

ABCD STUDY: STUDY DESIGN, DATA SHARING & DEAP

Wesley K. Thompson | August 20, 2019

STUDY DESIGN

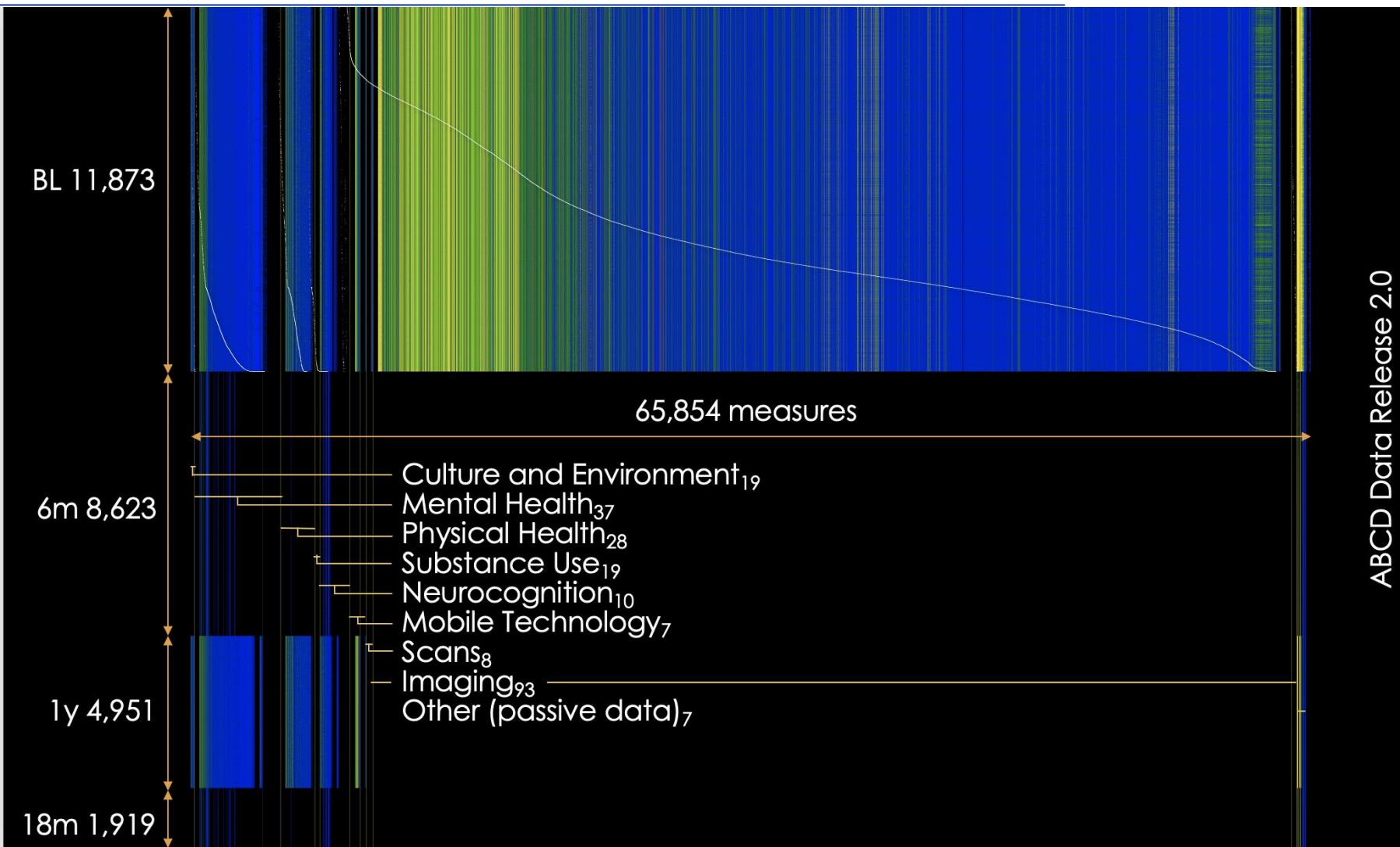
ABCD STUDY DESIGN

- The complete collection of baseline data was released on the NIMH Data Archive (NDA) in March 2019.
- Baseline data are assessed on 11,878 subjects at 21 sites around the country.
- There are also follow-up assessments on a minority of these subjects.

ABCD data dictionary (release 2.0)

27,400 x 65,000

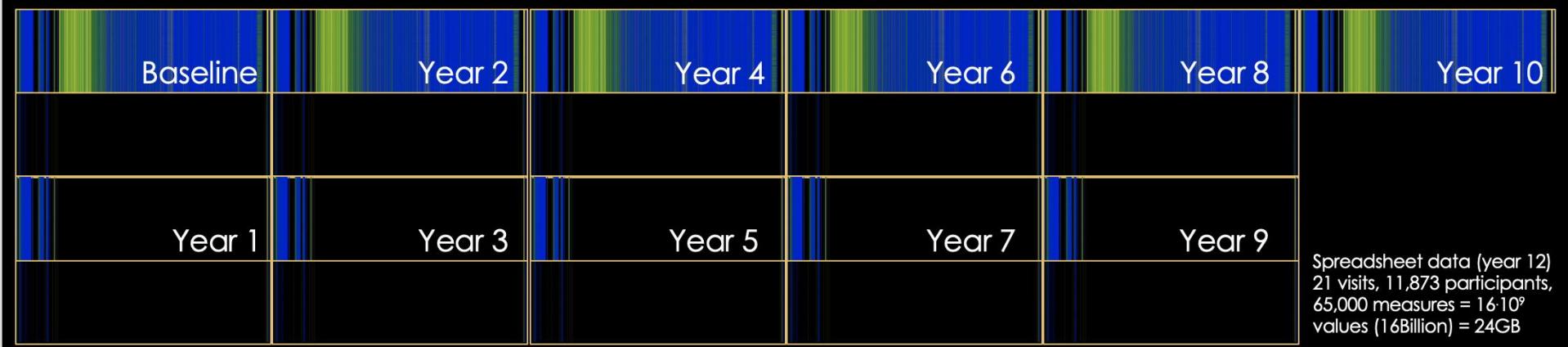
ABCD STUDY DESIGN (SHARED DATA IN 2.0)



ABCD STUDY DESIGN – DATA RELEASE SCHEDULE

| Release Year | Baseline | 6 month | 1 year | 18 month | 2 year | 36 month | 3 year | 48 month | 4 year | 60 month | 5 year | 72 month | 6 year | 84 month | 7 year | 96 month | 8 year | 108 month | 9 year | 120 month | 10 year | 132 month | 11 year | 144 month | 12 year |
|-----------------|----------|---------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|-----------|--------|-----------|---------|-----------|---------|-----------|---------|
| 1 | 4,951 | 0 | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | 11,873 | 8,623 | 4,951 | 1,919 | 0 | | | | | | | | | | | | | | | | | | | | |
| 3 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | | | | | | | | | | | | | |
| 4 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | | | | | | | | | | | |
| 5 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | | | | | | | | | |
| 6 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | | | | | | | | |
| 7 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | | | | | | | |
| 8 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | | | | | | |
| 9 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | | | | | |
| 10 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | | | | |
| 11 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | 5,937 | 2,968 | 0 | | | | | | |
| 12 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 0 | | |
| 13 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 8,905 | | |
| 14 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | 11,873 | | |
| Collection year | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 | 5.5 | 6 | 6.5 | 7 | 7.5 | 8 | 8.5 | 9 | 9.5 | 10 | 10.5 | 11 | 11.5 | 12 |

Yearly (rolling) release schedule



ABCD STUDY DESIGN

NESTED STUDY DESIGN IN ABCD

Dependencies between groups of participants violate the independence assumption. Any analysis should therefore account for the known dependencies before estimating unknown variables of interest.



Missing data

- ✓ I don't know
- ✓ I don't want to tell you
- ✓ Truly missing
 - ✓ Messed up, never asked
 - ✓ Lost in transmission
 - ✓ We have answers but no participant ID
- ✓ Missingness by design (not missing)
 - ✓ By event type (e.g. no imaging data at non-imaging events)
 - ✓ New questionnaires/Variables are introduced – missing before date
 - ✓ Missing because of branching logic

DATA SHARING

Shared data, opportunities/challenges

- ABCD Policy: All data is shared on an ongoing basis – no holdout data.
Any results published require a pre-release of that data.
 - Single channel for data release on National Data Archive.
- Share *standard* results such as results from QC pipelines and derived scores is good
 - lower barrier for analysis entry
 - use the community to provide feedback
 - promote best practices
 - reduce researchers degrees of freedom
- Requires additional resources for data curation, additional documentation, data sharing and communication towards the community. Exposes study to more challenging events.

A study centric view of data harmonization

Harmonization of no interest

Name changes require extensive coupling lists for quality assurance

Harmonization of value

Coding of complex data during acquisition to allow for linkage to external information sources

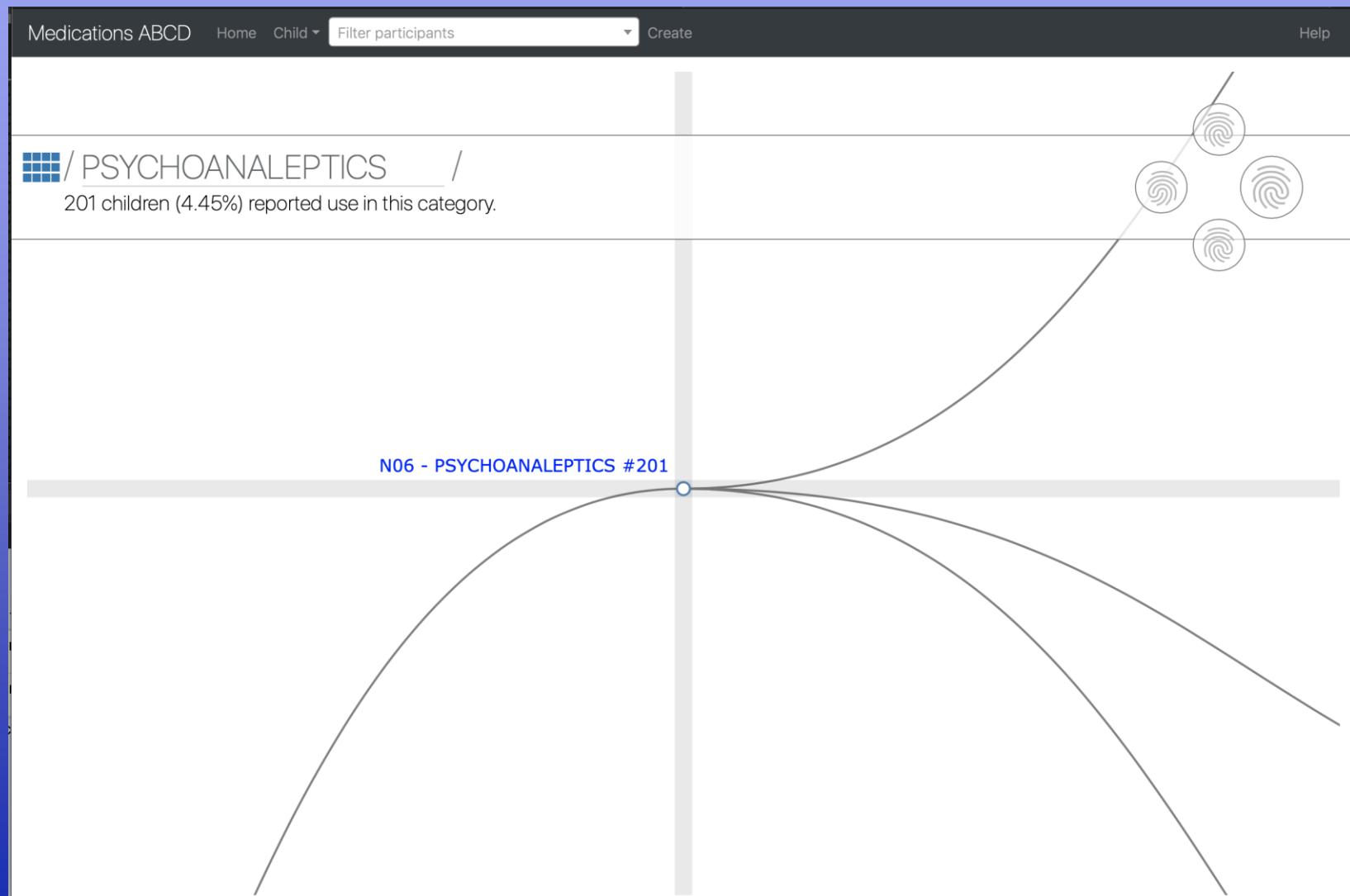
Supported now by NDA:

- Alias fields in data dictionary
- Study specific download packages

Supported by ABCD:

- Use of RxNorm for medication inventory
- Use of consistent names for brain ROIs

DEAP applications for specialized domains



Medications

Anatomical Therapeutic Chemical (ATC1-4)

NERVOUS SYSTEM

Not a drug

RESPIRATORY SYSTEM

PSYCHOANALEPTICS

PSYCHOSTIMULANTS, AGENTS USED FOR ADHD AND NOOTROPICS

Centrally acting sympathomimetics

ANTIHISTAMINES FOR SYSTEMIC USE

CARDIOVASCULAR SYSTEM

Vitamin

The screenshot shows a search interface with a query bar containing "ibuprofen". Below the query, a list of search results is displayed, each with a small thumbnail image and a product name. The results include:

- Ibuprofen 100 MG
- Ibuprofen 200 MG
- Ibuprofen 100 MG/ML
- Ibuprofen 100 MG Chewable Tablet [Advil]
- Ibuprofen Pill
- Ibuprofen 400 MG
- Ibuprofen 200 MG Oral Tablet
- Ibuprofen 50 MG
- Ibuprofen 800 MG
- Ibuprofen Oral Suspension
- Ibuprofen 20 MG/ML Oral Suspension [Motrin]
- Ibuprofen Chewable Product

At the bottom of the interface, there are two small circular icons with numbers: 1162653 Enalapril Pill #1 and 316151 Lisinopril 10 MG #1.

DATA EXPLORATION AND ANALYSIS PORTAL (DEAP)

Data Exploration and Analysis Portal

Web-based interface, cloud deployment

NIMH's NDA data sharing platform as data source

Access to all ABCD measures shared in NDA17

Build-in nesting for multi-level covariates of choice

Access to visualizations and statistical model summary

Shared ABCD data

Available on National Data Archive (nda.nih.gov)
requires signup and support from institution

11,875 participants data available since early 2019
3.2GB spreadsheet data (*.tsv)
23TB MRI (300Gb T1/T2)

65,000 measures per participant
(>67% from imaging)

Resources:

Source code repositories - github.com/ABCD-STUDY/
Data Analysis and Exploration Portal

ABCD open science

[1 Team, 15 members, 33 git repositories]

enroll

Participant enrollment system hosting sensitive data

● PHP Updated 7 hours ago



DEAP

Data Exploration and Analysis Portal of the ABCD Study

● JavaScript ★ 2 ♫ 1 Updated a day ago



analysis-ndar17

Collection of scripts to analyze ABCD release data

ndar17 abcd-study

★ 6 ♫ 5 Updated 2 days ago



redcap_rewrite_history

Change the name of REDCap items in an existing project and attempt to rewrite the projects history.

● PHP ★ 1 GPL-3.0 Updated 5 days ago



FIONASITE

Data upload site for FIONA site computer

● JavaScript ♫ 1 Updated 7 days ago



complete_row

REDCap extension module: Colors each row of an instrument - if a value has been provided. This will highlight rows with missing values to improve their visibility.

● PHP GPL-3.0 Updated on Aug 20



CIFTI-Analysis

Scripts to enable vertex-wise (CIFTI) analysis of ABCD with FSL/PALM and HCP Workbench

● Python ★ 1 Updated on Aug 16



redcap-to-ndar

Exporting REDCap data dictionaries and data to the NIMH National Data Archive (NDA)

● JavaScript ♫ 1 Updated on Jul 31



auto-scoring

Visual programming to calculate derived scores for REDCap

● JavaScript Updated on Jun 29



eprime-data-clean

Convert E-Prime generated files (exported as csv) to proper CSV

● Python Updated on Jun 25



ABCDWorkshop-reproducible-science

Material collection for the ABCD-DAIC workshops on data science

● JavaScript Updated on Jun 16



timeline-followback

Online timeline-followback subject test

● JavaScript Updated on Apr 24



nih-ipad-app-end-point

An end-point for centrally storing data from the NIH iPad app

● PHP ♫ 1 Updated on Apr 20



geocoding

A framework for adding geolocation derived data to the ABCD study.

● R GPL-3.0 MIT Updated on Apr 4



FIONA-QC-PHANTOM

Online QC operations performed on Phantom MRI data

Matlab ★ 1 1 Updated on Mar 23

Fast-Track-Image-Sharing

The ABCD study shares data on the National Data Archive. This project provides the tools for sharing.

ndaa dicom-images anonymization

Python ★ 1 Updated on Feb 7

Minimally-Processed-Image-Sharing

Python ★ 2 Updated on Dec 5, 2017

little-man-task

The little man task web-based instrument

JavaScript Updated on Nov 30, 2017

redcap-completion

Measure item level completion in a large REDCap project

JavaScript Updated on Nov 11, 2017

simple-t1-motion-detection

Measures the amount of ghosting artifacts in T1-weighted images

C++ Updated on Jul 27, 2017

tick-tock

Study Observation system monitoring events per day

JavaScript Updated on Jun 23, 2017

numerical-fitting

Client side numerical computation library written in javascript

JavaScript Updated on Dec 28, 2016

aux-file-upload

PHP Updated on Dec 2, 2016

FIONA-protocol-compliance

Matlab script for ABCD study protocol compliance

Matlab Updated on Nov 28, 2016

redcap-hook-framework

Forked from 123andy/redcap-hook-framework

The REDCap hook framework is a means to organize and deploy custom hooks in a single project or across the entire instance.

PHP ★ 1 16 Updated on Nov 4, 2016

ABCReport

PHP Updated on Sep 6, 2016

pearson-central-end-point

An end-point for centrally storing data from the Pearson's Q-interactive.

PHP Updated on Jun 7, 2016

delay-discounting

Delay-discounting task measuring impulsivity

JavaScript Updated on Aug 11, 2016

ABCD

Adolescent Brain Cognitive Development
Data Exploration and Analysis Portal

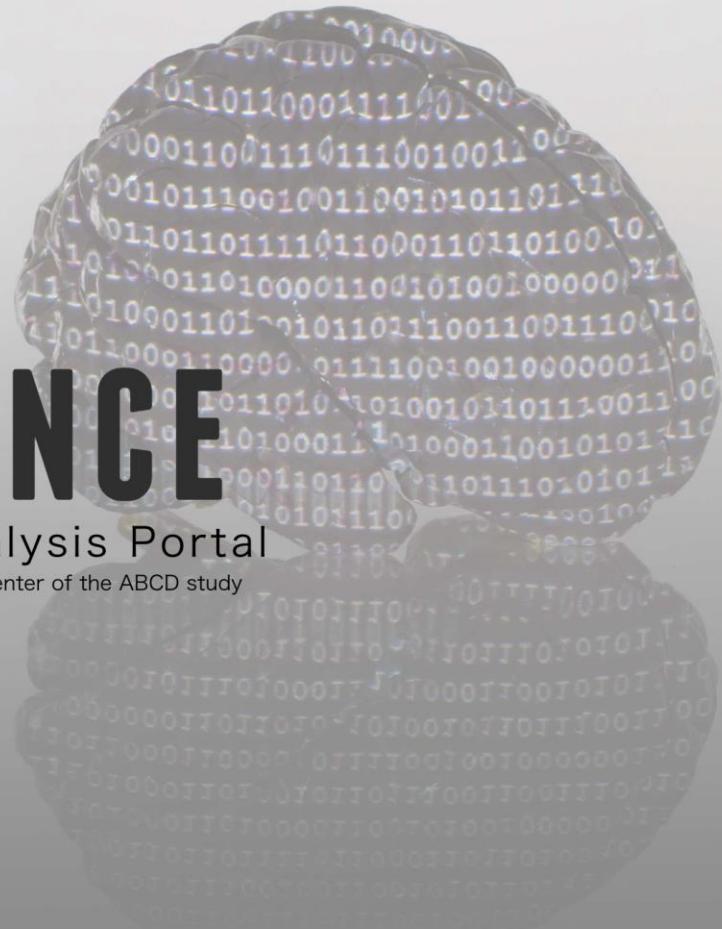
USERNAME: **ADMIN**

GETTING STARTED
00 PLAN
01 EXPLORE
02 LIMIT
03 ANALYSE
04 EXTEND

DEAP SCIENCE

Data Exploration and Analysis Portal

A service provided by the Data Analysis and Informatics Center of the ABCD study

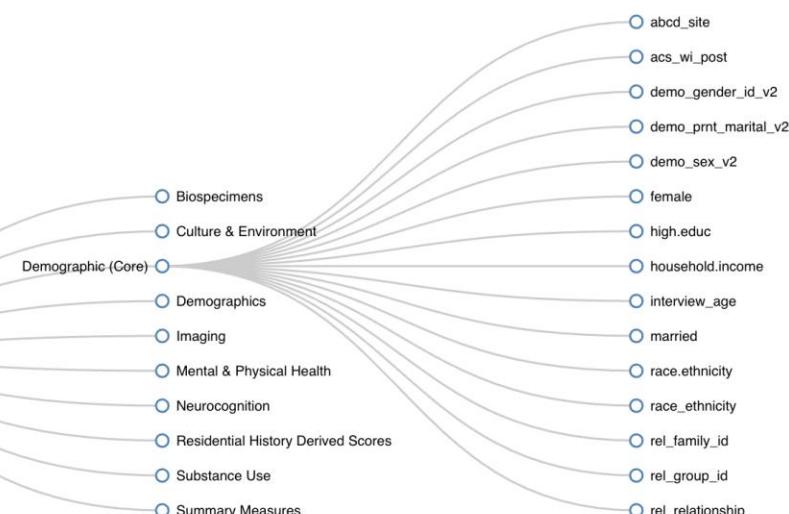


2018 NOV 17

DEAP web-interface

ABCD Ontology

click to expand or collapse, drag to pan, scroll wheel to zoom



interview_age

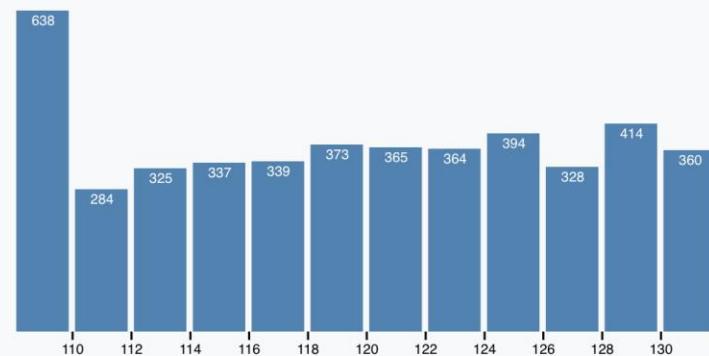
?

examples: intelligence, schizophrenia, ADHD

More than 101 results (0.11 seconds)

interview_age in ABCD Children's Report of Parental Behavioral Inventory / crpb01 [Parenting]

Min. 108
1st Qu. 114
Median 120
Mean 120.02
3rd Qu. 126
Max. 131



[Open NDA for crpb01 (new tab)]

search term: interview_age - matches element name

Age in months at the time of the interview/test/sampling/imaging.

Age is rounded to chronological month. If the research participant is 15-days-old at time of interview, the appropriate value would be 0 months. If the participant is 16-days-old, the value would be 1 month.

interview_age in ABCD Cash Choice Task / cct01 [Task Based]

search term: interview_age - matches element name

Females only

sex="F"

sex M...F

Run

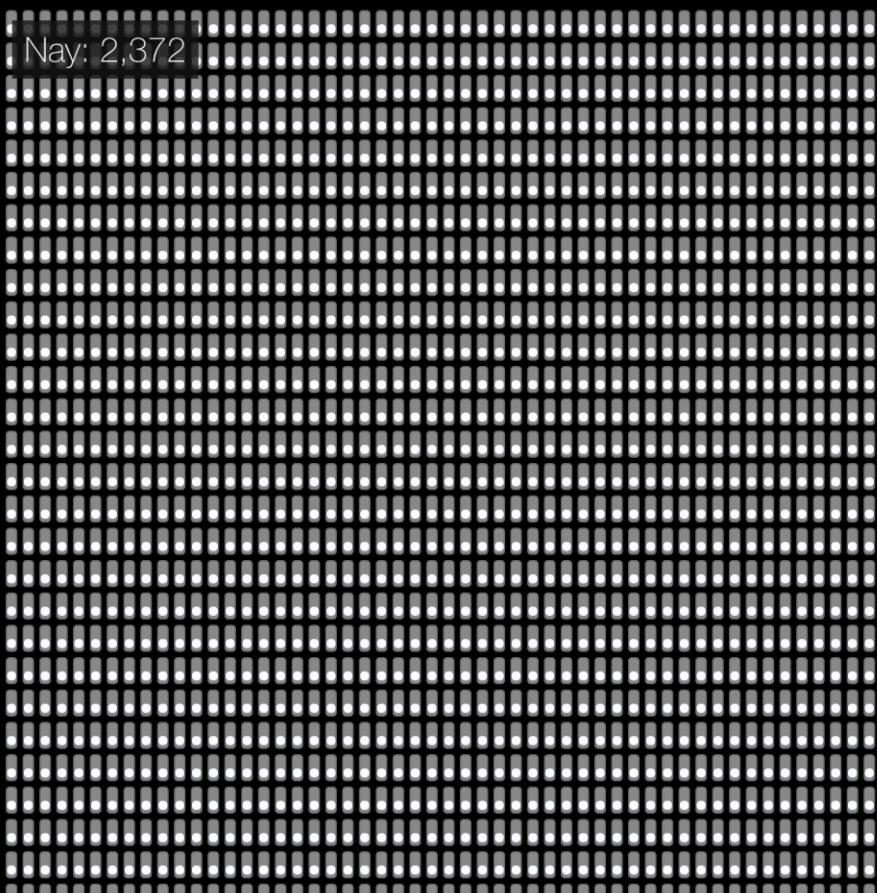
Save

Result of the current restriction

Yea: 2,152
key: #9172



Nay: 2,372



Element Name (user admin - public score)

bmi_calc_example

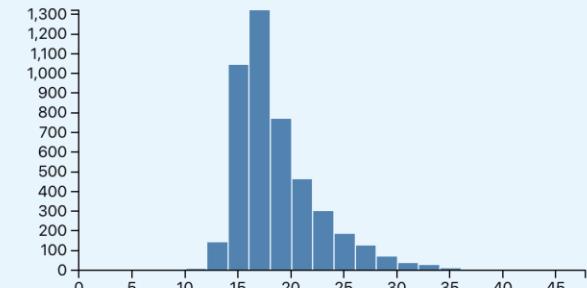
Axis label

The body mass index calculated from the height and weigh

 Save as private

Save

| bmi_calc_example | anthroheightcalc | anthroweightcalc | eventname |
|------------------|------------------|-------------------|---------------|
| 18.234 | 56.5 | 82.8 | baseline_year |
| 20.15 | 56.5 | 91.5 | baseline_year |
| 15.174 | 57.3 | 70.86666666666667 | baseline_year |
| 19.993 | 53.5 | 81.4 | baseline_year |
| 17.663 | 58.3 | 85.4 | baseline_year |
| 16.213 | 54.5 | 68.5 | baseline_year |
| 20.468 | 55.35 | 89.2 | baseline_year |
| 34.171 | 63.5 | 196 | baseline_year |



The Body-Mass Index (BMI)

The body-mass-index (BMI) depends on the height and weight of the participant. These two values exist for each participant in DEAP. We need to copy these values by calling the `use`-function into our browser window. As a return value `use` returns a list of promises that are fulfilled once all the data arrives.

```
var promises = use(["anthroweightcalc", "anthroheightcalc"]);
```

The BMI can be calculated using the following formula - assuming pounds as units for weight (w) and inches as units for height (h):

$$703 \frac{w}{h^2}$$

We can implement this calculation in a function called `calc` that gets two arguments, the weight of a participant in w and the height of a participant in h . The function then returns the calculated values.

```
function calc(w, h) {
  return w/(h*h) * 703;
}
```

Now we wait until the promises have been resolved, which indicates that the weight and height values are available. At this point we can get the data and compute the new variable `anthro_weight_calc` using map. The map function computes for each row of the data spreadsheet the value of the new variable. It is sufficient to `row.set` the new value to have it show up in the histogram and table of this variable:

```
Promise.all(promises).then(function() {
  var data = new DataFrame(allMeasurements);
```

Autosaved: 2:30 pm lines: 27 words: 297

0:0 2 Keystrokes Cmd-P

Multilevel Data Analysis

Multilevel statistical models for baseline data reflect the multilevel study design (GAMM4).

$$Y_{sfi} = \beta_0 + \mathbf{x}_{sfi}\boldsymbol{\beta} + \mathbf{z}_{sfi}\boldsymbol{\gamma} + a_s + b_{f(s)} + \epsilon_{sfi}$$

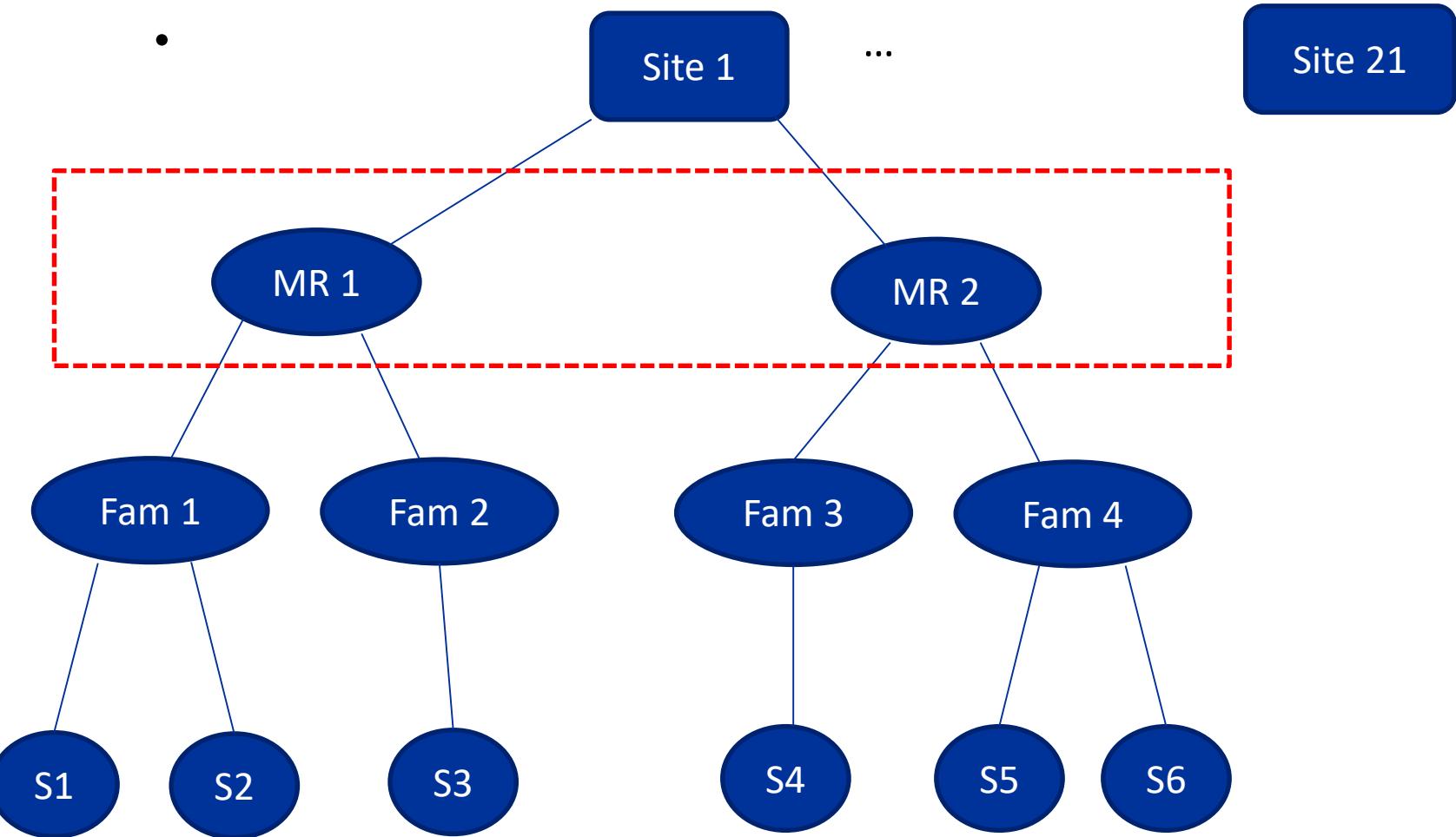
- x_{sfi} are covariates (e.g., demographics)
- z_{sfi} are independent variables of interest
- a_s is a site-specific random effect
- $b_{f(s)}$ is a family random effect nested within site

This model is extendable to non-normal outcomes (e.g., discrete, count variables).

ABCD STUDY DESIGN

- Of these 11,875 subjects, family units include:
 - 8,150 singletons
 - 1,600 non-twin siblings
 - 2,100 twins (1,050 pairs)
 - 30 triplets (10 sets)

ABCD STUDY DESIGN



Tutorial Mode on DEAP

Not familiar with generalized additive mixed models for the analysis of longitudinal data in a multi-site project with a complex family structure? Deap provides a training-wheel mode with in-depth explanations on how to interpret your model.

an independent variable again. Use the buttons to toggle off the inclusion of any of the fixed effect covariates. Both site and family are always included into the model as random effects as they are part of the study design.

Data Display and Summaries

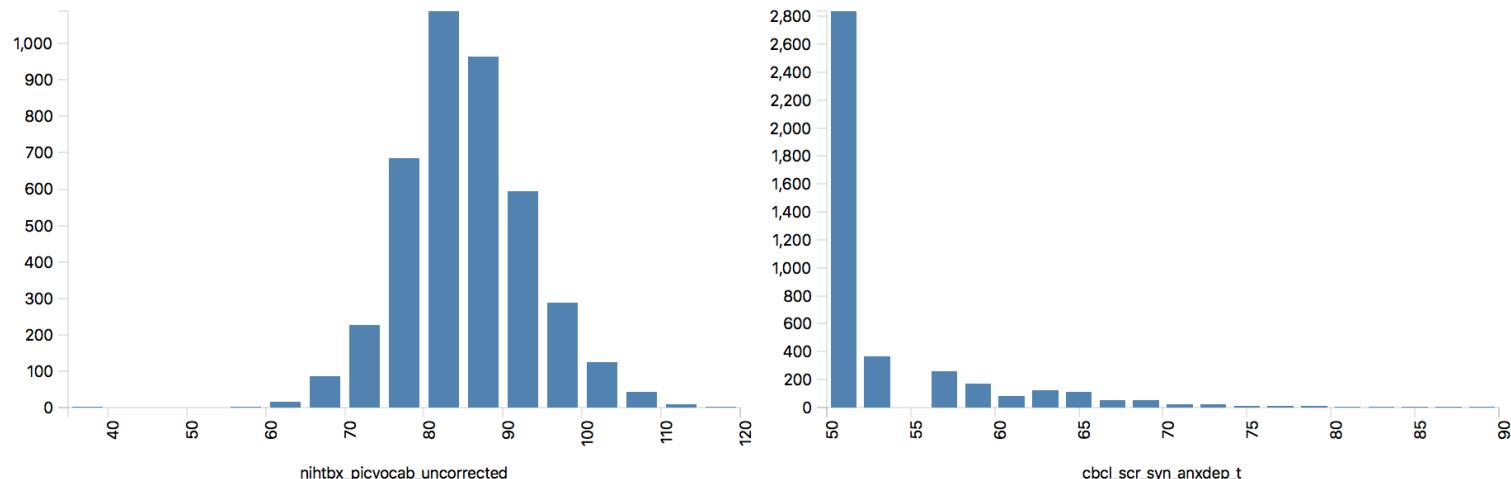


Fig. 2: Data distributions for dependent (left) and independent variable (right).

Histograms are used to inspect the distributions of the data used in the model. For the dependent variable (Fig. 2, left) we want to make sure that they are roughly normally distributed (bell-shaped). In particular we want to check if there are outliers or, if the distribution is highly skewed. If large

Hypothesis Testing on DEAP

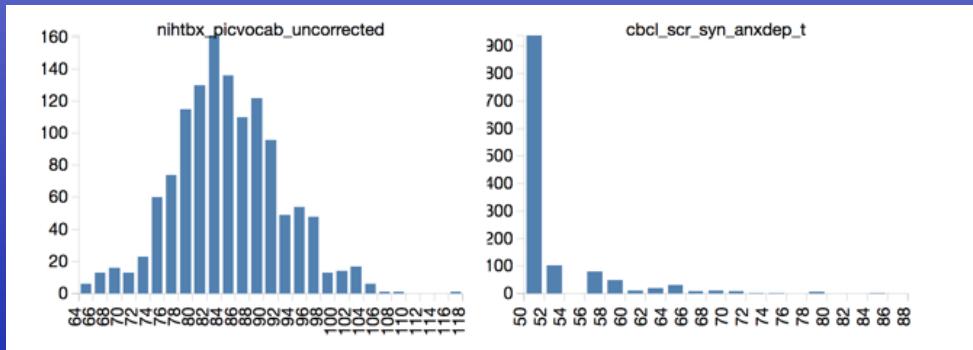
Can changes in anxiety be explained by cognitive development scores measured in the picture vocabulary test, if one corrects for known covariates?

A Model specification

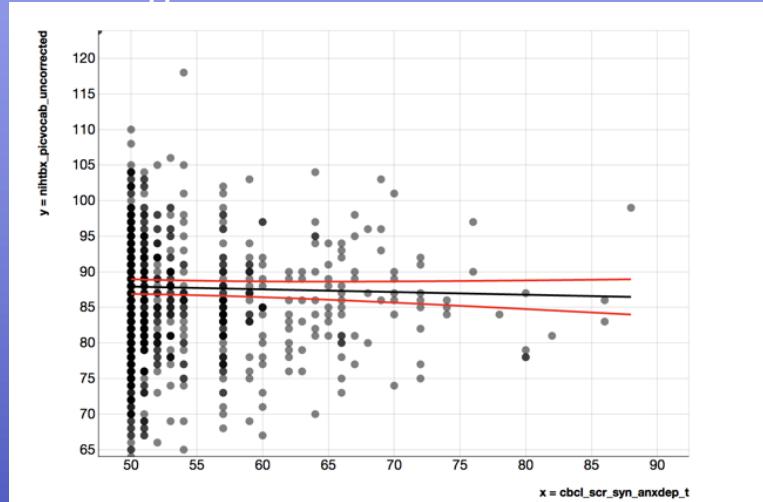
| | |
|-------------------------|---|
| Independent Variable | cbcl_scr_syn_anxdep_t |
| Dependent Variable | nihtbx_picvocab_uncorrected |
| User Covariates | |
| Fixed Effect Covariates | Race/Ethnicity GENDER EDU INC MARITAL AGE |
| Random Effects | SITE FAMILY |

Submit

B Data used in the model



C Regression model fit



D Result tables / Model comparisons

| | Estimate | Std. Error | t value | Pr(> t) | sig |
|-----------------------------------|----------|------------|---------|-------------|-----|
| (Intercept) | 52.27064 | 1.77974 | 29.37 | < 1e-6 | *** |
| nihtbx_picvocab_uncorrected | 0.02316 | 0.01322 | 1.75 | 0.0798201 | . |
| race.ethnicityBlack | -1.15741 | 0.37474 | -3.09 | 0.0020246 | ** |
| race.ethnicityHispanic | -0.14640 | 0.30244 | -0.48 | 0.628372 | |
| race.ethnicityAsian | -1.21511 | 0.66369 | -1.83 | 0.071952 | . |
| race.ethnicityOther | 0.13576 | 0.33444 | 0.41 | 0.6848096 | |
| genderM | 0.67781 | 0.18458 | 3.67 | 0.0002436 | *** |
| high.educBachelor | -0.05391 | 0.54923 | -0.10 | 0.9218111 | |
| high.educHS Diploma/GED | -0.90738 | 0.57636 | -1.57 | 0.154924 | |
| high.educPost Graduate Degree | -0.17039 | 0.56453 | -0.30 | 0.7628061 | |
| high.educSome College | -0.06243 | 0.52201 | -0.12 | 0.9048016 | |
| marriayedyes | -0.40529 | 0.24155 | -1.68 | 0.0926505 | . |
| interview.age | -0.00946 | 0.01301 | -0.73 | 0.4672105 | |
| household.income[< 50K] | 1.12847 | 0.32764 | 3.44 | 0.0005784 | *** |
| household.income[> =50K & < 100K] | 0.48843 | 0.24194 | 2.02 | 0.0435734 | * |

Table 3: Statistical parameter table.

Feature: Expert Mode

Access to the (R) source code behind the GAMM4 model. Can be edited by the user and becomes part of a sharable resource for download and to other DEAP users.



GAMM4 Home Tutorial

This application fits generalized additive mixed models using the R package [GAMM4](#) (Simon Wood, Fabian Scheipl). The GAMM model is appropriate for both cross-sectional and longitudinal regression analyses and allows for an explicit modelling of aspects of the study design such as nesting of subjects within sites data collection and family structures such as twin pairs and other siblings.

Dependent Variable (Y-axis)

nihtbx_fluidcomp_uncorrected +

Submit

Independent Variable (X-axis, for plotting)

dti_fiber_fa_l_ifo +

Grouping Variable (for interaction and plotting)

▼

Other Independent Variables

+

Fixed Effect Covariates

Race/Ethnicity SEX EDU INC MARITAL AGE

Random Effects

SITE FAMILY

Expert Mode (testing and debugging)



```
170 #form_arr = c(independendVar, usercovVar,covfixedVar, smoothVar, logVar, interactionVar, sqVar, sqVar_SQUA
171 ##### similarly for the independent variable...
172 #if independendVar is a smooth variable, log variable, or squared variable remove independendVar
173 if(independendVar %in% c( substring(smoothVar,3,nchar(smoothVar)-1),
174   smoothVarInt.stripped.terml,
175   smoothVarLogVar,5,nchar(logVar)-1),
176   sqVar )) {
177
178   form_arr = c(usercovVar,covfixedVar, smoothVar.all, logVar, interactionVar, sqVar, sqVar_SQUARED);
179   #form_arr = form_arr[form_arr != independendVar]
180 }
181 form_arr = form_arr[form_arr!=""]
182 #take out duplicate variables
183 form_arr = form_arr[!duplicated(form_arr)]
184
185 formulastr = paste(dependendVar, " ~ ",paste(form_arr,collapse='+'))
186 #if(length(smoothVar) > 0){
187 # formulastr = paste(dependendVar, " ~ ",paste(form_arr,collapse='+'),"+",smoothVar)
188 #}
189
190 #get variables involve in the formula
191 varList = all.vars(as.formula(formulastr));
192 varList.independent = varList[-1]
193 print(varList)
194 print(formulastr)
195
196 ##########
197 ## data manipulation ##
198 #####
199 #data = data[data$eventname == "baseline_year_1_arm_1",]
200 print(summary(data[[independendVar]]))
201 #if independent variable has 5 or less unique values change it to character/factor variable
202 categorical.independent = FALSE
203 #if( length(table(data[[independendVar]])) < 6 ){
204 # data[[independendVar]] = as.character(data[[independendVar]])
205 # categorical.independent = TRUE
206 #} else{
207 # data[[independendVar]] = as.numeric(as.character(data[[independendVar]]))
208 #
209
210 if(class(data[[independendVar]]) != "numeric"){
211   categorical.independent = TRUE
212 }
213
214 #user defined covariates
215 #for(ucov in unlist(inputs[['usercov']])){
216 # data[[ucov]] = as.numeric(as.character(data[[ucov]]))
217 #
218
219 print(summary(data[[independendVar]]))
220 #determine if logistic regression or not
221 categorical_dependent = FALSE
222 data[[dependendVar]][data[[dependendVar]] == "" ] = NA
223 if( length(table(data[[dependendVar]]) == 2 ){
224   data[[dependendVar]] = as.factor(data[[dependendVar]])
225   categorical_dependent = TRUE
226 } else{
227   data[[dependendVar]] = as.numeric(as.character(data[[dependendVar]]))
228 }
229 #data[[dependendVar]] = as.numeric(as.character(data[[dependendVar]]))
230
231 #if("demo_prnt_marital_v2" %in% colnames(data)){
```

Fig. 1: Model specification used to define and execute the statistical model.

DEAP Updates

- Docker deployment of DEAP (github.com/ABCD-STUDY/DEAP).
- Pre-registration workflow supporting model specification with variable selection and appropriate variable transformations. Text is provided for sampling, design, and analysis plan as well as for the analysis scripts.
- Subset analysis of participants.
- User defined derived variables with data dictionary entries and scoring algorithms (sharable).
- Upcoming:
- Allow for
 - additional projects shared on DEAP (NDA17, NDA18),
 - additional participants (add to our replace ABCD cohort)

submit

Dependent Variable (Y-axis)

nihtbx_fluidcomp_uncorrected



Independent Variable (X-axis, for plotting)

nihtbx_picvocab_uncorrected



Grouping Variable (for interaction and plotting)



Select subset of sessions



Other Independent Variables



Fixed Effect Covariates

Race/Ethnicity SEX EDU Income Marital AGE

Random Effects

FAMILY SITE

Expert Mode (testing and debugging)



Analyze

This application fits generalized additive mixed models using the R package **GAMM4** (Simon Wood, Fabian Scheipl). The GAMM model is appropriate for both cross-sectional and longitudinal regression analyses and allows for an explicit modelling of aspects of the study design such as nesting of subjects within sites and family, twin pairs and other siblings.

Dependent Variable (Y-axis)

submit

Independent Variable (X-axis, for plotting)

Grouping Variable (for interaction and plotting)

Select subset of sessions

Other Independent Variables

Fixed Effect Covariates

Race/Ethnicity

Random Effects

Expert Mode (testing and debugging)



Fig. 1: Model specification used to define and execute the statistical model.

10.17sec for calculation

The model specified in Fig. 1 is used to estimate the statistical relationship between an independent variable and a measured dependent variable. In the generated model plot (Fig. 3) the independent variable is displayed on the X-axis and the dependent variable appears on the Y-axis. Both measures are user defined and can be selected from a list of available measures. Whereas the independent variable can be of any type (categorical or continuous), and there are no

Save As Clear Delete

Load recipe

GAMM4
(drag & drop entries →)

Data

NDA17

Model

GAMM4

R Code

Imputation

Input

Measure All (single)
Measure All (multi)
Measure Fixed
Measure Categorical (single)
Measure Continuous
Covariates Fixed

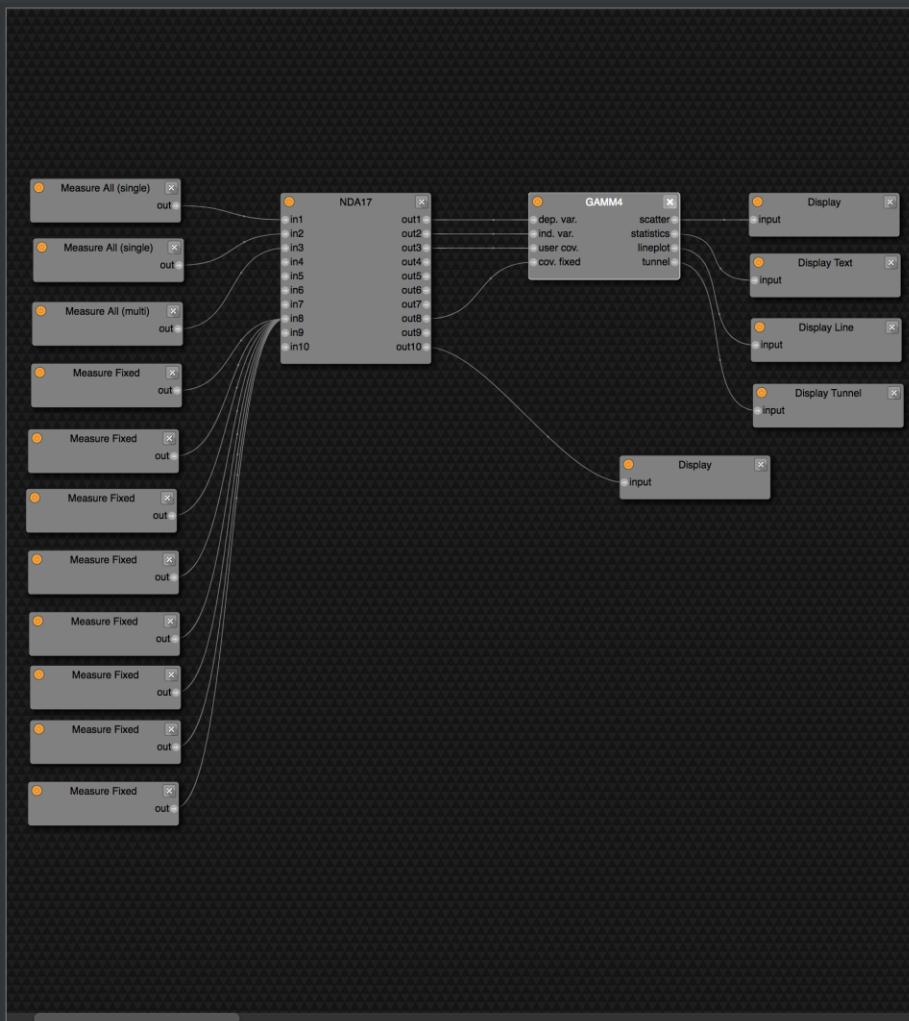
Transforms

Log Transform

Display

Display Scatterplot
Display Text
Display Line
Display Tunnel

model



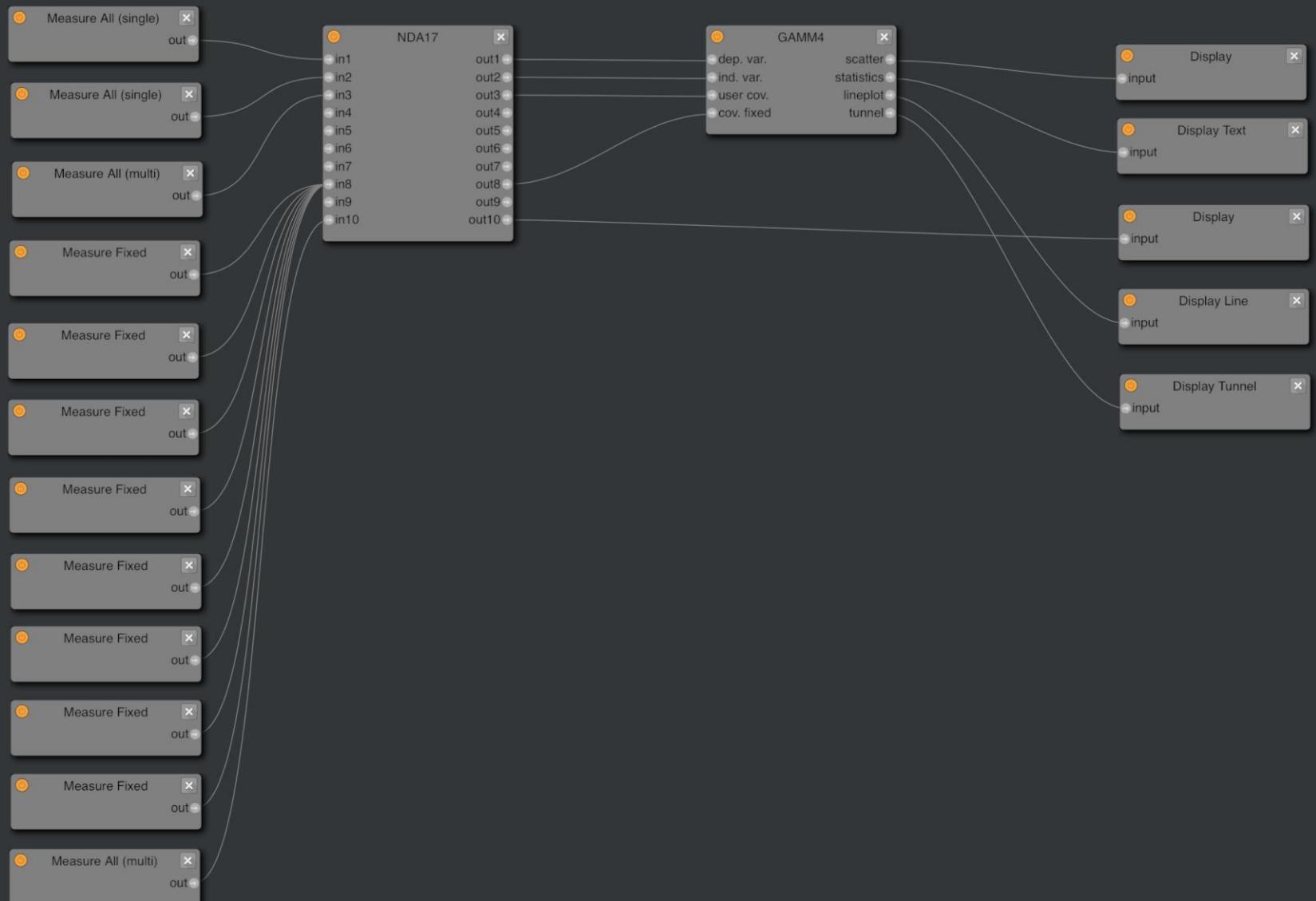
```

14
15 library(gamm4)
16
17 dependendVar = unlist(inputs[['dep.var.']])
18 independendVar = unlist(inputs[['ind.var.']])
19 usercovVar = paste(unlist(inputs[['usercov.']]), sep='+')
20 #nestVar = c("Site", "FamilyID")
21 #usercovVar = usercovVar[!(usercovVar %in% nestVar)]
22
23 #TODO: seperate Site and Family to another catagory of random effects
24 inputs[['cov.fixed']][][[which(unlist(inputs[['cov.fixed.']]) == "Site")]] = "Site"
25 inputs[['cov.fixed']][][[which(unlist(inputs[['cov.fixed.']]) == "FamilyID")]] = "FamilyID"
26
27 covfixedVar = paste(unlist(inputs[['cov.fixed.']]), sep='+')
28
29 form_arr = c(independendVar, usercovVar,covfixedVar);
30 form_arr = form_arr[form_arr!=""]
31
32 formulastr = paste(dependendVar," ~ ",paste(form_arr,collapse='+'))
33
34 ##### data manipulation #####
35 ## data manipulation ##
36 #####
37 data = data[data$VisitID == "baseline_year_1_arm_1",]
38 summary(data[independendVar])
39
40 #if independent variable has 5 or less unique values change it to categorical
41 categorical = FALSE
42 if( length(table(data[[independendVar]])) < 6 ){
43   data[[independendVar]] = as.character(data[[independendVar]])
44   categorical = TRUE
45 } else{
46   data[[independendVar]] = as.numeric(as.character(data[[independendVar]]))
47 }
48
49
50 summary(data[[independendVar]])
51 #age
52 if("Age" %in% colnames(data)){
53   data[["Age"]] = as.numeric(as.character(data[["Age"]]))
54 }
55 #dependent var
56 data[[dependendVar]] = as.numeric(as.character(data[[dependendVar]]))
57
58 #income
59 if("demo_comb_income_v2" %in% colnames(data)){
60   data$demo_comb_income_v2 = as.character(data$demo_comb_income_v2)
}

```

Advanced Usage (Model Builder)

A collaborative environment to integrate advanced statistical analysis features into ABCD. The model builder is software agnostic. R modules coexist next to python/pandas, Matlab. Data frames are used for inter-nodal communication. System provides computational cloud resources and each block can be extracted from the system (data and source-code) for documentation and offline analysis.



Data flow graph (graphical programming) of the Model Builder on DEAP

ACKNOWLEDGEMENTS

- The NIMH Data Archive (NDA)
 - Greg Farber
 - Rebecca Rosen
 - Brian Koser
 - Trevor Griffiths
- NIH/NIDA
 - Gaya Dowling
 - Steve Grant
 - Elizabeth Hoffman
 - Vani Pariyadath
- Anders Dale (PI of the ABCD DAIC: U24 DA041123)
- The DAIC-DEAP Team:
 - Hauke Bartsch
 - Fangzhou Hu
 - Chase Reuter
- ABCD Biostatistics Work Group