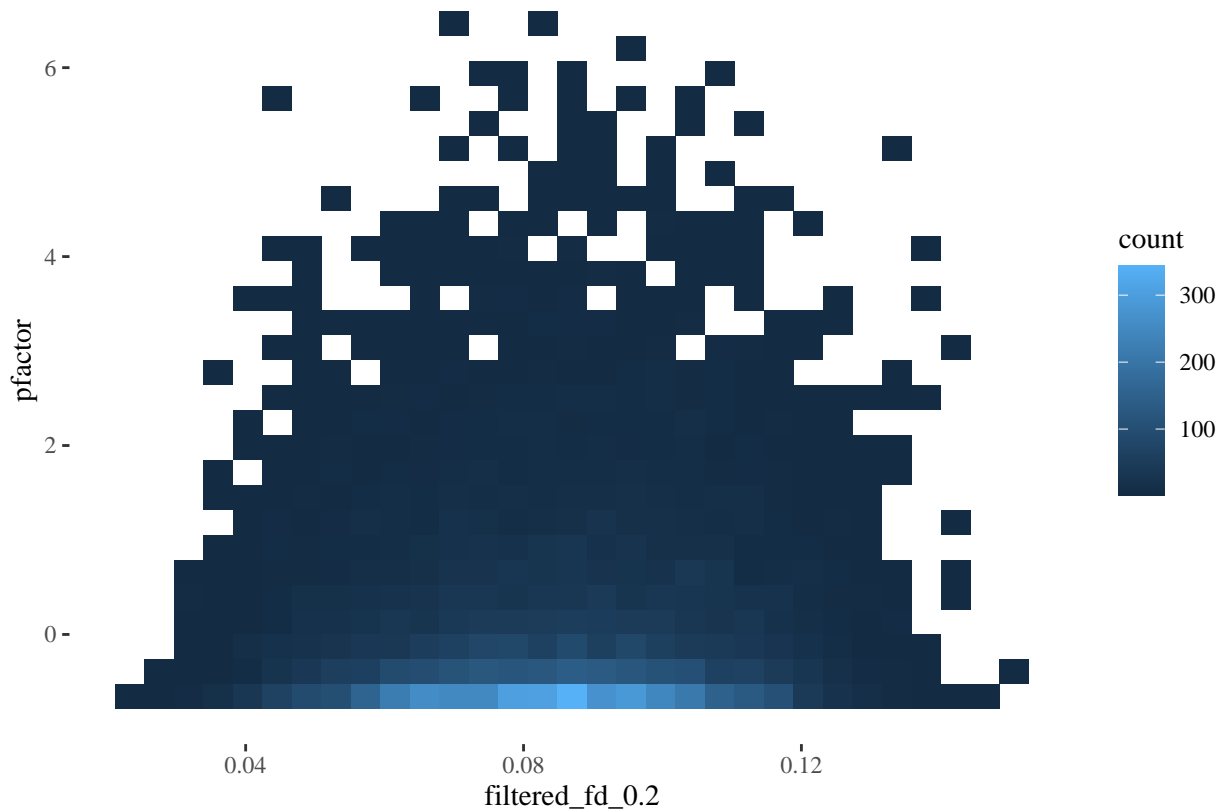
```



```

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_1)) %>%
  rename(c("Child Opportunity Index"="coi","Household Income"="household.income","Area Disadvantage Index"="adi","NIH Toolbox Total"="nihtbx_total"))
  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")
```

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",  
  "  >$200k",  
  "**Highest Parental Education**",  
  "  <High School",  
  "  HS Grad.",  
  "  Some College",  
  "  Graduate",  
  "**Census Race/Ethnicity**",  
  "  Black",  
  "  Hispanic",  
  "  Asian",  
  "  Other",  
  "**Trauma Count**",  
  "  1 Trauma",  
  "  >=2 Trauma",  
  "**Pubertal Status**",  
  "  Early Puberty",  
  "  Mid Puberty",  
  "  Late Puberty",  
  "  Post Pubertal",  
  "Area Disadvantage",  
  "Child Opportunity",  
  "NIHTB Flanker",  
  "NIHTB Crystallized",  
  "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") %>%
  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")
```

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
```

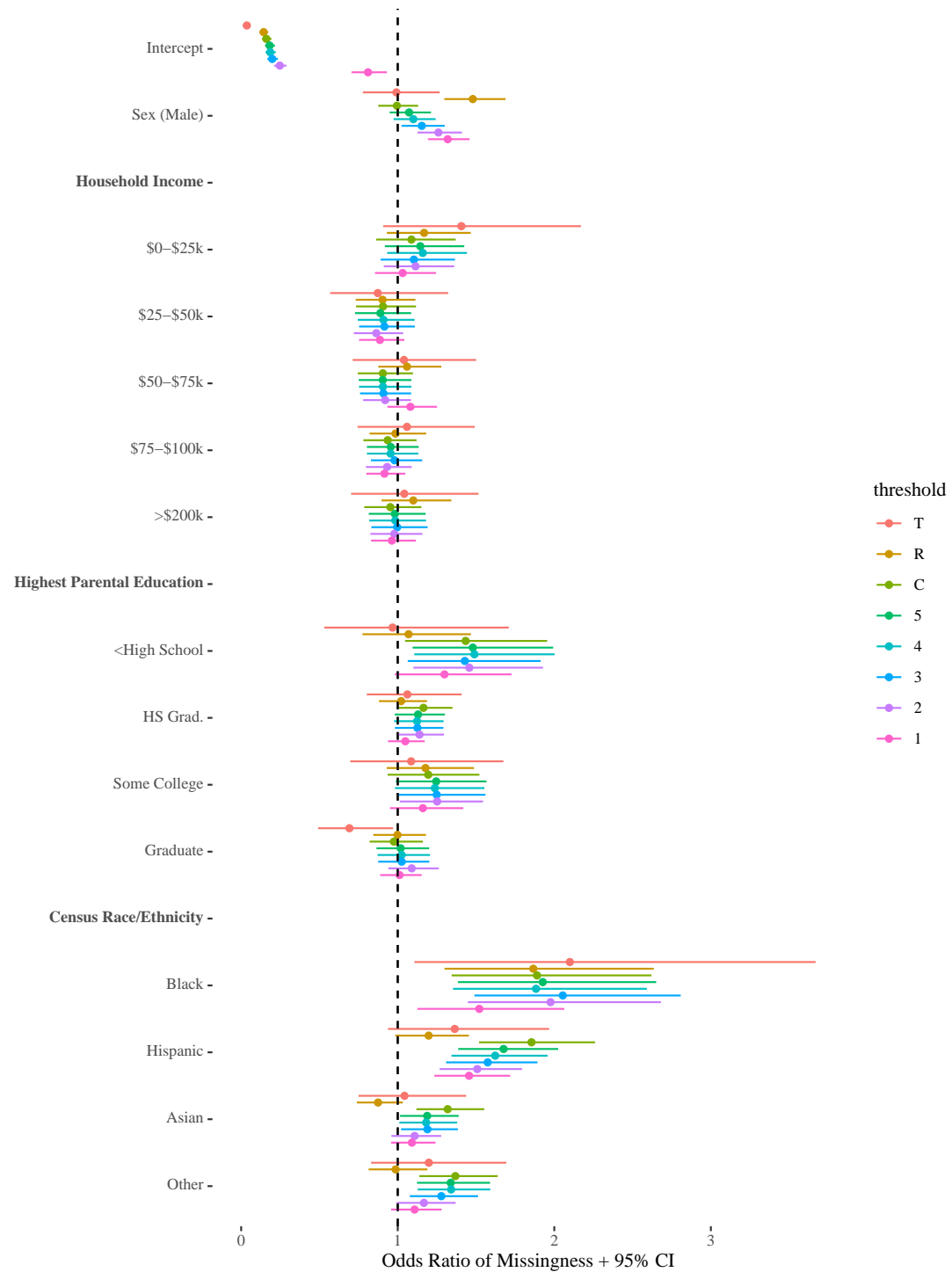


Figure 4: Odds ratios for missingness by threshold – Adjusted Models (part 1)

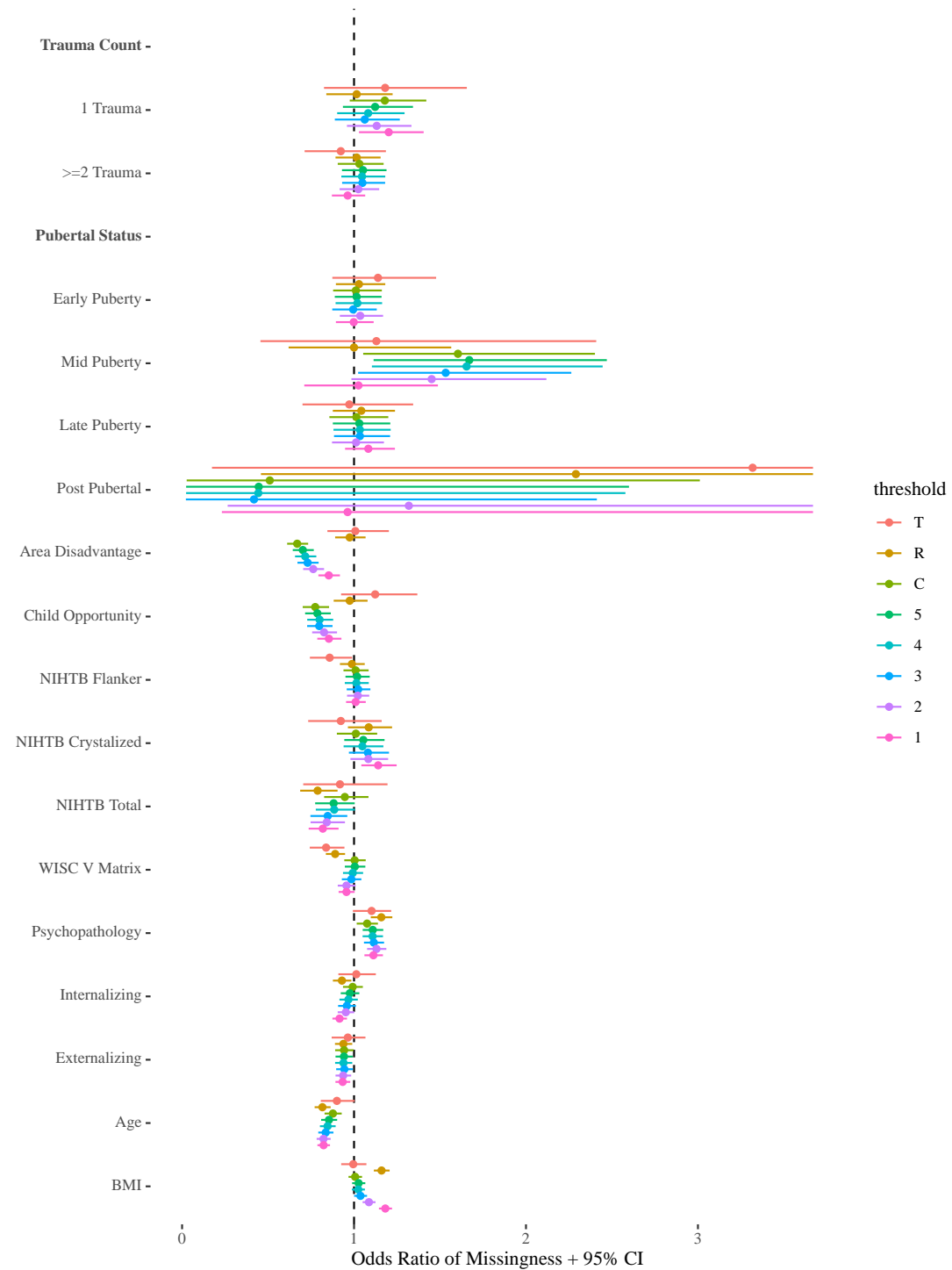


Figure 5: Odds ratios for missingness by threshold – Adjusted Models (part 2)

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")
```

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 comparability with Cosgrove et. al. (2020)
- The Behavioral Inhibition scale was removed. It is not available in the ABCD baseline data and was included in error
- Three 'QC' conditions were added in addition to the motion thresholds, to provide additional detail to inform study design. However, the H2 hypothesis test was performed without these conditions (as originally planned).

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

```
model.h1prereg.5mm<-glm(!gframe5_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nih
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(
  !gframe4_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrect
  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(
  !gframe3_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrect
  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(
  !gframe2_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrect
  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(
  !gframe1_375 ~ sex + household.income + p.edu + race_ethnicity.factor + ksads_factor + pds_category + adi.z + coi.z + nihtbx_flanker_agecorrect
  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

```

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

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 Crystallized",
"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\n[90% CI]\np"),5)),after=0) %>%
  insert_row(c("", ".5", ".4", ".3", ".2", ".1"))

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

h1prereghux %>%
  set_caption("Adjusted Models Output, as Pre-registered")

```

4 Save Tables

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

5 SessionInfo

R version 4.3.1 (2023-06-16)

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

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

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

other attached packages: *ggtext(v.0.1.2), emmeans(v.1.8.8), lmerTest(v.3.1-3), lme4(v.1.1-33), Matrix(v.1.6-0), ggupset(v.0.3.0), openxlsx(v.4.2.5.2), patchwork(v.1.1.2), viridis(v.0.6.3), 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.1.9.2), forcats(v.1.0.0), stringr(v.1.5.0), dplyr(v.1.1.2), purrr(v.1.0.1), readr(v.2.1.4), tidyr(v.1.3.0), tibble(v.3.2.1), ggplot2(v.3.4.2) and tidyverse(v.2.0.0)*

loaded via a namespace (and not attached): *rstudioapi(v.0.14), magrittr(v.2.0.3), TH.data(v.1.1-2), estimability(v.1.4.1), ggstance(v.0.3.6), farver(v.2.1.1), nloptr(v.2.0.3), rmarkdown(v.2.21), ragg(v.1.2.5), vctrs(v.0.6.2), memoise(v.2.0.1), minqa(v.1.2.5), sdsFunctions(v.0.2.0), base64enc(v.0.1-3), htmltools(v.0.5.5), broom(v.1.0.4), cellranger(v.1.1.0), Formula(v.1.2-5), htmlwidgets(v.1.6.2), plyr(v.1.8.8), sandwich(v.3.0-2), rootSolve(v.1.8.2.3), zoo(v.1.8-12), cachem(v.1.0.8), commonmark(v.1.9.0), mime(v.0.12), lifecycle(v.1.0.3), pkgconfig(v.2.0.3), R6(v.2.5.1), fastmap(v.1.1.1), shiny(v.1.7.4), digest(v.0.6.31), Exact(v.3.2), numDeriv(v.2016.8-1.1), colorspace(v.2.1-0), textshaping(v.0.3.6), Hmisc(v.5.1-0), labeling(v.0.4.2), fansi(v.1.0.4), timechange(v.0.2.0), mgcv(v.1.9-0), httr(v.1.4.5), compiler(v.4.3.1), proxy(v.0.4-27), withr(v.2.5.0), htmlTable(v.2.4.1), backports(v.1.4.1), highr(v.0.10), Rttf2pt1(v.1.3.12), MASS(v.7.3-60), gld(v.2.6.6), tools(v.4.3.1), foreign(v.0.8-82), zip(v.2.3.0), httpuv(v.1.6.9), extrafontdb(v.1.0), nnet(v.7.3-19), glue(v.1.6.2), nlme(v.3.1-162), promises(v.1.2.0.1), gridtext(v.0.1.5), grid(v.4.3.1), checkmate(v.2.2.0), cluster(v.2.1.4), generics(v.0.1.3), gtable(v.0.3.3), tzdb(v.0.3.0), class(v.7.3-22), data.table(v.1.14.8), lmom(v.2.9), hms(v.1.1.3), xml2(v.1.3.4), utf8(v.1.2.3), pillar(v.1.9.0), markdown(v.1.6), later(v.1.3.1), splines(v.4.3.1), lattice(v.0.21-8), survival(v.3.5-5), tidyselect(v.1.2.0), miniUI(v.0.1.1.1), knitr(v.1.42), gridExtra(v.2.3), svglite(v.2.1.1), xfun(v.0.39), expm(v.0.999-7), stringi(v.1.7.12), yaml(v.2.3.7), boot(v.1.3-28), evaluate(v.0.21), codetools(v.0.2-19), extrafont(v.0.19), cli(v.3.6.1), rpart(v.4.1.19), xtable(v.1.8-4), DescTools(v.0.99.49), systemfonts(v.1.0.4), munsell(v.0.5.0), Rcpp(v.1.0.10), readxl(v.1.4.2), coda(v.0.19-4), ellipsis(v.0.3.2), assertthat(v.0.2.1), QuantPsyc(v.1.6), mvtnorm(v.1.1-3), scales(v.1.2.1), e1071(v.1.7-13), crayon(v.1.5.2), rlang(v.1.1.1) and multcomp(v.1.4-23)*

Table 4: Categorical Values at Each Level of Stringency (Subjects excluded with < 375 Frames)

		Full		QC1			0.5mm Scrub			0.4mm Scrub			0.2mm Scrub			0.1mm Scrub		
		n	%	n	%	ΣΔ%	n	%	ΣΔ%	n	%	ΣΔ%	n	%	ΣΔ%	n	%	ΣΔ%
Total	Total	11,876	1e+02	0	0	-1e+02	9,600	81	-19	9,320	78	-22	8,507	72	-28	5,250	44	-56
Sex	F	5,680	48	0	NaN	NaN	4,575	48	-0.17	4,475	48	0.19	4,199	49	1.5	2,703	51	3.7
	M	6,196	52	0	NaN	NaN	5,025	52	0.17	4,845	52	-0.19	4,308	51	-1.5	2,547	49	-3.7
	Missing	0	0	0	NaN	NaN	0	0	0	0	0	0	0	0	0	0	0	0
Income	\$100k to \$200k	3,314	28	0	NaN	NaN	2,735	28	0.58	2,673	29	0.78	2,470	29	1.1	1,622	31	3
	\$0 to \$25k	1,635	14	0	NaN	NaN	1,206	13	-1.2	1,139	12	-1.5	988	12	-2.2	538	10	-3.5
	\$25k to \$50k	1,588	13	0	NaN	NaN	1,291	13	0.076	1,250	13	0.041	1,135	13	-0.03	676	13	-0.5
	\$50k to \$75k	1,499	13	0	NaN	NaN	1,254	13	0.44	1,215	13	0.41	1,106	13	0.38	630	12	-0.62
	\$75k to \$100k	1,572	13	0	NaN	NaN	1,316	14	0.47	1,281	14	0.51	1,188	14	0.73	773	15	1.5
	Over \$200k	1,250	11	0	NaN	NaN	1,014	11	0.037	995	11	0.15	930	11	0.41	626	12	1.4
	Missing	1,018	8.6	0	NaN	NaN	784	8.2	-0.41	767	8.2	-0.34	690	8.1	-0.46	385	7.3	-1.2
	Parent Ed.	College Degree	3,015	25	0	NaN	NaN	2,504	26	0.7	2,441	26	0.8	2,268	27	1.3	1,438	27
< HS	593	5	0	NaN	NaN	388	4	-0.95	370	4	-1	324	3.8	-1.2	173	3.3	-1.7	
HS Graduate	1,132	9.5	0	NaN	NaN	855	8.9	-0.63	815	8.7	-0.79	721	8.5	-1.1	373	7.1	-2.4	
Race/Ethnicity	Some College	3,079	26	0	NaN	NaN	2,528	26	0.41	2,432	26	0.17	2,179	26	-0.31	1,292	25	-1.3
	Graduate Degree	4,043	34	0	NaN	NaN	3,313	35	0.47	3,250	35	0.83	3,007	35	1.3	1,971	38	3.5
	Missing	14	0.12	0	NaN	NaN	12	0.12	0.0071	12	0.13	0.011	8	0.094	-0.024	3	0.057	-0.061
	White	6,180	52	0	NaN	NaN	5,238	55	2.5	5,104	55	2.7	4,710	55	3.3	3,073	59	6.5
	Black	1,784	15	0	NaN	NaN	1,338	14	-1.1	1,279	14	-1.3	1,122	13	-1.8	569	11	-4.2
	Hispanic	2,411	20	0	NaN	NaN	1,845	19	-1.1	1,799	19	-1	1,630	19	-1.1	960	18	-2
	Asian	252	2.1	0	NaN	NaN	180	1.9	-0.25	175	1.9	-0.24	155	1.8	-0.3	101	1.9	-0.2
	Other	1,247	11	0	NaN	NaN	997	10	-0.11	961	10	-0.19	888	10	-0.062	545	10	-0.12
Missing	2	0.017	0	NaN	NaN	2	0.021	0.004	2	0.021	0.0046	2	0.024	0.0067	2	0.038	0.021	
Trauma Count	0	7,723	65	0	NaN	NaN	6,275	65	0.33	6,102	65	0.44	5,586	66	0.63	3,472	66	1.1
	1	3,004	25	0	NaN	NaN	2,422	25	-0.066	2,345	25	-0.13	2,147	25	-0.057	1,338	25	0.19
	>2	1,149	9.7	0	NaN	NaN	903	9.4	-0.27	873	9.4	-0.31	774	9.1	-0.58	440	8.4	-1.3
	Missing	0	0	0	NaN	NaN	0	0	0	0	0	0	0	0	0	0	0	0
Puberty	pre-pubertal	5,938	50	0	NaN	NaN	4,845	50	0.47	4,707	51	0.5	4,284	50	0.36	2,672	51	0.9
	early puberty	2,815	24	0	NaN	NaN	2,275	24	-0.0054	2,207	24	-0.023	1,998	23	-0.22	1,262	24	0.33
	mid puberty	2,798	24	0	NaN	NaN	2,253	23	-0.091	2,190	23	-0.062	2,025	24	0.24	1,195	23	-0.8
	late puberty	188	1.6	0	NaN	NaN	135	1.4	-0.18	127	1.4	-0.22	118	1.4	-0.2	75	1.4	-0.15
	post pubertal	12	0.1	0	NaN	NaN	10	0.1	0.0031	9	0.097	-0.0045	7	0.082	-0.019	3	0.057	-0.044
	Missing	125	1.1	0	NaN	NaN	82	0.85	-0.2	80	0.86	-0.19	75	0.88	-0.17	43	0.82	-0.23

Table 5: Categorical Values at Each Level of Stringency (Subjects excluded with < 375 frames) - omitted columns

		0.3mm Scrub			QC2		
		n	%	$\Sigma\Delta\%$	n	%	$\Sigma\Delta\%$
Total	Total	9,262	78	-22	9,098	77	-23
Sex	F	4,458	48	0.3	4,408	48	0.62
	M	4,804	52	-0.3	4,690	52	-0.62
	Missing	0	0	0	0	0	0
Income	\$100k to \$200k	2,660	29	0.81	2,623	29	0.93
	\$0 to \$25k	1,126	12	-1.6	1,106	12	-1.6
	\$25k to \$50k	1,239	13	0.0057	1,212	13	-0.05
	\$50k to \$75k	1,208	13	0.42	1,189	13	0.45
	\$75k to \$100k	1,276	14	0.54	1,250	14	0.5
	Over \$200k	992	11	0.19	978	11	0.22
	Missing	761	8.2	-0.36	740	8.1	-0.44
Parent Ed.	College Degree	2,429	26	0.84	2,393	26	0.92
	< HS	365	3.9	-1.1	359	3.9	-1
	HS Graduate	809	8.7	-0.8	786	8.6	-0.89
	Some College	2,410	26	0.094	2,361	26	0.025
	Graduate Degree	3,238	35	0.92	3,188	35	1
	Missing	11	0.12	0.00088	11	0.12	0.003
Race/Ethnicity	White	5,075	55	2.8	4,999	55	2.9
	Black	1,269	14	-1.3	1,241	14	-1.4
	Hispanic	1,788	19	-1	1,746	19	-1.1
	Asian	175	1.9	-0.23	168	1.8	-0.28
	Other	953	10	-0.21	942	10	-0.15
	Missing	2	0.022	0.0048	2	0.022	0.0051
Trauma Count	0	6,059	65	0.39	5,957	65	0.45
	1	2,330	25	-0.14	2,285	25	-0.18
	>2	873	9.4	-0.25	856	9.4	-0.27
	Missing	0	0	0	0	0	0
Puberty	pre-pubertal	4,671	50	0.43	4,575	50	0.29
	early puberty	2,195	24	-0.0043	2,160	24	0.038
	mid puberty	2,181	24	-0.012	2,150	24	0.071
	late puberty	126	1.4	-0.22	126	1.4	-0.2
	post pubertal	9	0.097	-0.0039	9	0.099	-0.0021
	Missing	80	0.86	-0.19	78	0.86	-0.2

Table 6: 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 Crystallized	338	2.85
WISC V	249	2.1
NIH Toolbox Flanker	153	1.29
Pubertal Status	125	1.05

Table 7: 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

Table 8: Adjusted Models Output, as Pre-registered

	0.5	0.4	0.3	0.2	0.1
Variable	OR[90% CI] _p	OR[90% CI] _p	OR[90% CI] _p	OR[90% CI] _p	OR[90% CI] _p
Intercept	.18***.15-.22.000	.18***.16-.22.000	.20***.17-.23.000	.25***.21-.29.000	.80***.70-.92.001
Sex (Male)	1.09.96-1.23.183	1.11.99-1.26.083	1.17**1.04-1.32.008	1.31***1.17-1.46.000	1.42***1.29-1.57.000
Household Income (ref: \$100-\$200k)					
\$0-\$25k	1.15.93-1.44.202	1.17.94-1.45.160	1.11.90-1.38.323	1.13.93-1.38.227	1.06.88-1.28.556
\$25-\$50k	.89.73-1.09.273	.91.75-1.11.371	.92.76-1.12.398	.87.73-1.04.133	.89.76-1.05.174
\$50-\$75k	.92.76-1.10.354	.92.76-1.10.355	.92.77-1.10.365	.93.79-1.10.417	1.09.94-1.26.241
\$75-\$100k	.96.80-1.14.619	.96.81-1.13.609	.98.83-1.16.831	.94.80-1.09.403	.92.80-1.05.227
>\$200k	.98.81-1.18.838	.98.82-1.18.861	1.00.83-1.19.960	.98.82-1.15.771	.96.82-1.11.549
Highest Parental Education (ref: College Degree)					
<High School	1.49**1.10-2.00.009	1.50**1.11-2.01.008	1.44*1.07-1.93.015	1.49**1.12-1.97.005	1.35*1.02-1.79.036
HS Grad.	1.25.99-1.58.055	1.24.99-1.56.061	1.26*1.01-1.58.041	1.29*1.04-1.59.019	1.23*1.01-1.50.044
Some College	1.02.86-1.20.821	1.03.87-1.21.763	1.03.88-1.20.733	1.10.95-1.27.205	1.03.91-1.18.626
Graduate	1.13.98-1.30.092	1.12.98-1.29.105	1.12.98-1.29.093	1.141.00-1.29.051	1.04.93-1.17.454
Census Race/Ethnicity (ref: White)					
Black	1.69***1.40-2.04.000	1.63***1.35-1.97.000	1.59***1.32-1.91.000	1.53***1.29-1.82.000	1.48***1.26-1.75.000
Hispanic	1.20*1.02-1.40.023	1.19*1.02-1.39.027	1.21*1.04-1.40.015	1.14.99-1.32.064	1.16*1.02-1.32.024
Asian	1.92***1.38-2.64.000	1.88***1.35-2.58.000	2.05***1.48-2.80.000	1.96***1.44-2.66.000	1.49**1.10-2.01.010
Other	1.34***1.13-1.59.001	1.34***1.13-1.59.001	1.28**1.08-1.52.004	1.18*1.00-1.38.049	1.12.97-1.29.125
KSADS Trauma Count (ref: 0 Exposures)					
1 Trauma	1.06.93-1.19.383	1.05.93-1.19.427	1.05.93-1.18.402	1.03.92-1.15.643	.96.87-1.06.453
>=2 Trauma	1.12.93-1.34.214	1.08.90-1.29.397	1.06.89-1.26.511	1.13.96-1.33.146	1.20*1.03-1.40.023
Pubertal Status (ref: pre-pubertal)					
Early Puberty	1.02.90-1.17.720	1.03.90-1.17.681	1.01.89-1.15.908	1.07.95-1.20.286	1.07.96-1.19.233
Mid Puberty	1.05.90-1.24.511	1.06.90-1.24.497	1.07.91-1.24.410	1.08.94-1.25.277	1.25***1.10-1.42.001