

Exploring covariates, rs-fMRI usability, and functional connectivity

Liwei Wang and MB Nebel

2022-03-25

```
#for flowchart
library(DiagrammeR)
#to export flowchart as .png
#webshot::install_phantomjs()
library(DiagrammeRsvg)
library(rsvg)
library(grid)
library(xtable)
library(dplyr)
library(ggplot2)
library(gridExtra)
#nest/unnest
library(tidyr)
#map function (kind of like a for loop)
library(purrr)
#tidy model summary
library(broom)
library(readxl)
#for tableby
library(arsenal)

#Also uses functions from plyr, scales, mgcv, cowplot
```

Define geom_split_violin()

Based on <https://debruine.github.io/post/plot-comparison/> (<https://debruine.github.io/post/plot-comparison/>)

```

GeomSplitViolin <- ggproto("GeomSplitViolin", GeomViolin, draw_group = function(self,
  data, ..., draw_quantiles = NULL) {
  data <- transform(data, xminv = x - violinwidth * (x - xmin), xmaxv = x + violinwidth *
  h *
    (xmax - x))
  grp <- data[1, "group"]
  newdata <- plyr::arrange(transform(data, x = if (grp%%2 == 1)
    xminv else xmaxv), if (grp%%2 == 1)
    y else -y)
  newdata <- rbind(newdata[1, ], newdata, newdata[nrow(newdata), ], newdata[1,
  ])
  newdata[c(1, nrow(newdata) - 1, nrow(newdata)), "x"] <- round(newdata[1, "x"])
  if (length(draw_quantiles) > 0 & !scales::zero_range(range(data$y))) {
    stopifnot(all(draw_quantiles >= 0), all(draw_quantiles <= 1))
    quantiles <- ggplot2::create_quantile_segment_frame(data, draw_quantiles)
    aesthetics <- data[rep(1, nrow(quantiles)), setdiff(names(data), c("x", "y")),
      drop = FALSE]
    aesthetics$alpha <- rep(1, nrow(quantiles))
    both <- cbind(quantiles, aesthetics)
    quantile_grob <- GeomPath$draw_panel(both, ...)
    ggplot2::ggname("geom_split_violin", grid::grobTree(GeomPolygon$draw_panel(newdata,
    ata,
      ...), quantile_grob))
  } else {
    ggplot2::ggname("geom_split_violin", GeomPolygon$draw_panel(newdata, ...))
  }
})

geom_split_violin <- function(mapping = NULL, data = NULL, stat = "ydensity", position =
"identity",
  ..., draw_quantiles = NULL, trim = TRUE, scale = "area", na.rm = FALSE, show.legend
= NA,
  inherit.aes = TRUE) {
  layer(data = data, mapping = mapping, stat = stat, geom = GeomSplitViolin, position
= position,
    show.legend = show.legend, inherit.aes = inherit.aes, params = list(trim = trim,
    scale = scale, draw_quantiles = draw_quantiles, na.rm = na.rm, ...))
}

```

Load initial data set

```

load('./Data/noImputation/DataWithPropensities_seed1.RData')

#convert PrimaryDiagnosis to factor
dat3$PrimaryDiagnosis <- factor(dat3$PrimaryDiagnosis, levels = c("Autism", "None"))
levels(dat3$PrimaryDiagnosis)

```

[1] "Autism" "None"

```

dat3$PrimaryDiagnosis <- relevel(dat3$PrimaryDiagnosis, "None")
levels(dat3$PrimaryDiagnosis) = c("TD", "ASD")

tabInit<- tableby(PrimaryDiagnosis ~ Sex + AgeAtScan,
                  data=dat3)
summary(tabInit,
        title='Summary of diagnosis and sex of all participants who attempted a scan',di
        gits=1, digits.p=4,digits.pct=1, numeric.simplify=TRUE, total=FALSE, test=FALSE)

```

Summary of diagnosis and sex of all participants who attempted a scan

	TD (N=372)	ASD (N=173)
Sex		
F	114 (30.6%)	25 (14.5%)
M	258 (69.4%)	148 (85.5%)
AgeAtScan		
Mean (SD)	10.4 (1.2)	10.4 (1.4)
Range	8.0 - 13.0	8.0 - 13.0

Our initial cohort is an aggregate of **545** children between 8 and 13-years old who participated in one of several neuroimaging studies at Kennedy Krieger Institute (KKI) between 2007 and 2020.

Reshape data to combine motion quality control (QC) levels

```

# create dummy factor to include all subjects
dat3$noExclusion <- ifelse(dat3$ID > 0, "Pass", "Pass")
dat3$noExclusion <- factor(dat3$noExclusion, levels = c("Pass", "Fail"))

# convert KKI_criteria to factor with reference level 'Pass'
dat3$KKI_criteria <- factor(dat3$KKI_criteria, levels = c("Pass", "Fail"))

# convert Ciric_length to factor with reference level 'Pass' to match
# KKI_criteria
dat3$Ciric_length <- factor(dat3$Ciric_length, levels = c("Pass", "Fail"))

# combine Ciric_length, KKI, and noExclusion exclusion into one variable
allVariables = c("ID", "PrimaryDiagnosis", "AgeAtScan", "Ciric_length", "KKI_criteria",
  "noExclusion", "PANESS.TotalOverflowNotAccountingForAge", "SRS.Score", "WISC.GAI",
  "DuPaulHome.InattentionRaw", "DuPaulHome.HyperactivityRaw", "ADOS.Comparable.Total",
  "CurrentlyOnStimulants", "HeadCoil", "Sex", "ADHD_Secondary", "SES.Family", "Race2",
  "handedness", "CompletePredictorCases", "YearOfScan", "MeanFramewiseDisplacement.KK
I")

idVariables = c("ID", "PrimaryDiagnosis", "AgeAtScan", "PANESS.TotalOverflowNotAccountin
gForAge",
  "SRS.Score", "WISC.GAI", "DuPaulHome.InattentionRaw", "DuPaulHome.HyperactivityRaw",
  "ADOS.Comparable.Total", "CurrentlyOnStimulants", "HeadCoil", "Sex", "ADHD_Secondar
y",
  "SES.Family", "Race2", "handedness", "CompletePredictorCases", "YearOfScan",
  "MeanFramewiseDisplacement.KKI")

qcMelt <- reshape2::melt(dat3[, allVariables], id.vars = names(dat3)[which(names(dat3) %
in%
  idVariables)], variable.name = "Motion.Exclusion.Level", value.name = "Included")

# rename exclusion levels NOTE: need None to be highest level for
# geom_split_violin
levels(qcMelt$Motion.Exclusion.Level) <- c("Strict", "Lenient", "None")

# convert Included to factor with pass as reference
qcMelt$Included <- factor(qcMelt$Included, levels = c("Pass", "Fail"))

# rename levels of value
levels(qcMelt$Included) <- c("Included", "Excluded")

```

Motion QC levels:

1. **Strict motion QC** = Ciric_length

In the strict case, scans were excluded if mean FD exceeded .2 mm or they included less than five minutes of data free from frames with FD exceeding .25 mm

2. **Lenient motion QC** = KKI_criteria

In the lenient case, scans were excluded if the participant had less than 5 minutes of continuous data after removing frames in which the participant moved more than the nominal size of a voxel between any two frames (3 mm) or their head rotated 3. This procedure was modeled after common head motion exclusion criteria for task

fMRI data, which rely on voxel size to determine thresholds for unacceptable motion.

3. **None** = all participants

Limit initial dataset to complete cases

```
dat3 <- filter(dat3, CompletePredictorCases==1)
```

2.1.4 Determine number of participants in set of complete cases who did not attempt or aborted the scan early

```
dat3$aborted = rep("No", length=nrow(dat3))
dat3$aborted[is.na(dat3$MeanFramewiseDisplacement) & dat3$KKI_criteria=="Fail"] = "Yes"

tabAbort<- tableby(PrimaryDiagnosis ~ aborted,
                   data=dat3)
summary(tabAbort,
        title='Scans aborted by diagnosis',digits=1, digits.p=4,digits.pct=1, numeric.simplify=TRUE, total=FALSE, test=FALSE)
```

Scans aborted by diagnosis

	TD (N=348)	ASD (N=137)
aborted		
No	344 (98.9%)	134 (97.8%)
Yes	4 (1.1%)	3 (2.2%)

Scans were either not attempted after two unsuccessful mock scan training sessions or aborted due to non-compliance for 7 of the participants (3 ASD) in the set of complete cases.

2.1.5 Determine number of participants in complete cases set who attempted more than one scan

```
load('./Data/noImputation/nScans.RData')

dat3 <- merge(dat3, nScans, all.x = TRUE)

dat3$n <- factor(dat3$n)

tabNScans<- tableby(PrimaryDiagnosis ~ n,
                   data=dat3)
summary(tabNScans,
        title='Number of scans attempted',digits=1, digits.p=4,digits.pct=1, numeric.simplify=TRUE, total=FALSE, test=FALSE)
```

Number of scans attempted

	TD (N=348)	ASD (N=137)
--	------------	-------------

	TD (N=348)	ASD (N=137)
n		
1	282 (81.0%)	120 (87.6%)
2	62 (17.8%)	17 (12.4%)
3	3 (0.9%)	0 (0.0%)
5	1 (0.3%)	0 (0.0%)

```

getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}

nMode = getmode(as.numeric(as.character(dat3$n[dat3$n!=1])))

```

83 of the complete cases (66 typically developing, 17 ASD) attempted more than one resting-state fMRI scan. Most participants with multiple attempts had 2 scans.

Table 1. Summarize socio-demographic characteristics of the complete predictor cases for paper

```

completeCases <- filter(qcMelt, CompletePredictorCases==1)

#make M reference level for sex
completeCases$Sex <- relevel(as.factor(completeCases$Sex), "M")

#labels for table
labels(completeCases) <- c(PrimaryDiagnosis = 'Diagnosis',
                           AgeAtScan = 'Age in Years',
                           Sex = 'Sex',
                           handedness = 'Handedness',
                           Race2 = 'Race',
                           SES.Family = 'Socioeconomic Status',
                           CurrentlyOnStimulants = 'Currently on Stimulants?')

#use chisq for Sex and handedness, kruskal-wallis rank test for Age
tabSex<- tableby( PrimaryDiagnosis~Sex+AgeAtScan+handedness, data=filter(completeCases,
  Motion.Exclusion.Level=="None" & Included=="Included"), control=tableby.control(numeri
c.test="kwt", cat.test="chisq", total=FALSE))

#use fisher's exact test for Race, k-w for SES
tabRace<- tableby( PrimaryDiagnosis~Race2+SES.Family, data=filter(completeCases, Motion.
Exclusion.Level=="None" & Included=="Included"), control=tableby.control(numeric.test="k
wt", cat.test="fe", total=FALSE))

tab12 <- merge(tabSex, tabRace)

#Currently on Stimulants - no test because no TDs are currently on stimulants by design
completeCases$CurrentlyOnStimulants <- factor(completeCases$CurrentlyOnStimulants)

#rename factor levels
levels(completeCases$CurrentlyOnStimulants)[levels(completeCases$CurrentlyOnStimulants)=
=="0"] <- "No"
levels(completeCases$CurrentlyOnStimulants)[levels(completeCases$CurrentlyOnStimulants)=
=="1"] <- "Yes"

tabStim<- tableby( PrimaryDiagnosis~CurrentlyOnStimulants, data=filter(completeCases, Mo
tion.Exclusion.Level=="None" & Included=="Included"), control=tableby.control(total=FALS
E, test=FALSE))

tab123 <- merge(tab12, tabStim)
summary(tab123)

```

	TD (N=348)	ASD (N=137)	p value
Sex			0.002
M	242 (69.5%)	114 (83.2%)	
F	106 (30.5%)	23 (16.8%)	
Age in Years			0.664
Mean (SD)	10.353 (1.249)	10.286 (1.344)	

	TD (N=348)	ASD (N=137)	p value
Range	8.020 - 12.980	8.010 - 12.990	
Handedness			0.308
Left	17 (4.9%)	10 (7.3%)	
Mixed	19 (5.5%)	11 (8.0%)	
Right	312 (89.7%)	116 (84.7%)	
Race			0.005
African American	36 (10.3%)	7 (5.1%)	
Asian	27 (7.8%)	3 (2.2%)	
Biracial	45 (12.9%)	12 (8.8%)	
Caucasian	240 (69.0%)	115 (83.9%)	
Socioeconomic Status			0.007
Mean (SD)	54.072 (9.408)	51.883 (9.356)	
Range	18.500 - 66.000	27.000 - 66.000	
CurrentlyOnStimulants			
No	348 (100.0%)	89 (65.0%)	
Yes	0 (0.0%)	48 (35.0%)	

Generate version of Table 1 to paste into overleaf

```
tab <- xtable(as.data.frame(summary(tab123)))
print(tab, type="latex")
```



```

## % latex table generated in R 4.1.2 by xtable 1.8-4 package
## % Fri Mar 25 19:19:48 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rllll}
## \hline
## & & TD (N=348) & ASD (N=137) & p value \\
## \hline
## 1 & Sex & & & 0.002 \\
## 2 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;M & 242 (69.5\%) & 114 (83.2\%) & \\
## 3 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;F & 106 (30.5\%) & 23 (16.8\%) & \\
## 4 & Age in Years & & & 0.664 \\
## 5 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Mean (SD) & 10.353 (1.249) & 10.286 (1.344) & \\
## 6 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Range & 8.020 - 12.980 & 8.010 - 12.990 & \\
## 7 & Handedness & & & 0.308 \\
## 8 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Left & 17 (4.9\%) & 10 (7.3\%) & \\
## 9 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Mixed & 19 (5.5\%) & 11 (8.0\%) & \\
## 10 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Right & 312 (89.7\%) & 116 (84.7\%) & \\
## 11 & Race & & & 0.005 \\
## 12 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;African American & 36 (10.3\%) & 7 (5.1\%) & \\
## 13 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Asian & 27 (7.8\%) & 3 (2.2\%) & \\
## 14 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Biracial & 45 (12.9\%) & 12 (8.8\%) & \\
## 15 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Caucasian & 240 (69.0\%) & 115 (83.9\%) & \\
## 16 & Socioeconomic Status & & & 0.007 \\
## 17 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Mean (SD) & 54.072 (9.408) & 51.883 (9.356) & \\
## 18 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Range & 18.500 - 66.000 & 27.000 - 66.000 & \\
## 19 & CurrentlyOnStimulants & & & \\
## 20 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;No & 348 (100.0\%) & 89 (65.0\%) & \\
## 21 & \&nbsp;\&nbsp;\&nbsp;\&nbsp;Yes & 0 (0.0\%) & 48 (35.0\%) & \\
## \hline
## \end{tabular}
## \end{table}

```

3.1.1. The impact of motion QC on sample size

Figure 3a. Exclusion flowchart

Proportion of complete cases included and excluded by motion QC

	Strict (N=485)	Lenient (N=485)
Included		
Included	165 (34.0%)	390 (80.4%)
Excluded	320 (66.0%)	95 (19.6%)

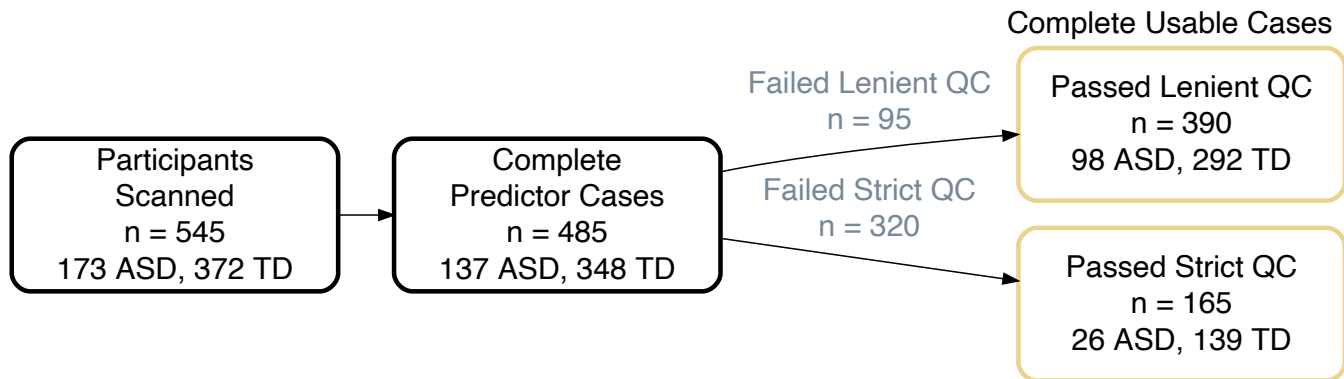


Figure 3a. Motion quality control leads to dramatic reductions in sample size. Flow chart of inclusion criteria for this study showing the number of participants remaining after each exclusion step. Lenient motion quality control (QC) excluded 20% of complete predictor cases, while strict motion QC excluded 66% of complete predictor cases.

Figure 3b. Proportion excluded stratified by Primary Diagnosis and motion QC level

Define theme for proportion excluded plots

```

My_Theme_prop = theme_light()+theme(
  legend.title =element_blank(),
  axis.title.x = element_text(size = 12),
  axis.title.y = element_text(size = 11),
  plot.title = element_text(size = 30),
  axis.text.x = element_text(size = 10),
  axis.text.y = element_text(size = 10),
  strip.text.x = element_text(size = 12,color="black"),
  strip.background = element_rect(fill = "white"))
  
```

Figure 3b. Plot proportions

```

motion <- filter(completeCases, Motion.Exclusion.Level != "None")
motion$Motion.Exclusion.Level <- droplevels(motion$Motion.Exclusion.Level)

motion <- group_by(motion, PrimaryDiagnosis, Motion.Exclusion.Level, Included)

dx_proportions <- ggplot(motion, aes(x = PrimaryDiagnosis, fill = Included)) + geom_bar
(position = "fill",
  alpha = 0.6) + facet_grid(~Motion.Exclusion.Level) + scale_fill_manual(values = c("#
FDE599",
  "#9FB0CC")) + scale_color_manual(values = c("#E9D38D", "#8C9AB4")) + My_Theme_prop +
theme(legend.title = element_blank()) + theme(legend.title = element_blank()) +
  ylab("Proportion of Children") + theme(legend.position = "bottom") + theme(legend.ma
rgin = margin(t = 0,
  r = 0, b = -1, l = -1)) + theme(legend.key.size = unit(0.15, "in"), legend.text = el
ement_text(size = 11)) +
  theme(axis.title.x = element_blank()) + theme(axis.text.x = element_text(size = 10))

png("./CovariatesAndRS-fMRIUsability/fig_propExcludedDx_cc.png", width = 3, height = 2.5
,
  units = "in", res = 200)
dx_proportions
invisible(dev.off())

# Pearson's chi squared tests
extib <- tibble(motion)

exNest <- extib %>%
  select(c("PrimaryDiagnosis", "Motion.Exclusion.Level", "Included")) %>%
  group_by(Motion.Exclusion.Level) %>%
  tidyr::nest()

# nested models
ex_chisq <- exNest %>%
  mutate(stats = map(data, ~broom::tidy(chisq.test(.x$PrimaryDiagnosis, .x$Include
d)))) %>%
  unnest(stats)

ex_chisq

```

```

## # A tibble: 2 × 6
## # Groups:   Motion.Exclusion.Level [2]
##   Motion.Exclusion.Level data                statistic  p.value parameter method
##   <fct>                  <list>                <dbl>    <dbl>    <int> <chr>
## 1 Strict                 <tibble [485 × 2]>    18.3  0.0000186      1 Pears...
## 2 Lenient                <tibble [485 × 2]>     8.79  0.00303      1 Pears...

```

```
dx_proportions
```

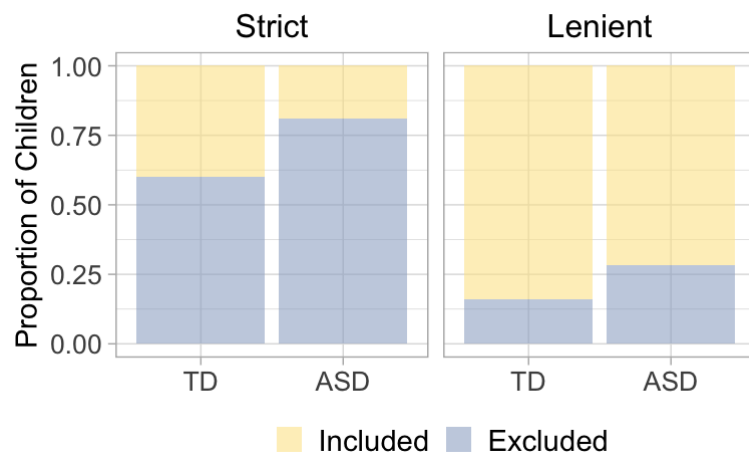


Figure 3b. The proportion of children in each diagnosis group whose scans were included (yellow) and excluded (slate blue) using the strict (left) and lenient (right panel) gross motion QC. A larger proportion of children in the autism spectrum disorder (ASD) group had unusable data and were excluded compared to typically developing (TD) children using lenient motion QC ($\chi^2=8.8$, $df=1$, $p=0.003$) and strict ($\chi^2=18.3$, $df=1$, $p=0.003$).

Figure 3b. Numbers for main text

```
tabASD<- tableby(Motion.Exclusion.Level~Included, data=filter(completeCases,
                                                             PrimaryDiagnosis=="A
SD"))
summary(tabASD,
        title = "Proportion of ASD complete cases included/excluded",
        digits=0, digits.p=4,digits.pct=1, numeric.simplify=TRUE, total=FALSE, test=FALS
E)
```

Proportion of ASD complete cases included/excluded

	Strict (N=137)	Lenient (N=137)	None (N=137)
Included			
Included	26 (19.0%)	98 (71.5%)	137 (100.0%)
Excluded	111 (81.0%)	39 (28.5%)	0 (0.0%)

```
tabTD<- tableby(Motion.Exclusion.Level~Included, data=filter(completeCases,
                                                             PrimaryDiagnosis=="T
D"))
summary(tabTD,
        title = "Proportion of TD complete cases included/excluded",
        digits=0, digits.p=4,digits.pct=1, numeric.simplify=TRUE, total=FALSE, test=FALS
E)
```

Proportion of TD complete cases included/excluded

	Strict (N=348)	Lenient (N=348)	None (N=348)
Included			

	Strict (N=348)	Lenient (N=348)	None (N=348)
Included	139 (39.9%)	292 (83.9%)	348 (100.0%)
Excluded	209 (60.1%)	56 (16.1%)	0 (0.0%)

The proportion of children excluded differs across diagnostic groups using both the lenient and strict motion QC.

3.1.2 rs-fMRI exclusion probability changes with phenotype and age

Covariates

```
phenoVariables <- c("ID", "PrimaryDiagnosis",
                  "ADOS.Comparable.Total",
                  "SRS.Score",
                  "PANESS.TotalOverflowNotAccountingForAge",
                  "DuPaulHome.InattentionRaw",
                  "DuPaulHome.HyperactivityRaw",
                  "AgeAtScan",
                  "WISC.GAI", "SES.Family",
                  "Motion.Exclusion.Level", "Included", "Sex")

phenoIDs <- c("ID", "PrimaryDiagnosis", "Motion.Exclusion.Level", "Included", "Sex", "SES.Family")

aim1 <- reshape2::melt(completeCases[, phenoVariables],
                      id.vars=names(completeCases)[which(names(completeCases) %in% phenoIDs)])
```

```
## Warning: attributes are not identical across measure variables; they will be
## dropped
```

```
levels(aim1$variable) <- c("ADOS", "SRS", "Motor Overflow", "Inattention",
                          "Hyperactivity", "Age", "GAI")

aim1G <- group_by(aim1, PrimaryDiagnosis, Motion.Exclusion.Level, Included, variable)
```

Fit univariate GAMs.

We used univariate models rather than a model with all covariates simultaneously because some of the variables are correlated, such that the impact of each variable on rs-fMRI usability may be difficult to estimate. These models are related to the propensity models that will be used in the estimation of the deconfounded group difference.

We used univariate models rather than a model with all covariates simultaneously because some of the variables are correlated, such that the impact of each variable on rs-fMRI usability may be difficult to estimate. These models are related to the propensity models that will be used in the estimation of the deconfounded group difference.

```

aim1$delta = rep(NA,length=nrow(aim1))
aim1$delta = ifelse(aim1$Included=="Included",1,0)
aim1tib <- tibble(filter(aim1, Motion.Exclusion.Level!="None"))
aim1tib$Motion.Exclusion.Level <- droplevels(aim1tib$Motion.Exclusion.Level)
aim1Nest <- aim1tib %>%
  group_by(Motion.Exclusion.Level, variable) %>%
  tidyr::nest()
#nested models
nested_gams <- aim1Nest %>%
  mutate(model = map(data, ~mgcv::gam(1-delta~s(value, k=-1), data = na.omit(.x),
                                     family=binomial(link=logit), method="REML")),
         coefs = map(model, tidy, conf.int = FALSE),
         Rsq = map_dbl(model, ~summary(.)$r.sq)) %>%
  unnest(coefs)
#Ben: correct for 7 lenient and 7 strict
nested_gams_len <- nested_gams %>%
  filter(Motion.Exclusion.Level=="Lenient")
nested_gams_len$p.fdr = p.adjust(nested_gams_len$p.value, method = "BH")
nested_gams_strict <- nested_gams %>%
  filter(Motion.Exclusion.Level=="Strict")

nested_gams_strict$p.fdr = p.adjust(nested_gams_strict$p.value, method = "BH")
#combine to print
nested_gams <- rbind(nested_gams_len, nested_gams_strict)
#list adjusted p values
nested_gams[, c(1:2,6:11)]

```

```

## # A tibble: 14 × 8
## # Groups:   Motion.Exclusion.Level, variable [14]
##   Motion.Exclusion.Level variable   edf ref.df statistic p.value   Rsq   p.fdr
##   <fct>                <fct>   <dbl> <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 Lenient              ADOS     1.54  1.87    15.3  1.44e-3 0.0303 2.77e-3
## 2 Lenient              SRS      1.78  2.23    14.9  8.95e-4 0.0411 2.77e-3
## 3 Lenient              Motor O... 1.00  1.00    15.6  7.58e-5 0.0308 5.31e-4
## 4 Lenient              Inatten... 1.56  1.93     7.83  1.41e-2 0.0167 1.41e-2
## 5 Lenient              Hyperac... 1.88  2.36    13.1  2.94e-3 0.0251 4.12e-3
## 6 Lenient              Age      1.00  1.00     8.17  4.28e-3 0.0142 4.99e-3
## 7 Lenient              GAI      1.00  1.00     9.98  1.59e-3 0.0188 2.77e-3
## 8 Strict               ADOS     1.00  1.00    20.3  7.46e-6 0.0426 2.25e-5
## 9 Strict               SRS      1.00  1.00    21.0  5.10e-6 0.0547 2.25e-5
## 10 Strict              Motor O... 1.00  1.00    10.2  1.42e-3 0.0194 1.99e-3
## 11 Strict              Inatten... 1.00  1.00    16.1  5.96e-5 0.0339 1.04e-4
## 12 Strict              Hyperac... 1.65  2.05    23.4  9.64e-6 0.0516 2.25e-5
## 13 Strict              Age      1.87  2.34     8.05  2.93e-2 0.0139 2.93e-2
## 14 Strict              GAI      1.00  1.00     5.90  1.51e-2 0.0101 1.77e-2

```

```

#max p value for 7 lenient models
max(nested_gams_len$p.fdr)

```

```
## [1] 0.01409489
```

```
#max p value for 7 strict models
max(nested_gams_strict$p.fdr)
```

```
## [1] 0.02926959
```

```
nested_gams <- nested_gams %>%
  mutate(LB = map(data, ~round(min(na.omit(.x$value)))),
         UB = map(data, ~round(max(na.omit(.x$value)))),
         range = map2(LB, UB, ~seq(from=.x, to=.y, by=1)),
         logpredict = map2(model, range, ~predict(.x, newdata = data.frame(valu
e = .y), type="link",se.fit=TRUE)),
         fit = map(logpredict, ~plogis(.x$fit)),
         lCI = map(logpredict, ~plogis(.x$fit-1.96*.x$se.fit)),
         hCI = map(logpredict, ~plogis(.x$fit+1.96*.x$se.fit)))
```

Define theme for Figure 4 top row

```
gam_theme = theme(
  axis.title.x=element_text(size=12),
  axis.title.y=element_text(size=12),
  axis.text.x=element_text(size=8),
  axis.text.y=element_text(size=10),
  plot.title = element_text(size = 16),
  plot.caption = element_text(size = 16,hjust = 0),
  legend.title = element_blank(), legend.position = "none")
```

Figure 4a top. Probability of exclusion as a function of ADOS

```
ados <- nested_gams %>%
  filter(variable=="ADOS") %>%
  select("variable", "Motion.Exclusion.Level", "range", "fit", 'lCI', 'hCI') %>%
  unnest(c(range, fit, lCI, hCI))
p_ados <- ggplot(ados, aes(x=range, y=fit))+
  geom_line(aes(colour = Motion.Exclusion.Level),size=1.2)+ylim(0,1)+theme_bw()+
  geom_ribbon(aes(ymin=lCI, ymax=hCI, fill=Motion.Exclusion.Level), linetype='blank', al
pha=0.2)+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154","#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154","#9FB0CC"))+
  labs(x='', y='Probability of Exclusion', fill='Motion Control', colour='Motion Contro
l')+
  scale_x_continuous(expand = c(0, 0))+
  gam_theme+
  ggtitle("ADOS")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))
```

Figure 4b top. Probability of exclusion as a function of SRS

```

srs <- nested_gams %>%
  filter(variable=="SRS") %>%
  select("variable", "Motion.Exclusion.Level", "range", "fit", 'lCI', 'hCI') %>%
  unnest(c(range, fit, lCI, hCI))
p_srs <- ggplot(srs, aes(x=range, y=fit))+
  geom_line(aes(colour = Motion.Exclusion.Level),size=1.2)+ylim(0,1)+theme_bw()+
  geom_ribbon(aes(ymin=lCI, ymax=hCI, fill=Motion.Exclusion.Level), linetype='blank', al
pha=0.2)+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  labs(x='', y='', fill='Motion Control', colour='Motion Control')+
  scale_x_continuous(expand = c(0, 0))+
  gam_theme+
  ggtitle("SRS")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))+
  theme(axis.title.y = element_blank())

```

Figure 4c top. Probability of exclusion as a function of Inattention

```

ina <- nested_gams %>%
  filter(variable=="Inattention") %>%
  select("variable", "Motion.Exclusion.Level", "range", "fit", 'lCI', 'hCI') %>%
  unnest(c(range, fit, lCI, hCI))
p_in <- ggplot(ina, aes(x=range, y=fit))+
  geom_line(aes(colour = Motion.Exclusion.Level),size=1.2)+ylim(0,1)+theme_bw()+
  geom_ribbon(aes(ymin=lCI, ymax=hCI, fill=Motion.Exclusion.Level), linetype='blank', al
pha=0.2)+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  labs(x='', y='', fill='Motion Control', colour='Motion Control')+
  scale_x_continuous(expand = c(0, 0))+
  gam_theme+
  ggtitle("Inattention")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))+
  theme(axis.title.y = element_blank())

```

Figure 4d top. Probability of exclusion as a function of Hyperactivity


```

hi <- nested_gams %>%
  filter(variable=="Hyperactivity") %>%
  select("variable", "Motion.Exclusion.Level", "range", "fit", 'lCI', 'hCI') %>%
  unnest(c(range, fit, lCI, hCI))
p_hi <- ggplot(hi, aes(x=range, y=fit))+
  geom_line(aes(colour = Motion.Exclusion.Level),size=1.2)+ylim(0,1)+theme_bw()+
  geom_ribbon(aes(ymin=lCI, ymax=hCI, fill=Motion.Exclusion.Level), linetype='blank', al
pha=0.2)+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154","#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154","#9FB0CC"))+
  labs(x='', y='', fill='Motion Control', colour='Motion Control')+
  scale_x_continuous(expand = c(0, 0))+
  gam_theme+
  ggtitle("Hyperactivity")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))+
  theme(axis.title.y = element_blank())

```

Figure 4e top. Probability of exclusion as a function of Motor Overflow

```

mo <- nested_gams %>%
  filter(variable=="Motor Overflow") %>%
  select("variable", "Motion.Exclusion.Level", "range", "fit", 'lCI', 'hCI') %>%
  unnest(c(range, fit, lCI, hCI))
p_mo <- ggplot(mo, aes(x=range, y=fit))+
  geom_line(aes(colour = Motion.Exclusion.Level),size=1.2)+ylim(0,1)+theme_bw()+
  geom_ribbon(aes(ymin=lCI, ymax=hCI, fill=Motion.Exclusion.Level), linetype='blank', al
pha=0.2)+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154","#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154","#9FB0CC"))+
  labs(x='', y='', fill='Motion Control', colour='Motion Control')+
  scale_x_continuous(expand = c(0, 0))+
  gam_theme+
  ggtitle("Motor Overflow")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))+
  theme(axis.title.y = element_blank())

```

Figure 4f top. Probability of exclusion as a function of Age

```

age<- nested_gams %>%
  filter(variable=="Age") %>%
  select("variable", "Motion.Exclusion.Level", "range", "fit", 'lCI', 'hCI') %>%
  unnest(c(range, fit, lCI, hCI))
p_age <- ggplot(age, aes(x=range, y=fit))+
  geom_line(aes(colour = Motion.Exclusion.Level),size=1.2)+ylim(0,1)+theme_bw()+
  geom_ribbon(aes(ymin=lCI, ymax=hCI, fill=Motion.Exclusion.Level), linetype='blank', al
pha=0.2)+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  labs(x='', y='', fill='Motion Control', colour='Motion Control')+
  scale_x_continuous(expand = c(0, 0))+
  gam_theme+
  ggtitle("Age")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))+
  theme(axis.title.y = element_blank())

```

Figure 4g top. Probability of exclusion as a function of GAI

```

gai <- nested_gams %>%
  filter(variable=="GAI") %>%
  select("variable", "Motion.Exclusion.Level", "range", "fit", 'lCI', 'hCI') %>%
  unnest(c(range, fit, lCI, hCI))
p_gai <- ggplot(gai, aes(x=range, y=fit))+
  geom_line(aes(colour = Motion.Exclusion.Level),size=1.2)+ylim(0,1)+theme_bw()+
  geom_ribbon(aes(ymin=lCI, ymax=hCI, fill=Motion.Exclusion.Level), linetype='blank', al
pha=0.2)+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  labs(x='', y='', fill='Motion Control', colour='Motion Control')+
  scale_x_continuous(expand = c(0, 0))+
  gam_theme+
  ggtitle("GAI")+
  theme(plot.title = element_text(size = 11, hjust = 0.5))+
  theme(axis.title.y = element_blank())
p_legend = cowplot::get_legend(p_gai + guides(color = guide_legend(nrow = 1))+
  theme(legend.position = "bottom", legend.text = element
_text(size = 11),
  legend.key.size=unit(.15, "in")))

```

Figure 4 bottom row: Density plots of covariates used to fit GAMs (across included & excluded children)

Define theme for density plots of covariates across included and excluded children

Figure 4a bottom. ADOS density

```

ddata <- nested_gams %>%
  filter(variable=="ADOS") %>%
  filter(Motion.Exclusion.Level=="Lenient") %>%
  select("variable", "Motion.Exclusion.Level", "data", "LB", "UB") %>%
  unnest(c(data, LB, UB)) %>%
  filter(PrimaryDiagnosis=="ASD")

ddata$PrimaryDiagnosis <- droplevels(ddata$PrimaryDiagnosis)

d_ados=ggplot(ddata, aes(x=value, fill=PrimaryDiagnosis, color=PrimaryDiagnosis))+
  geom_density(alpha=0.5, inherit.aes=TRUE)+
  scale_x_continuous(expand = c(0, 0), limits = c(ddata$LB[1], ddata$UB[1]), breaks=seq(
0, ddata$UB[1], by=5))+
  scale_y_continuous(expand = c(0, 0), limits = c(0, .09), breaks=seq(0, .08, by=.02))+
  labs(x='', y='Density')+
  scale_fill_manual(values = c("#FDE599"))+
  scale_color_manual(values = c("#E9D38D"))+
  den_theme

```

Figure 4b bottom. SRS density

```

ddata <- nested_gams %>%
  filter(variable=="SRS") %>%
  filter(Motion.Exclusion.Level=="Lenient") %>%
  select("variable", "Motion.Exclusion.Level", "data") %>%
  unnest(data)

d_srs=ggplot(ddata, aes(x=value, fill=PrimaryDiagnosis, color=PrimaryDiagnosis))+
  geom_density(alpha=0.5, inherit.aes=TRUE)+
  scale_x_continuous(expand = c(0, 0), limits=c(0,max(srs$range)),breaks = seq(0, 100 ,
by = 50))+
  scale_y_continuous(expand = c(0, 0))+
  scale_fill_manual(labels=c('TD','ASD'), values = c("#009E73", "#FDE599"))+
  scale_color_manual(labels=c('TD','ASD'), values = c("#05634a", "#E9D38D"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  den_theme+
  theme(axis.title.y = element_blank())+
  labs(x='')

```

Figure 4c bottom. Inattention density

```

ddata <- nested_gams %>%
  filter(variable=="Inattention") %>%
  filter(Motion.Exclusion.Level=="Lenient") %>%
  select("variable", "Motion.Exclusion.Level", "data") %>%
  unnest(data)

d_in=ggplot(ddata, aes(x=value, fill=PrimaryDiagnosis, color=PrimaryDiagnosis))+
  geom_density(alpha=0.5, inherit.aes=TRUE)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  scale_fill_manual(labels=c('TD','ASD'), values = c("#009E73", "#FDE599"))+
  scale_color_manual(labels=c('TD','ASD'), values = c("#05634a", "#E9D38D"))+
  labs(x='')+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  theme(axis.title.y = element_blank())+
  den_theme+
  theme(axis.title.y = element_blank())+
  labs(x='')

```

Figure 4d bottom. Hyperactivity/Impulsivity Density

```

ddata <- nested_gams %>%
  filter(variable=="Hyperactivity") %>%
  filter(Motion.Exclusion.Level=="Lenient") %>%
  select("variable", "Motion.Exclusion.Level", "data") %>%
  unnest(data)

d_hi=ggplot(ddata, aes(x=value, fill=PrimaryDiagnosis, color=PrimaryDiagnosis))+
  geom_density(alpha=0.5, inherit.aes=TRUE)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  scale_fill_manual(labels=c('TD','ASD'), values = c("#009E73", "#FDE599"))+
  scale_color_manual(labels=c('TD','ASD'), values = c("#05634a", "#E9D38D"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  den_theme+
  theme(axis.title.y = element_blank())+
  labs(x='')

```

Figure 4e bottom. Motor Overflow Density

```

ddata <- nested_gams %>%
  filter(variable=="Motor Overflow") %>%
  filter(Motion.Exclusion.Level=="Lenient") %>%
  select("variable", "Motion.Exclusion.Level", "data") %>%
  unnest(data)

d_mo=ggplot(ddata, aes(x=value, fill=PrimaryDiagnosis, color=PrimaryDiagnosis))+
  geom_density(alpha=0.5, inherit.aes=TRUE)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  theme(axis.title.y = element_blank())+
  scale_fill_manual(labels=c('TD','ASD'), values = c("#009E73", "#FDE599"))+
  scale_color_manual(labels=c('TD','ASD'), values = c("#05634a", "#E9D38D"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  den_theme+
  theme(axis.title.y = element_blank())+
  labs(x='')

```

Figure 4f bottom. Age Density

```

ddata <- nested_gams %>%
  filter(variable=="Age") %>%
  filter(Motion.Exclusion.Level=="Lenient") %>%
  select("variable", "Motion.Exclusion.Level", "data") %>%
  unnest(data)

d_age=ggplot(ddata, aes(x=value, fill=PrimaryDiagnosis, color=PrimaryDiagnosis))+
  geom_density(alpha=0.5, inherit.aes=TRUE)+
  scale_x_continuous(expand = c(0, 0), limits=c(8,13), breaks = seq(8, 13 , by = 1))+
  scale_y_continuous(expand = c(0, 0), limits=c(0,.29), breaks=seq(0, .25, by = .05))+
  scale_fill_manual(labels=c('TD','ASD'), values = c("#009E73", "#FDE599"))+
  scale_color_manual(labels=c('TD','ASD'), values = c("#05634a", "#E9D38D"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  den_theme+
  theme(axis.title.y = element_blank())+
  labs(x='')

```

Figure 4g bottom. GAI Density

```

ddata <- nested_gams %>%
  filter(variable=="GAI") %>%
  filter(Motion.Exclusion.Level=="Lenient") %>%
  select("variable", "Motion.Exclusion.Level", "data") %>%
  unnest(data)

d_gai=ggplot(ddata, aes(x=value, fill=PrimaryDiagnosis, color=PrimaryDiagnosis))+
  geom_density(alpha=0.5, inherit.aes=TRUE)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0), limits = c(0, .035), breaks=seq(0., .03, by=.01))
+
  scale_fill_manual(labels=c('TD','ASD'), values = c("#009E73", "#FDE599"))+
  scale_color_manual(labels=c('TD','ASD'), values = c("#05634a", "#E9D38D"))+

  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  den_theme+
  labs(x='')+
  theme(axis.title.y = element_blank())

hist_legend = cowplot::get_legend(d_gai + guides(color = guide_legend(nrow = 1)))+theme(
legend.position = "bottom", legend.text = element_text(size = 11), legend.key.size=unit
(.15, "in"))

```

combine gam plots with densities & print

```

top_row <- cowplot::plot_grid(p_ados, p_srs, p_in, p_hi, p_mo, p_age, p_gai, ncol=7, rel
_widths=c(1.18/7, .97/7, .97/7, .97/7, .97/7, .97/7, .97/7))
bottom_row <- cowplot::plot_grid(d_ados, d_srs, d_in, d_hi, d_mo, d_age, d_gai, ncol=7,
rel_widths=c(1.18/7, .97/7, .97/7, .97/7, .97/7, .97/7, .97/7))

```

```
## Warning: Removed 89 rows containing non-finite values (stat_density).
```

```

png("./CovariatesAndRS-fMRIUsability/fig_probExclusion_allGAM_TD_ASD_cc.png",width=10,he
ight=6,units="in",res=200)
cowplot::plot_grid(p_legend, top_row, NULL, bottom_row, NULL, hist_legend, nrow=6, rel_h
eights=c(.1, 1, -.01, .5, -.07, .1))
dev.off()

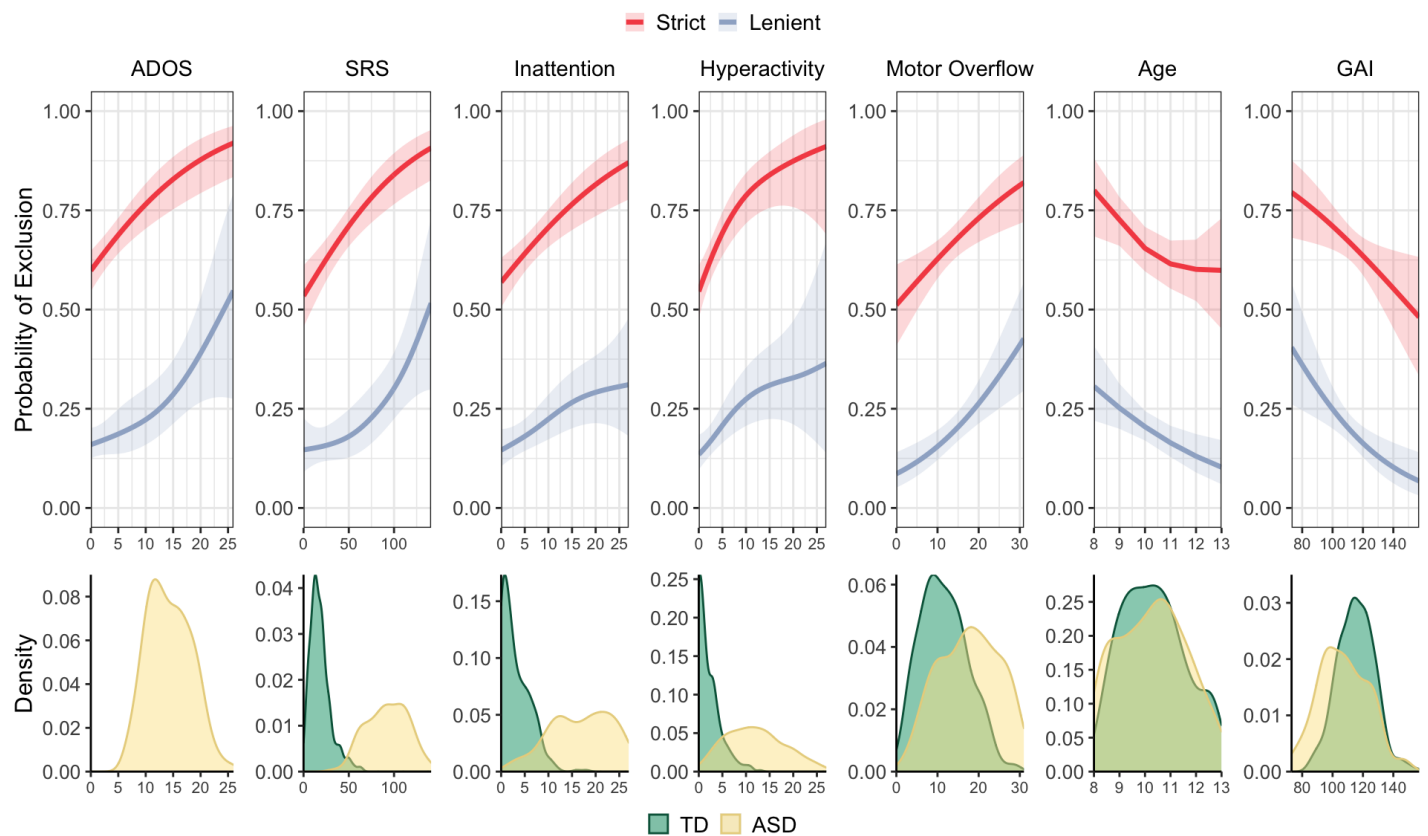
```

```
## quartz_off_screen
## 2
```

```

#png("~/Dropbox/Apps/Overleaf/MotionSelectionBias_rsFMRI/Figures/fig_probExclusion_allGA
M_TD_ASD_cc.png",width=10,height=6,units="in",res=200)
cowplot::plot_grid(p_legend, top_row, NULL, bottom_row, NULL, hist_legend, nrow=6, rel_h
eights=c(.1, 1, -.05, .5, -.07, .1))

```



```
#dev.off()
```

NOTE: 19 rows = # of participants with missing Motor Overflow scores (imputed for the DRTMLE of the deconfounded group difference)

Figure 4. rs-fMRI exclusion probability changes with phenotype and age but not SES. Univariate analysis of rs-fMRI exclusion probability as a function of participant characteristics. From left to right: Autism Diagnostic Observation Schedule (ADOS), social responsiveness scale (SRS) scores, inattentive symptoms, hyperactive/impulsive symptoms, total motor overflow, age, general ability index (GAI), and socioeconomic status (SES) using the lenient (light blue lines, all FDR-adjusted $p < 0.01$), and strict (red lines) motion quality control (all FDR-adjusted $p < 0.03$). Variable distributions for each diagnosis group (included and excluded scans) are displayed across the bottom panel (TD=typically developing, green; ASD=autism spectrum disorder, yellow).

3.1.3. Phenotype and age representations differ between included and excluded children

3.1.3. Mann-Whitney U tests to compare included vs excluded participants using lenient motion QC (13 tests)

```

#run lenient tests first
aimltib <- tibble(aiml)

aimlMW <- aimltib %>%
  filter(Motion.Exclusion.Level=="Lenient") %>%
  group_by(PrimaryDiagnosis, variable) %>%
  tidyr::nest()

#hypothesis: included children will have less severe symptoms. NOTE: ADOS only collected
in ASD (9 tests)
nested_mw_less <- aimlMW %>%
  filter(variable %in% c("SRS", "Inattention", "Hyperactivity",
    "Motor Overflow")|(variable=="ADOS"&PrimaryDiagnosis=="ASD")) %>%
  mutate(mwm = map(data, ~wilcox.test(value~Included, alternative="less", correct = TRUE
, data = na.omit(.x))),
    idata = map(data, ~filter(., Included=="Included")),
    edata = map(data, ~filter(., Included=="Excluded")),
    includedMedian = map(idata, ~median(.x$value, na.rm=TRUE)),
    excludedMedian = map(edata, ~median(.x$value, na.rm=TRUE)),
    coefs = map(mwm, tidy),
    Za = map(coefs, ~qnorm(.x$p.value)),
    r = map2(Za, data, ~(abs(.x)/sqrt(length(na.omit(.y$value)))))) %>%
  unnest(c(coefs, includedMedian, excludedMedian, r))

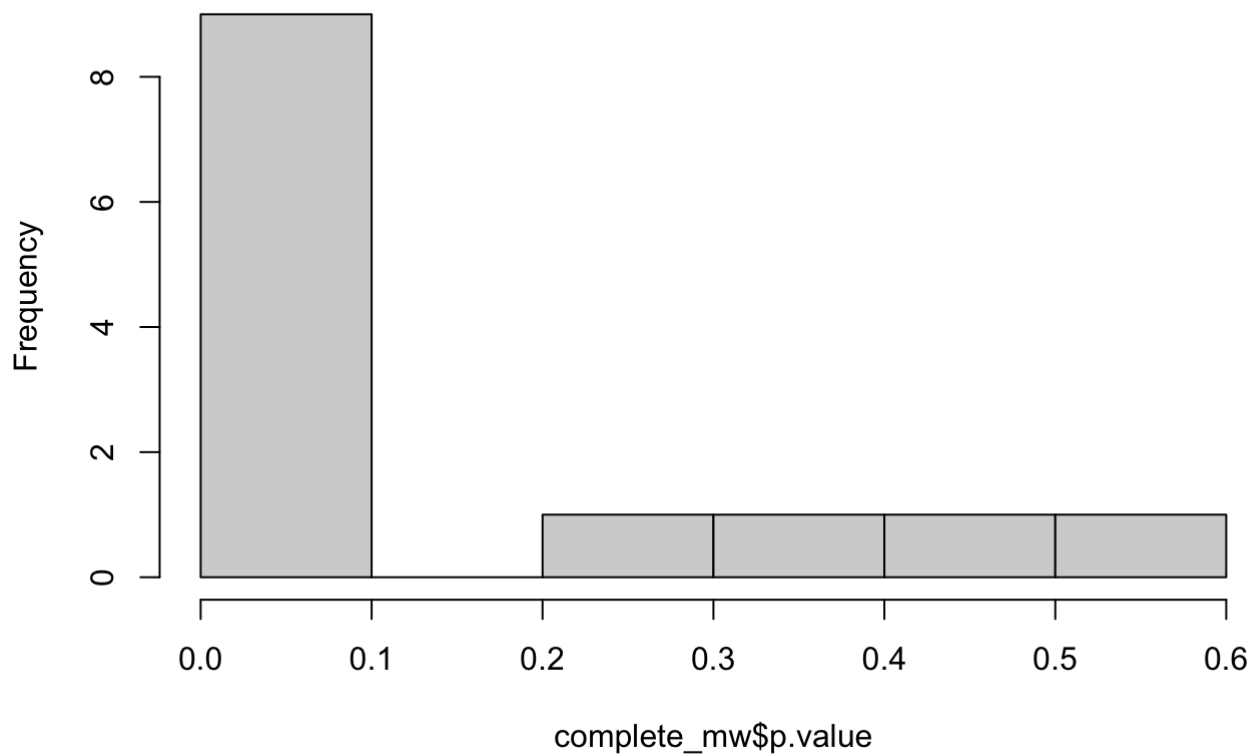
#hypothesis: included children will be older and have greater GAI (4 tests)
nested_mw_greater<- aimlMW %>%
  filter(variable %in% c("Age", "GAI")) %>%
  mutate(mwm = map(data, ~wilcox.test(value~Included, alternative="greater", correct = T
RUE, data = na.omit(.x))),
    idata = map(data, ~filter(., Included=="Included")),
    edata = map(data, ~filter(., Included=="Excluded")),
    includedMedian = map(idata, ~median(.x$value, na.rm=TRUE)),
    excludedMedian = map(edata, ~median(.x$value, na.rm=TRUE)),
    coefs = map(mwm, tidy),
    Za = map(coefs, ~qnorm(.x$p.value)),
    r = map2(Za, data, ~(abs(.x)/sqrt(length(na.omit(.y$value)))))) %>%
  unnest(c(coefs, includedMedian, excludedMedian, r))

complete_mw <- rbind(nested_mw_less, nested_mw_greater)

hist(complete_mw$p.value,
  main = "Mann-Whitney U test p values Using Lenient Motion QC")

```


Mann-Whitney U test p values Using Lenient Motion QC



```
complete_mw$p.fdr <- p.adjust(complete_mw$p.value, method = "BH")  
names(complete_mw)[which(names(complete_mw)=="statistic")]="U"  
  
#complete_mw[order(complete_mw$PrimaryDiagnosis, decreasing=TRUE), c(1:2, 7:9, 13)]  
  
#sort by q-value/p.fdr to ease interpretation of fdr adjusted p values  
complete_mw[order(complete_mw$p.fdr, decreasing=FALSE), c(1:2, 7:10, 14:15)]
```

```
## # A tibble: 13 × 8
## # Groups:   PrimaryDiagnosis, variable [13]
##   PrimaryDiagnosis variable includedMedian excludedMedian      U p.value      r
##   <fct>           <fct>           <dbl>           <dbl> <dbl> <dbl> <dbl>
## 1 ASD             Motor O...         17              22    1260 9.48e-4 2.65e-1
## 2 ASD             SRS                91.2            100   1439 1.22e-2 1.92e-1
## 3 TD              Age                10.3            9.86  9757 1.10e-2 1.23e-1
## 4 ASD             ADOS               13              16    1522 3.16e-2 1.59e-1
## 5 TD              Hyperac...         1               2    6833 2.31e-2 1.07e-1
## 6 TD              GAI                116             112.   9490. 2.84e-2 1.02e-1
## 7 ASD             GAI                108             101   2280 3.93e-2 1.50e-1
## 8 ASD             Age                10.5            9.76  2263 4.68e-2 1.43e-1
## 9 TD              Motor O...         11              13    7086. 5.69e-2 8.48e-2
## 10 TD             Inatten...         2               2    7743 2.63e-1 3.40e-2
## 11 ASD            Hyperac...         11              12    1832. 3.53e-1 3.22e-2
## 12 TD             SRS                16              15    4464. 4.96e-1 5.65e-4
## 13 ASD            Inatten...         18              16    1936. 5.48e-1 1.02e-2
## # ... with 1 more variable: p.fdr <dbl>
```

#for the paper, sort by PrimaryDiagnosis

```
xtable(complete_mw[order(complete_mw$PrimaryDiagnosis, decreasing=TRUE), c(1:2, 7:8, 14:
15)])
```

```
## % latex table generated in R 4.1.2 by xtable 1.8-4 package
## % Fri Mar 25 19:19:57 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rllrrrr}
## \hline
## & PrimaryDiagnosis & variable & includedMedian & excludedMedian & r & p.fdr \\
## \hline
## 1 & ASD & ADOS & 13.00 & 16.00 & 0.16 & 0.07 \\
## 2 & ASD & SRS & 91.25 & 100.00 & 0.19 & 0.05 \\
## 3 & ASD & Motor Overflow & 17.00 & 22.00 & 0.27 & 0.01 \\
## 4 & ASD & Inattention & 18.00 & 16.00 & 0.01 & 0.55 \\
## 5 & ASD & Hyperactivity & 11.00 & 12.00 & 0.03 & 0.42 \\
## 6 & ASD & Age & 10.54 & 9.76 & 0.14 & 0.08 \\
## 7 & ASD & GAI & 108.00 & 101.00 & 0.15 & 0.07 \\
## 8 & TD & SRS & 16.00 & 15.00 & 0.00 & 0.54 \\
## 9 & TD & Motor Overflow & 11.00 & 13.00 & 0.08 & 0.08 \\
## 10 & TD & Inattention & 2.00 & 2.00 & 0.03 & 0.34 \\
## 11 & TD & Hyperactivity & 1.00 & 2.00 & 0.11 & 0.07 \\
## 12 & TD & Age & 10.32 & 9.87 & 0.12 & 0.05 \\
## 13 & TD & GAI & 116.00 & 112.50 & 0.10 & 0.07 \\
## \hline
## \end{tabular}
## \end{table}
```

```
#xtable(complete_mw[order(complete_mw$p.fdr, decreasing=FALSE), c(1:2, 7:10, 13)])

#aimlp = complete_mw[, c("PrimaryDiagnosis, variable", "includedMedian", "excludedMedia
n",
#                               "U", "p.value", "r", "p.fdr")]

aimlp = complete_mw[, c(1:2, 7:8, 10, 15)]

aimlp$Motion.Exclusion.Level = rep("Lenient", nrow(aimlp))
```

9 of the 13 tests have an FDR-adjusted p value < .2. We expect roughly 1.8 of these to be a false positive.

3.1.3 Supplemental Table S2. Mann-Whitney U tests to compare included vs excluded participants using strict motion QC (13 tests)

```

#run tests using strict motion QC
aim1MW <- aim1tib %>%
  filter(Motion.Exclusion.Level=="Strict") %>%
  group_by(variable, PrimaryDiagnosis) %>%
  tidyr::nest()

#hypothesis: included children will have less severe symptoms
nested_mw_less <- aim1MW %>%
  filter(variable %in% c("SRS", "Inattention", "Hyperactivity",
    "Motor Overflow")|(variable=="ADOS"&PrimaryDiagnosis=="ASD")) %>%
  mutate(mwm = map(data, ~wilcox.test(value~Included, alternative="less", data = na.omit
    (.x))),
    idata = map(data, ~filter(., Included=="Included")),
    edata = map(data, ~filter(., Included=="Excluded")),
    includedMedian = map(idata, ~median(.x$value, na.rm=TRUE)),
    excludedMedian = map(edata, ~median(.x$value, na.rm=TRUE)),
    coefs = map(mwm, tidy),
    Za = map(coefs, ~qnorm(.x$p.value)),
    r = map2(Za, data, ~(abs(.x)/sqrt(length(na.omit(.y$value)))))) %>%
  unnest(c(coefs, includedMedian, excludedMedian, r))

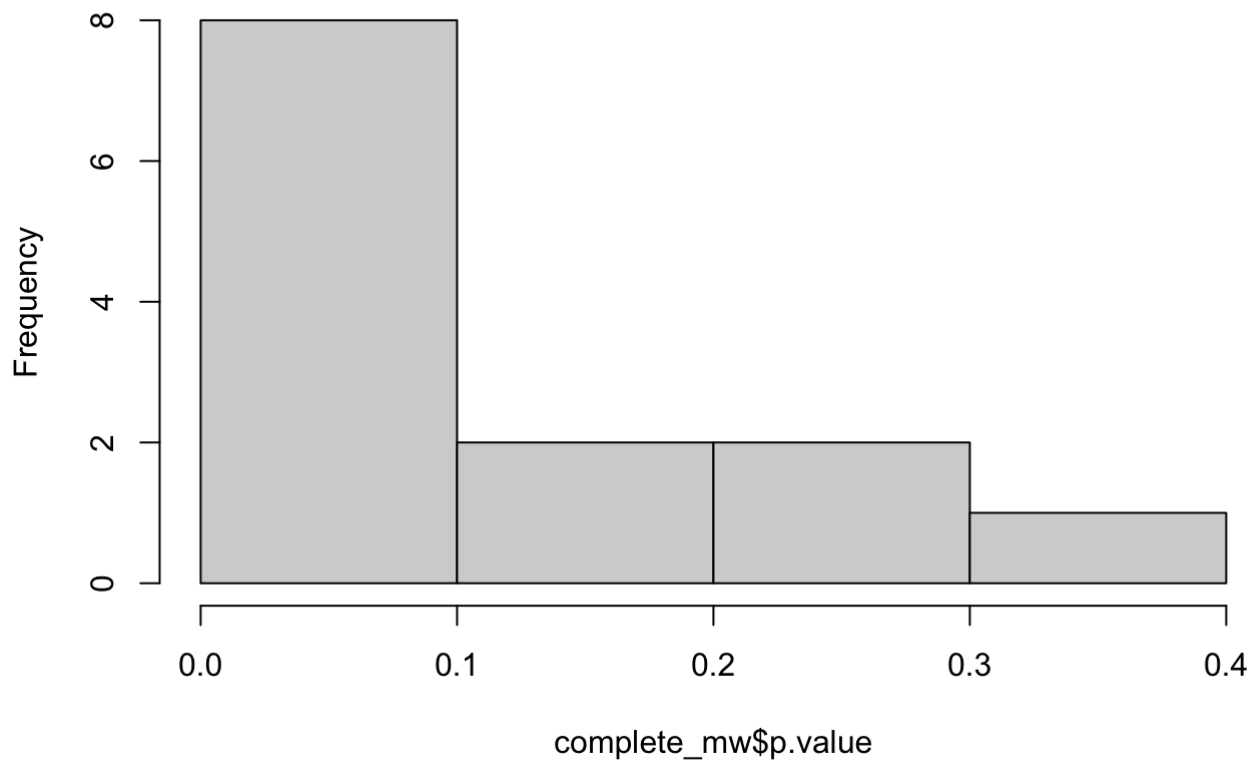
#hypothesis: included children will be older and have greater GAI
nested_mw_greater<- aim1MW %>%
  filter(variable %in% c("Age", "GAI")) %>%
  mutate(mwm = map(data, ~wilcox.test(value~Included, alternative="greater", data = na.o
    mit(.x))),
    idata = map(data, ~filter(., Included=="Included")),
    edata = map(data, ~filter(., Included=="Excluded")),
    includedMedian = map(idata, ~median(.x$value, na.rm=TRUE)),
    excludedMedian = map(edata, ~median(.x$value, na.rm=TRUE)),
    coefs = map(mwm, tidy),
    Za = map(coefs, ~qnorm(.x$p.value)),
    r = map2(Za, data, ~(abs(.x)/sqrt(length(na.omit(.y$value)))))) %>%
  unnest(c(coefs, includedMedian, excludedMedian, r))

complete_mw <- rbind(nested_mw_less, nested_mw_greater)

hist(complete_mw$p.value,
  main = "Mann-Whitney U test p values Using Strict Motion QC")

```

Mann-Whitney U test p values Using Strict Motion QC



```
complete_mw$p.fdr <- p.adjust(complete_mw$p.value, method = "BH")  
  
names(complete_mw)[which(names(complete_mw)=="statistic")]="U"  
  
#sort by q-value/p.fdr to ease interpretation of fdr adjusted p values  
complete_mw[order(complete_mw$p.fdr, decreasing=FALSE), c(1:2, 7:10, 14:15)]
```

```
## # A tibble: 13 × 8
## # Groups:   PrimaryDiagnosis, variable [13]
##   PrimaryDiagnosis variable includedMedian excludedMedian      U p.value      r
##   <fct>           <fct>           <dbl>           <dbl> <dbl> <dbl> <dbl>
## 1 TD              Hyperac...           1             2 11884. 0.00163 0.158
## 2 ASD             Age                 11.0          10.1 1880 0.00829 0.205
## 3 ASD             ADOS                12.5          14   1080. 0.0231 0.170
## 4 ASD             SRS                 84.2          93.5 1080. 0.0231 0.170
## 5 TD             SRS                 14.2          17   7047 0.0516 0.101
## 6 ASD             Motor O...          15           18   1132. 0.0437 0.146
## 7 TD             Age                 10.4          10.1 16000. 0.0545 0.0859
## 8 TD             GAI                 116           115  15797 0.0833 0.0742
## 9 TD             Motor O...          11           12   13570. 0.149 0.0557
## 10 TD            Inatten...          2             2   13598. 0.154 0.0546
## 11 ASD            Inatten...          15.5          18   1306 0.227 0.0641
## 12 ASD            Hyperac...          11            12   1297 0.212 0.0683
## 13 ASD            GAI                 108.          107  1491 0.397 0.0223
## # ... with 1 more variable: p.fdr <dbl>
```

#for the paper, sort by PrimaryDiagnosis

```
xtable(complete_mw[order(complete_mw$PrimaryDiagnosis, decreasing=TRUE), c(1:2, 7:8, 14:
15)])
```

```
## % latex table generated in R 4.1.2 by xtable 1.8-4 package
## % Fri Mar 25 19:19:57 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{rllrrrr}
## \hline
## & PrimaryDiagnosis & variable & includedMedian & excludedMedian & r & p.fdr \\
## \hline
## 1 & ASD & ADOS & 12.50 & 14.00 & 0.17 & 0.08 \\
## 2 & ASD & SRS & 84.25 & 93.50 & 0.17 & 0.08 \\
## 3 & ASD & Motor Overflow & 15.00 & 18.00 & 0.15 & 0.10 \\
## 4 & ASD & Inattention & 15.50 & 18.00 & 0.06 & 0.25 \\
## 5 & ASD & Hyperactivity & 11.00 & 12.00 & 0.07 & 0.25 \\
## 6 & ASD & Age & 11.04 & 10.13 & 0.20 & 0.05 \\
## 7 & ASD & GAI & 107.50 & 107.00 & 0.02 & 0.40 \\
## 8 & TD & SRS & 14.25 & 17.00 & 0.10 & 0.10 \\
## 9 & TD & Motor Overflow & 11.00 & 12.00 & 0.06 & 0.20 \\
## 10 & TD & Inattention & 2.00 & 2.00 & 0.05 & 0.20 \\
## 11 & TD & Hyperactivity & 1.00 & 2.00 & 0.16 & 0.02 \\
## 12 & TD & Age & 10.38 & 10.13 & 0.09 & 0.10 \\
## 13 & TD & GAI & 116.00 & 115.00 & 0.07 & 0.14 \\
## \hline
## \end{tabular}
## \end{table}
```

```
#xtable(complete_mw[order(complete_mw$p.fdr, decreasing=FALSE), c(1:2, 7:10, 13)])

temp = complete_mw[, c(1:2,7:8,14:15)]
temp$Motion.Exclusion.Level = rep("Strict", nrow(temp))

aimlp <- rbind(aimlp, temp)
aimlp$p.fdr = round(aimlp$p.fdr, 3)
aimlp$p.signif = rep("", nrow(aimlp))
aimlp$p.signif[aimlp$p.fdr<.2]="^"
aimlp$p.signif[aimlp$p.fdr<.1]="*"
aimlp$p.signif[aimlp$p.fdr<.05]="**"

#think I need this for add_pvalue?
aimlp$Included = rep("Included", nrow(aimlp))
```

Figure 5. Split violin plots

```
My_Theme = theme_light()+theme(
  legend.title = element_blank(),
  axis.title.x = element_blank(),
  axis.title.y = element_blank(),
  axis.text.x = element_text(size = 5),
  axis.text.y = element_text(size = 8),
  strip.text.x = element_text(size = 12, face = "bold", color="black"),
  strip.text.y = element_text(size = 10, color="black"),
  strip.background = element_rect(fill="white"),
  plot.title = element_text(size = 9, hjust = 0.5))
```

Figure 5a. ADOS (ASD only) split violin

NOTE: Missing ADOS scores for one participant evaluated at CARD (no included in complete predictor cases)

```

ados <- aim1 %>%
  filter(PrimaryDiagnosis=="ASD" & variable=="ADOS") %>%
  dplyr::select(-c("PrimaryDiagnosis", "variable"))

stat.test <- aim1p %>% filter(PrimaryDiagnosis=="ASD" & variable=="ADOS")
stat.test$group1 = rep("Included", nrow(stat.test))
stat.test$group2 = rep("null model", nrow(stat.test))
stat.test <- select(stat.test, -p.fdr)
stat.test$y.position = c(27, 27)

paper_ados <- ggplot(ados, aes(Motion.Exclusion.Level, value, fill = Included, color = I
ncluded)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE, adjust = 1.5) +
  stat_summary(fun = "mean", position = position_dodge(width = 0.5),
              color="black", geom="point", aes(y=value))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.5),
              color="black", geom="errorbar", width=.2)+
  #geom_mark_rect(aes(filter = Motion.Exclusion.Level == "Lenient"))+
  ggprism::add_pvalue(stat.test,
                    x = "Motion.Exclusion.Level",
                    y.position = "y.position",
                    color="black",
                    label.size = 6)+
  scale_y_continuous(limits=c(0,30),breaks = seq(0, 30 , by = 10))+
  scale_fill_manual(values = c("#FDE599", "#9FB0CC"))+
  scale_color_manual(values = c("#E9D38D", "#8C9AB4"))+
  My_Theme+
  theme(legend.text = element_text(size = 8))+
  theme(legend.position = "none")+
  ggtitle("ADOS")

```

Checking the positivity assumption for DRTMLE (reported in Discussion section of the paper)

```

adosMaxAll = ados %>% filter(Motion.Exclusion.Level=="Lenient" & Included=="Excluded") %
>% select(value) %>% max()
adosMaxUsable = ados %>% filter(Motion.Exclusion.Level=="Lenient" & Included=="Included"
) %>% select(value) %>% max()

```

The highest ADOS score among included children using lenient motion QC is 23; among all children, 26.

Figure 5b. SRS split violin


```

srs <- filter(aimlG, variable=="SRS")

stat.test <- aimlp %>% filter(variable=="SRS")
stat.test$group1 = rep("Included", nrow(stat.test))
stat.test$group2 = rep("null model", nrow(stat.test))
stat.test <- select(stat.test, -p.fdr)
stat.test$y.position = c(142, 70, 142, 70)

aiml_srs <- ggplot(srs, aes(Motion.Exclusion.Level, value, fill = Included, color=Included)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE) +
  facet_grid(PrimaryDiagnosis~., scales = "fixed")+
  stat_summary(fun = "mean", position = position_dodge(width = 0.5),
              color="black", geom="point", aes(y=value))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.5),
              color="black", geom="errorbar", width=.2)+
  ggprism::add_pvalue(stat.test,
                    x = "Motion.Exclusion.Level",
                    y.position = "y.position",
                    color="black",
                    label.size = 6)+
  scale_y_continuous(limits=c(0, 150),breaks = seq(0, 100 , by = 50))+
  # geom_jitter(position=position_jitterdodge(jitter.height = .25), alpha = .3)+
  # geom_pointrange(
  # data = summary_pheno,
  # aes(Motion.Exclusion.Level, mean, ymin=min, ymax=max),
  # shape = 20,
  # position = position_dodge(width = 0.9))+
  scale_fill_manual(values = c("#FDE599", "#9FB0CC"))+
  scale_color_manual(values = c("#E9D38D", "#8C9AB4"))+
  My_Theme+
  theme(strip.text.y =element_blank()+
  theme(legend.position = "none")+
  ggtitle("SRS")

v_legend = cowplot::get_legend(aiml_srs + guides(color = guide_legend(nrow = 2)))+
  theme(legend.position = "left",
        legend.text = element_text(size = 10),
        legend.key.size=unit(.1, "in"))

```

```
## Warning: Removed 267 rows containing non-finite values (stat_ydensity).
```

```
## Warning: Removed 267 rows containing non-finite values (stat_summary).
```

```
## Warning: Removed 267 rows containing non-finite values (stat_summary).
```

NOTE: 267/3 = 89, # of participants missing SRS

```
with(dat3, table(PrimaryDiagnosis, is.na(SRS.Score)))
```

```
##
## PrimaryDiagnosis FALSE TRUE
##           TD      259    89
##           ASD      137     0
```

Figure 5c. Inattention Split Violin

```
inat <- filter(aim1G, variable=="Inattention")

stat.test <- aim1p %>% filter(variable=="Inattention")
stat.test$group1 = rep("Included", nrow(stat.test))
stat.test$group2 = rep("null model", nrow(stat.test))
stat.test <- select(stat.test, -p.fdr)
stat.test$y.position = c(20, 20, 20, 20)

paper_inat <- ggplot(inat, aes(Motion.Exclusion.Level, value, fill = Included, color=Included)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE) +
  facet_grid(PrimaryDiagnosis~.)+
  stat_summary(fun = "mean", position = position_dodge(width = 0.5),
              color="black", geom="point", aes(y=value))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.5),
              color="black", geom="errorbar", width=.2)+
  ggprism::add_pvalue(stat.test,
                    x = "Motion.Exclusion.Level",
                    y.position = "y.position",
                    color="black",
                    label.size = 6)+
  # geom_jitter(position=position_jitterdodge(jitter.height = .25), alpha = .3)+
  #geom_pointrange(
  # data = summary_pheno,
  # aes(Motion.Exclusion.Level, mean, ymin=min, ymax=max),
  # shape = 20,
  # position = position_dodge(width = 0.9))+
  scale_fill_manual(values = c("#FDE599", "#9FB0CC"))+
  scale_color_manual(values = c("#E9D38D", "#8C9AB4"))+
  My_Theme+
  theme(strip.text.y = element_blank())+
  theme(legend.position = "none")+
  ggtitle("Inattention")
```

Figure 5d. Hyperactivity/Impulsivity Spilt Violin

```

hyp <- filter(aim1G, variable=="Hyperactivity")

stat.test <- aim1p %>% filter(variable=="Hyperactivity")
stat.test$group1 = rep("Included", nrow(stat.test))
stat.test$group2 = rep("null model", nrow(stat.test))
stat.test <- select(stat.test, -p.fdr)
stat.test$y.position = c(16, 16, 16, 16)

paper_hyp <- ggplot(hyp, aes(Motion.Exclusion.Level, value, fill = Included, color=Included)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE) +
  facet_grid(PrimaryDiagnosis~.)+
  stat_summary(fun = "mean", position = position_dodge(width = 0.5),
              color="black", geom="point", aes(y=value))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.5),
              color="black", geom="errorbar", width=.2)+
  ggprism::add_pvalue(stat.test,
                    x = "Motion.Exclusion.Level",
                    y.position = "y.position",
                    color="black",
                    label.size = 6)+
  # geom_jitter(position=position_jitterdodge(jitter.height = .25), alpha = .3)+
  #geom_pointrange(
  # data = summary_pheno,
  # aes(Motion.Exclusion.Level, mean, ymin=min, ymax=max),
  # shape = 20,
  # position = position_dodge(width = 0.9))+
  scale_fill_manual(values = c("#FDE599", "#9FB0CC"))+
  scale_color_manual(values = c("#E9D38D", "#8C9AB4"))+
  My_Theme+
  theme(strip.text.y = element_blank())+
  theme(legend.position = "none")+
  ggtitle("Hyperactivity")

```

Figure 5e. Motor Overflow Split Violin

```

overflow <- filter(aim1G, variable=="Motor Overflow")

stat.test <- aim1p %>% filter(variable=="Motor Overflow")
stat.test$group1 = rep("Included", nrow(stat.test))
stat.test$group2 = rep("null model", nrow(stat.test))
stat.test <- select(stat.test, -p.fdr)
stat.test$y.position = c(32, 31, 32, 31)

aim1_of <- ggplot(overflow, aes(Motion.Exclusion.Level, value, fill = Included, color=Included)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE) +
  facet_grid(PrimaryDiagnosis~.)+
  stat_summary(fun = "mean", position = position_dodge(width = 0.5),
              color="black", geom="point", aes(y=value))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.5),
              color="black", geom="errorbar", width=.2)+
  ggprism::add_pvalue(stat.test,
                    x = "Motion.Exclusion.Level",
                    y.position = "y.position",
                    color="black",
                    label.size = 6)+
  scale_y_continuous(limits=c(0,35),breaks = seq(0, 40 , by = 10))+
  # geom_jitter(position=position_jitterdodge(jitter.height = .25), alpha = .3)+
  #geom_pointrange(
  # data = summary_pheno,
  # aes(Motion.Exclusion.Level, mean, ymin=min, ymax=max),
  # shape = 20,
  # position = position_dodge(width = 0.9))+
  scale_fill_manual(values = c("#FDE599", "#9FB0CC"))+
  scale_color_manual(values = c("#E9D38D", "#8C9AB4"))+
  My_Theme+
  theme(strip.text.y =element_blank()+
  theme(legend.position = "none")+
  ggtitle("Motor Overflow")

```

Figure 5f. Age Split Violin

```

age <- filter(aim1G, variable=="Age")

stat.test <- aim1p %>% filter(variable=="Age")
stat.test$group1 = rep("Included", nrow(stat.test))
stat.test$group2 = rep("null model", nrow(stat.test))
stat.test <- select(stat.test, -p.fdr)
stat.test$y.position = c(13.1, 13.1, 13.1, 13.1)

aim1_age <- ggplot(age, aes(Motion.Exclusion.Level, value, fill = Included, color=Included)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE) +
  facet_grid(PrimaryDiagnosis~.)+
  stat_summary(fun = "mean", position = position_dodge(width = 0.5),
              color="black", geom="point", aes(y=value))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.5),
              color="black", geom="errorbar", width=.2)+
  ggprism::add_pvalue(stat.test,
                    x = "Motion.Exclusion.Level",
                    y.position = "y.position",
                    color="black",
                    label.size = 6)+
  scale_y_continuous(limits=c(8,14),breaks = seq(8, 14 , by = 1))+
  # geom_jitter(position=position_jitterdodge(jitter.height = .25), alpha = .3)+
  #geom_pointrange(
  # data = summary_pheno,
  # aes(Motion.Exclusion.Level, mean, ymin=min, ymax=max),
  # shape = 20,
  # position = position_dodge(width = 0.9))+
  scale_fill_manual(values = c("#FDE599", "#9FB0CC"))+
  scale_color_manual(values = c("#E9D38D", "#8C9AB4"))+
  My_Theme+
  theme(strip.text.y =element_blank()+
  theme(legend.position = "none")+
  ggtitle("Age")

```

Figure 5g. GAI Split Violin

```

gai <- filter(aim1G, variable=="GAI")

stat.test <- aim1p %>% filter(variable=="GAI")
stat.test$group1 = rep("Included", nrow(stat.test))
stat.test$group2 = rep("null model", nrow(stat.test))
stat.test <- select(stat.test, -p.fdr)
stat.test$y.position = c(155, 160, 155, 160)

aim1_gai <- ggplot(gai, aes(Motion.Exclusion.Level, value, fill = Included, color=Included)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE) +
  facet_grid(PrimaryDiagnosis~.)+
  stat_summary(fun = "mean", position = position_dodge(width = 0.5),
              color="black", geom="point", aes(y=value))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.5),
              color="black", geom="errorbar", width=.2)+
  ggprism::add_pvalue(stat.test,
                    x = "Motion.Exclusion.Level",
                    y.position = "y.position",
                    color="black",
                    label.size = 6)+
  scale_y_continuous(limits=c(75,165),breaks = seq(80, 160, by = 20))+
  # geom_jitter(position=position_jitterdodge(jitter.height = .25), alpha = .3)+
  #geom_pointrange(
  # data = summary_pheno,
  # aes(Motion.Exclusion.Level, mean, ymin=min, ymax=max),
  # shape = 20,
  # position = position_dodge(width = 0.9))+
  scale_fill_manual(values = c("#FDE599", "#9FB0CC"))+
  scale_color_manual(values = c("#E9D38D", "#8C9AB4"))+
  My_Theme+
  theme(legend.position = "none")+
  ggtitle("GAI")

```

Combine split violins into one figure & save

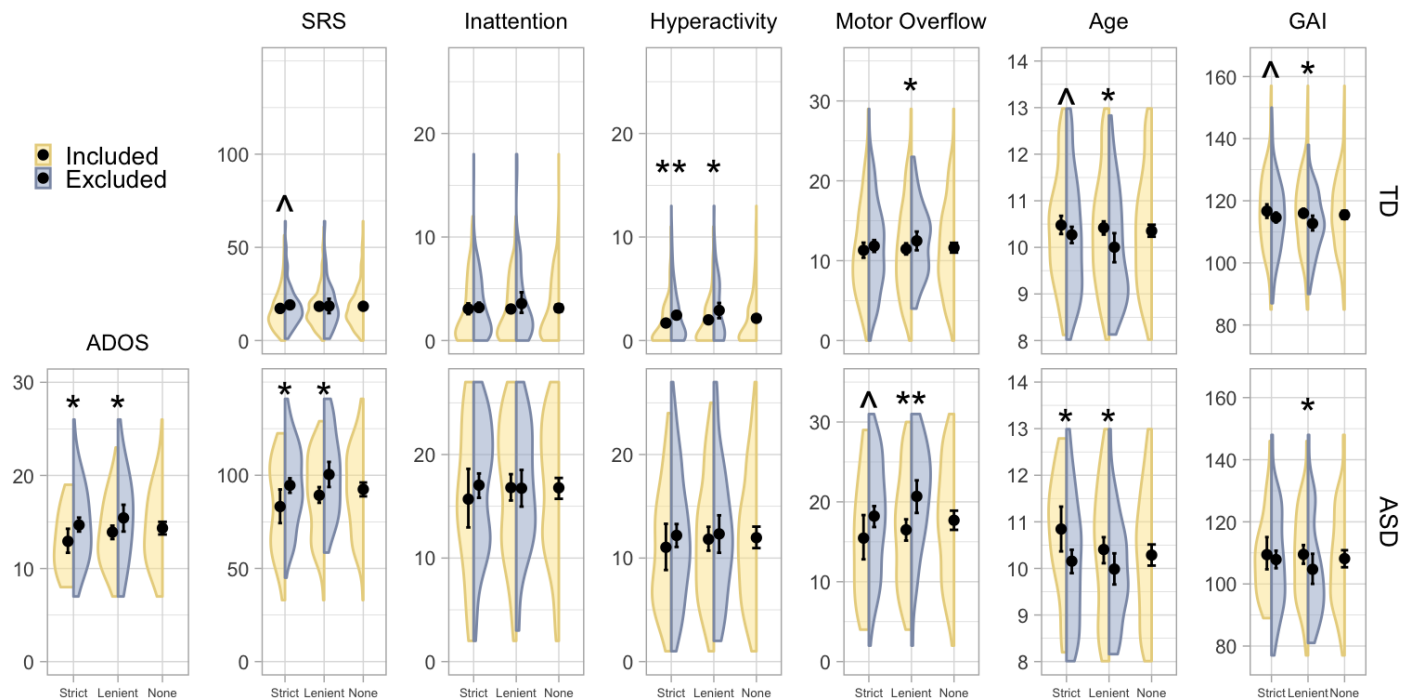


Figure 5. Participants with usable rs-fMRI data differed from participants with unusable rs-fMRI data. Comparison of Autism Diagnostic Observation Schedule (ADOS) scores, social responsiveness scale (SRS) scores, inattentive symptoms, hyperactive/impulsive symptoms, motor overflow, age, and general ability index (GAI) for included (yellow) and excluded (slate blue) participants stratified by diagnosis group and motion exclusion level. The deconfounded mean integrates across the diagnosis-specific distribution of usable and unusable covariates, which here is labeled as None. Mean values are indicated by a black dot; 95% bootstrap confidence intervals are indicated with black bars. We controlled for 13 comparisons performed for the lenient and strict motion QC cases using the false discovery rate (FDR). ** indicate differences between included and excluded participants with an FDR-adjusted p value <0.05; * indicate FDR-adjusted p values <0.1. ^ indicate FDR-adjusted p values <0.2. A larger number of significant differences are observed using the lenient motion QC than the strict motion QC, but very few participants pass strict motion QC. autism spectrum disorder (ASD), typically developing (TD).

3.1.4. Functional connectivity as a function of phenotype and age

Combine partial correlations with melted rs-fMRI usability and covariate info

```
startEdgeidx = which(names(dat3)=='r.ic1.ic2')
endEdgeidx = which(names(dat3)=='r.ic29.ic30')
names(dat3)[startEdgeidx:endEdgeidx]
```

```

## [1] "r.ic1.ic2" "r.ic1.ic4" "r.ic1.ic8" "r.ic1.ic13" "r.ic1.ic14"
## [6] "r.ic1.ic15" "r.ic1.ic17" "r.ic1.ic19" "r.ic1.ic21" "r.ic1.ic22"
## [11] "r.ic1.ic24" "r.ic1.ic25" "r.ic1.ic26" "r.ic1.ic27" "r.ic1.ic28"
## [16] "r.ic1.ic29" "r.ic1.ic30" "r.ic2.ic4" "r.ic2.ic8" "r.ic2.ic13"
## [21] "r.ic2.ic14" "r.ic2.ic15" "r.ic2.ic17" "r.ic2.ic19" "r.ic2.ic21"
## [26] "r.ic2.ic22" "r.ic2.ic24" "r.ic2.ic25" "r.ic2.ic26" "r.ic2.ic27"
## [31] "r.ic2.ic28" "r.ic2.ic29" "r.ic2.ic30" "r.ic4.ic8" "r.ic4.ic13"
## [36] "r.ic4.ic14" "r.ic4.ic15" "r.ic4.ic17" "r.ic4.ic19" "r.ic4.ic21"
## [41] "r.ic4.ic22" "r.ic4.ic24" "r.ic4.ic25" "r.ic4.ic26" "r.ic4.ic27"
## [46] "r.ic4.ic28" "r.ic4.ic29" "r.ic4.ic30" "r.ic8.ic13" "r.ic8.ic14"
## [51] "r.ic8.ic15" "r.ic8.ic17" "r.ic8.ic19" "r.ic8.ic21" "r.ic8.ic22"
## [56] "r.ic8.ic24" "r.ic8.ic25" "r.ic8.ic26" "r.ic8.ic27" "r.ic8.ic28"
## [61] "r.ic8.ic29" "r.ic8.ic30" "r.ic13.ic14" "r.ic13.ic15" "r.ic13.ic17"
## [66] "r.ic13.ic19" "r.ic13.ic21" "r.ic13.ic22" "r.ic13.ic24" "r.ic13.ic25"
## [71] "r.ic13.ic26" "r.ic13.ic27" "r.ic13.ic28" "r.ic13.ic29" "r.ic13.ic30"
## [76] "r.ic14.ic15" "r.ic14.ic17" "r.ic14.ic19" "r.ic14.ic21" "r.ic14.ic22"
## [81] "r.ic14.ic24" "r.ic14.ic25" "r.ic14.ic26" "r.ic14.ic27" "r.ic14.ic28"
## [86] "r.ic14.ic29" "r.ic14.ic30" "r.ic15.ic17" "r.ic15.ic19" "r.ic15.ic21"
## [91] "r.ic15.ic22" "r.ic15.ic24" "r.ic15.ic25" "r.ic15.ic26" "r.ic15.ic27"
## [96] "r.ic15.ic28" "r.ic15.ic29" "r.ic15.ic30" "r.ic17.ic19" "r.ic17.ic21"
## [101] "r.ic17.ic22" "r.ic17.ic24" "r.ic17.ic25" "r.ic17.ic26" "r.ic17.ic27"
## [106] "r.ic17.ic28" "r.ic17.ic29" "r.ic17.ic30" "r.ic19.ic21" "r.ic19.ic22"
## [111] "r.ic19.ic24" "r.ic19.ic25" "r.ic19.ic26" "r.ic19.ic27" "r.ic19.ic28"
## [116] "r.ic19.ic29" "r.ic19.ic30" "r.ic21.ic22" "r.ic21.ic24" "r.ic21.ic25"
## [121] "r.ic21.ic26" "r.ic21.ic27" "r.ic21.ic28" "r.ic21.ic29" "r.ic21.ic30"
## [126] "r.ic22.ic24" "r.ic22.ic25" "r.ic22.ic26" "r.ic22.ic27" "r.ic22.ic28"
## [131] "r.ic22.ic29" "r.ic22.ic30" "r.ic24.ic25" "r.ic24.ic26" "r.ic24.ic27"
## [136] "r.ic24.ic28" "r.ic24.ic29" "r.ic24.ic30" "r.ic25.ic26" "r.ic25.ic27"
## [141] "r.ic25.ic28" "r.ic25.ic29" "r.ic25.ic30" "r.ic26.ic27" "r.ic26.ic28"
## [146] "r.ic26.ic29" "r.ic26.ic30" "r.ic27.ic28" "r.ic27.ic29" "r.ic27.ic30"
## [151] "r.ic28.ic29" "r.ic28.ic30" "r.ic29.ic30"

```

```

signalFC <- dat3[, c(1,startEdgeidx:endEdgeidx)]

dat2 <- merge(aim1, signalFC, all=TRUE, by = "ID")
fcMelt <- reshape2::melt(dat2[,1:162],
                        id.vars=names(dat2)[1:9],
                        variable.name = "edge",
                        value.name = "fc")

```

3.1.4. Run nested gams


```
fctib <- tibble(filter(fcMelt, Motion.Exclusion.Level!="None"))
fctib$Motion.Exclusion.Level <- droplevels(fctib$Motion.Exclusion.Level)

fcNest <- fctib %>%
  filter(Included=="Included") %>%
  group_by(variable, Motion.Exclusion.Level, edge) %>%
  tidyr::nest()

#nested models
fc_gams <- fcNest %>%
  mutate(model = map(data, ~mgcv::gam(fc~s(value), data = na.omit(.x), method="REML")),
         coefs = map(model, tidy, conf.int = FALSE)) %>%
  unnest(coefs)
```

Figure 6 Plot histograms of edgewise p-values from GAMs

Figure 6a. Histogram of edgewise p-values for partial correlations as a function of ADOS.

NOTE: TD scores = 0

```
pdat <- fc_gams %>%
  ungroup() %>%
  filter(variable=="ADOS") %>%
  select(Motion.Exclusion.Level, p.value) %>%
  group_by(Motion.Exclusion.Level)

hist_ados_p=ggplot(pdat, aes(x=p.value, fill=Motion.Exclusion.Level, color=Motion.Exclusion.Level))+
  geom_histogram(position = "identity", alpha=0.5, inherit.aes=TRUE, binwidth = .05)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  My_Theme+
  ggtitle("ADOS")+
  theme(plot.title = element_text(size = 9, hjust = 0.5))+
  theme(legend.position = "none")+
  labs(x='p value', y='Count')+
  theme(axis.title.y = element_text(size = 9, angle=90))+
  theme(axis.title.x = element_text(size = 7))
```

Figure 6b. Histogram of edgewise p-values for partial correlations as a function of SRS

```

pdat <- fc_gams %>%
  ungroup() %>%
  filter(variable=="SRS") %>%
  select(Motion.Exclusion.Level, p.value) %>%
  group_by(Motion.Exclusion.Level)

hist_srs_p=ggplot(pdat, aes(x=p.value, fill=Motion.Exclusion.Level, color=Motion.Exclusion.Level))+
  geom_histogram(position = "identity", alpha=0.5, inherit.aes=TRUE, binwidth = .05)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  labs(x='p value', y='')+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  My_Theme+
  theme(axis.title.y = element_blank())+
  ggtitle("SRS")+
  theme(plot.title = element_text(size = 9, hjust = 0.5))+
  theme(legend.position = "none")+
  theme(axis.title.x = element_text(size = 7))

```

Figure 6c. Histogram of p-values for partial correlations as a function of inattention

```

pdat <- fc_gams %>%
  ungroup() %>%
  filter(variable=="Inattention") %>%
  select(Motion.Exclusion.Level, p.value) %>%
  group_by(Motion.Exclusion.Level)

hist_in_p=ggplot(pdat, aes(x=p.value, fill=Motion.Exclusion.Level, color=Motion.Exclusion.Level))+
  geom_histogram(position = "identity", alpha=0.5, inherit.aes=TRUE, binwidth = .05)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  labs(x='p value', y='')+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  My_Theme+
  theme(axis.title.y = element_blank())+
  ggtitle("Inattention")+
  theme(plot.title = element_text(size = 9, hjust = 0.5))+
  theme(legend.position = "none")+
  theme(axis.title.x = element_text(size = 7))

```

Number of edges with a p value < 0.05 for SRS: ##### Figure 6d. Histogram of p-values for partial correlations as a function of hyperactivity

```

pdat <- fc_gams %>%
  ungroup() %>%
  filter(variable=="Hyperactivity") %>%
  select(Motion.Exclusion.Level, p.value) %>%
  group_by(Motion.Exclusion.Level)

hist_hi_p=ggplot(pdat, aes(x=p.value, fill=Motion.Exclusion.Level, color=Motion.Exclusion.Level))+
  geom_histogram(position = "identity", alpha=0.5, inherit.aes=TRUE, binwidth = .05)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  labs(x='p value', y='')+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  My_Theme+
  theme(axis.title.y = element_blank())+
  ggtitle("Hyperactivity")+
  theme(plot.title = element_text(size = 9, hjust = 0.5))+
  theme(legend.position = "none")+
  theme(axis.title.x = element_text(size = 7))

```

Figure 6e. Histogram of p-values for partial correlations as a function of motor overflow

```

pdat <- fc_gams %>%
  ungroup() %>%
  filter(variable=="Motor Overflow") %>%
  select(Motion.Exclusion.Level, p.value) %>%
  group_by(Motion.Exclusion.Level)

hist_mo_p=ggplot(pdat, aes(x=p.value, fill=Motion.Exclusion.Level, color=Motion.Exclusion.Level))+
  geom_histogram(position = "identity", alpha=0.5, inherit.aes=TRUE, binwidth = .05)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  labs(x='p value', y='')+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  My_Theme+
  theme(axis.title.y = element_blank())+
  ggtitle("Motor Overflow")+
  theme(plot.title = element_text(size = 9, hjust = 0.5))+
  theme(legend.position = "none")+
  theme(axis.title.x = element_text(size = 7))

```

Figure 6f. Histogram of p-values for partial correlations as a function of age

```

pdat <- fc_gams %>%
  ungroup() %>%
  filter(variable=="Age") %>%
  select(Motion.Exclusion.Level, p.value) %>%
  group_by(Motion.Exclusion.Level)

hist_age_p=ggplot(pdat, aes(x=p.value, fill=Motion.Exclusion.Level, color=Motion.Exclusion.Level))+
  geom_histogram(position = "identity", alpha=0.5, inherit.aes=TRUE, binwidth = .05)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  labs(x='p value', y='')+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  My_Theme+
  theme(axis.title.y = element_blank())+
  ggtitle("Age")+
  theme(plot.title = element_text(size = 9, hjust = 0.5))+
  theme(legend.position = "none")+
  theme(axis.title.x = element_text(size = 7))

```

Figure 6g. Histogram of p-values for partial correlations as a function of GAI

```

pdat <- fc_gams %>%
  ungroup() %>%
  filter(variable=="GAI") %>%
  select(Motion.Exclusion.Level, p.value) %>%
  group_by(Motion.Exclusion.Level)

hist_gai_p=ggplot(pdat, aes(x=p.value, fill=Motion.Exclusion.Level, color=Motion.Exclusion.Level))+
  geom_histogram(position = "identity", alpha=0.5, inherit.aes=TRUE, binwidth = .05)+
  scale_x_continuous(expand = c(0, 0))+
  scale_y_continuous(expand = c(0, 0))+
  labs(x='p value', y='')+
  scale_color_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  scale_fill_manual(labels=c('Strict', 'Lenient'), values = c("#f55154", "#9FB0CC"))+
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), panel.background = element_blank(),
        axis.line = element_line(size=.5), panel.border = element_blank())+
  My_Theme+
  theme(axis.title.y = element_blank())+
  ggtitle("GAI")+
  theme(plot.title = element_text(size = 9, hjust = 0.5))+
  theme(legend.position = "none")+
  theme(axis.title.x = element_text(size = 7))

hist_p_legend <- cowplot::get_legend(hist_gai_p + guides(color = guide_legend(nrow = 1))
+theme(legend.position = "bottom", legend.text = element_text(size = 11), legend.key.size=unit(.15, "in"), legend.title = element_blank()))

```

Combine histograms and print

```

fc_hist <- cowplot::plot_grid(hist_ados_p, hist_srs_p, hist_in_p, hist_hi_p, hist_mo_p,
hist_age_p, hist_gai_p, ncol=7, rel_widths=c(1.18/7, .97/7, .97/7, .97/7, .97/7, .97/7,
.97/7))

png("./CovariatesAndRS-fMRIUsability/fig_hist_rfc_cc.png",width=8,height=3,units="in",res=200)

cowplot::plot_grid(fc_hist, hist_p_legend, nrow=2, rel_heights=c(1, .1))
dev.off()

```

```

## quartz_off_screen
##                2

```

```

cowplot::plot_grid(fc_hist, hist_p_legend, nrow=2, rel_heights=c(1, .1))

```

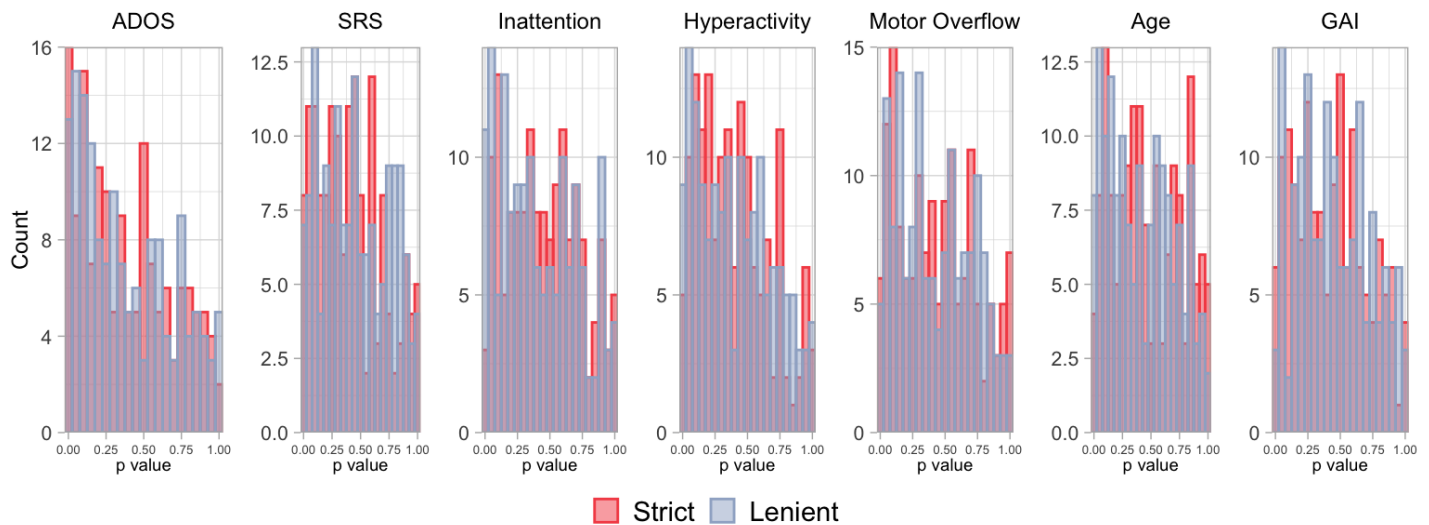


Figure 6. Some covariates related to rs-fMRI exclusion probability are also related to functional connectivity. *Histograms of p values for generalized additive models of the relationship between edgewise functional connectivity in participants with usable rs-fMRI data and (from left to right) ADOS, social responsiveness scale (SRS) scores, inattentive symptoms, hyperactive/impulsive symptoms, total motor overflow as assessed during the Physical and Neurological Exam for Subtle Signs, age, and general ability index (GAI). For a given covariate, a clustering of p values near zero suggests that covariate is associated with functional connectivity for a greater number of edges. Several covariates appear to be related to functional connectivity using both the lenient motion quality control (slate blue bins) and the strict motion quality control (red bins).

Create table summarizing number of edges showing a nominally significant relationship ($p < 0.05$ uncorrected) with each covariate, stratified by Primary Diagnosis

```
fc_gams_pCounts <- fc_gams %>%
  select(variable, Motion.Exclusion.Level, edge, p.value) %>%
  ungroup() %>%
  group_by(variable, Motion.Exclusion.Level) %>%
  summarise(pCount = sum(p.value < .05))
```

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

```
fc_gams_pCounts$clustering = rep("no clustering", nrow(fc_gams_pCounts))
fc_gams_pCounts$clustering[fc_gams_pCounts$pCount >= 10] = "some clustering"
fc_gams_pCounts$clustering[fc_gams_pCounts$pCount > 20] = "clustering"

fc_gams_pCounts
```

```
## # A tibble: 14 × 4
## # Groups:   variable [7]
##   variable      Motion.Exclusion.Level pCount clustering
##   <fct>         <fct>                <int> <chr>
## 1 ADOS          Strict                21 clustering
## 2 ADOS          Lenient              22 clustering
## 3 SRS           Strict                12 some clustering
## 4 SRS           Lenient              12 some clustering
## 5 Motor Overflow Strict                12 some clustering
## 6 Motor Overflow Lenient              10 some clustering
## 7 Inattention   Strict                 8 no clustering
## 8 Inattention   Lenient              19 some clustering
## 9 Hyperactivity Strict                11 some clustering
## 10 Hyperactivity Lenient              17 some clustering
## 11 Age          Strict                 8 no clustering
## 12 Age          Lenient              16 some clustering
## 13 GAI          Strict                12 some clustering
## 14 GAI          Lenient                 9 no clustering
```

```
#xtable(fc_gams_pCounts[order(fc_gams_pCounts$Motion.Exclusion.Level),])
```

NOTE: I used the following arbitrary cutoffs to describe Figure 6 in the section 3.1.4

no clustering < 10 edges

some evidence of clustering: between 10-20 edges

clustering: >20 edges

Extra. Impact of motion QC on framewise displacement metrics

```

#mean and max FD are different for lenient and strict because some frames at the beginning and/or end of the scan are excluded for scans to pass lenient motion QC
dat3$MeanFD.None = dat3$MeanFramewiseDisplacement
dat3$MaxFD.None = dat3$MaxFramewiseDisplacement

#same for all levels
dat3$FramesWithFDLessThanOrEqualTo250microns.None = dat3$FramesWithFDLessThanOrEqualTo250microns
dat3$FramesWithFDLessThanOrEqualTo250microns.KKI = dat3$FramesWithFDLessThanOrEqualTo250microns

meanFD = c("ID", "MeanFramewiseDisplacement", "MeanFramewiseDisplacement.KKI", "MeanFD.None")

maxFD = c("ID", "MaxFramewiseDisplacement", "MaxFramewiseDisplacement.KKI", "MaxFD.None")

framesFD = c("ID", "FramesWithFDLessThanOrEqualTo250microns", "FramesWithFDLessThanOrEqualTo250microns.KKI", "FramesWithFDLessThanOrEqualTo250microns.None")

fdID = c("ID")

meanFD.df <- reshape2::melt(dat3[, meanFD],
                           id.vars=names(dat3)[which(names(dat3) %in% fdID)],
                           variable.name = "Motion.Exclusion.Level",
                           value.name = "MeanFramewiseDisplacement")

levels(meanFD.df$Motion.Exclusion.Level)

```

```

## [1] "MeanFramewiseDisplacement"      "MeanFramewiseDisplacement.KKI"
## [3] "MeanFD.None"

```

```

#rename levels to match motion QC levels in completeCases
levels(meanFD.df$Motion.Exclusion.Level) <- c("Strict", "Lenient", "None")

#repeat for MaxFD
maxFD.df <- reshape2::melt(dat3[, maxFD],
                           id.vars=names(dat3)[which(names(dat3) %in% fdID)],
                           variable.name = "Motion.Exclusion.Level",
                           value.name = "MaxFramewiseDisplacement")

levels(maxFD.df$Motion.Exclusion.Level)

```

```

## [1] "MaxFramewiseDisplacement"      "MaxFramewiseDisplacement.KKI"
## [3] "MaxFD.None"

```



```

#rename levels to match motion QC levels in completeCases
levels(maxFD.df$Motion.Exclusion.Level) <- c("Strict", "Lenient", "None")

#merge meanFD.df and maxFD.df
fdMerg <- merge(meanFD.df, maxFD.df)

#repeat for FramesWithFDLessThanOrEqualTo250microns
frames.df <- reshape2::melt(dat3[, framesFD],
                           id.vars=names(dat3)[which(names(dat3) %in% fdID)],
                           variable.name = "Motion.Exclusion.Level",
                           value.name = "FramesWithFDLessThanOrEqualTo250microns")

levels(frames.df$Motion.Exclusion.Level)

```

```

## [1] "FramesWithFDLessThanOrEqualTo250microns"
## [2] "FramesWithFDLessThanOrEqualTo250microns.KKI"
## [3] "FramesWithFDLessThanOrEqualTo250microns.None"

```

```

#rename levels to match motion QC levels in qcMelt
levels(frames.df$Motion.Exclusion.Level) <- c("Strict", "Lenient", "None")

#merge with fdMerg
fdMerg <- merge(fdMerg, frames.df)

#merge with completeCases
completeCases <- merge(completeCases, fdMerg)

passOnly <- filter(completeCases, Included=="Included")

```

Filter out “None” from Motion.Exclusion.Level to make remaining group differences in FD metrics following motion QC easier to see

mean framewise displacement

```

passOnly <- filter(passOnly, Motion.Exclusion.Level!="None")
meanFD_violin <- ggplot(passOnly, aes(Motion.Exclusion.Level, MeanFrameWiseDisplacement,
                                     fill = PrimaryDiagnosis, color=PrimaryDiagnosis))
+
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE, adjust = 2) +
  stat_summary(fun = "mean", position = position_dodge(width = 0.18),
              color="black", geom="point", aes(y=MeanFrameWiseDisplacement))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.18),
              color="black", geom="errorbar", width=.15)+
  scale_fill_manual(values = c("#009E73", "#FDE599"))+
  scale_color_manual(values = c("#05634a", "#E9D38D"))+
  My_Theme_prop+theme(legend.title =element_blank()+
  theme(axis.text.y=element_text(size=8))+
  theme(legend.text = element_text(size = 7))+
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(legend.position = c(0.35, .7))+
  theme(legend.key.size=unit(.15, "in"))+
  theme(legend.box.margin = margin(-2, -2, -2, -2))+
  ggtitle("Mean FD")+
  theme(plot.title = element_text(size = 8, hjust=0.5))

```

max framewise displacement

```

maxFD_violin <- ggplot(passOnly, aes(Motion.Exclusion.Level, MaxFrameWiseDisplacement,
                                     fill = PrimaryDiagnosis, color=PrimaryDiagnosis)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE, adjust = 2) +
  stat_summary(fun = "mean", position = position_dodge(width = 0.18), color="black",
              geom="point", aes(y=MaxFrameWiseDisplacement))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.18),
              color="black", geom="errorbar", width=.15)+
  scale_fill_manual(values = c("#009E73", "#FDE599"))+
  scale_color_manual(values = c("#05634a", "#E9D38D"))+
  My_Theme_prop+
  theme(axis.text.y=element_text(size=8))+
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(legend.position = "none")+
  ggtitle("Max FD")+
  theme(plot.title = element_text(size = 8, hjust=0.5))

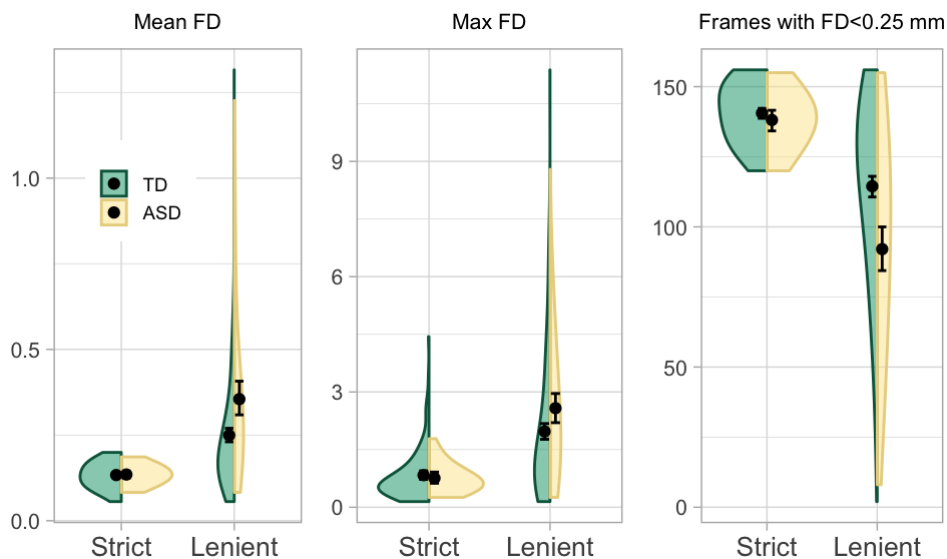
```

frames with FD <= 0.25 mm

```
frames_violin <- ggplot(passOnly, aes(Motion.Exclusion.Level, FramesWithFDLessThanOrEqual
lTo250microns,
                                fill = PrimaryDiagnosis, color=PrimaryDiagnosis)) +
  geom_split_violin(trim=TRUE, alpha = 0.5, inherit.aes = TRUE, adjust = 2) +
  stat_summary(fun = "mean", position = position_dodge(width = 0.18), color="black",
              geom="point", aes(y=FramesWithFDLessThanOrEqualTo250microns))+
  stat_summary(fun.data = "mean_cl_boot", position = position_dodge(width = 0.18),
              color="black", geom="errorbar", width=.15)+
  scale_fill_manual(values = c("#009E73", "#FDE599"))+
  scale_color_manual(values = c("#05634a", "#E9D38D"))+
  My_Theme_prop+
  theme(axis.text.y=element_text(size=8))+
  theme(axis.title.y = element_blank()+
  theme(axis.title.x = element_blank()+
  theme(legend.position = "none")+
  ggtitle("Frames with FD<0.25 mm")+
  theme(plot.title = element_text(size = 8, hjust=0.5))
```

Combine 3 FD plots with just lenient and strict Motion.Exclusion.Levels and save

```
## quartz_off_screen
##                2
```



Framewise displacement metrics are more similar across diagnosis groups using the strict motion QC, but very few participants are labeled as usable.