

ABCD Human Subject Study

Adolescent Brain Cognitive Development – ABCDSTUDY.org

Release Notes: Adolescent Brain Cognitive Development StudySM (ABCD Study[®]) Data Release 4.0

Linked External Data (LED)

<http://dx.doi.org/10.15154/1523041>

October 2021

Change Log

October 2021 – Data Release 4.0

- Initial release

List of Instruments

Name of Instrument	Short Name
Residential History Derived Scores	abcd_rhds01
	led_school_part_101
	led_school_part_201
ABCD Stanford Education Data Archive (SEDA)	led_school_part_301
	led_school_part_401
	led_school_part_501

General Information

The following information refers to the Adolescent Brain Cognitive Development StudySM (ABCD) Data Release 4.0 available from <https://nda.nih.gov/abcd>. An overview of the ABCD Study[®] is at <https://abcdstudy.org> and detailed descriptions of the assessment protocols can be viewed at <https://abcdstudy.org/scientists/protocols>.

This document describes the contents of various instruments available for download. To understand the context of this information, see *Release Notes ABCD README FIRST* and *Release Notes ABCD Imaging Instruments*.

Details of address data collection as well as included variables included are now available as a preprint of a recent manuscript that is currently under review at <https://psyarxiv.com/59z6b/>

Instrument Descriptions

Residential History Derived Scores

These LED data include information about both the physical and natural environment. The residential history is established according to the parents' report on where they lived. The variables were then retrieved based on the coordinates of their address. The query database was built based on digital archives from the federal government that are described below.

1. Smart location database from EPA.

The description for the smart location database can be found at: <https://www.epa.gov/smartgrowth/smart-location-mapping#walkability>. We used estimates based on 2010 census. The resulting variables include gross residential density and walkability index. The resolution of this database is at the census tract level.

2. Uniform crime report from FBI, compiled by ICPSR.

The description for this database can be found at: <http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/33523>. To maintain a stability on the crime estimates, here we used three-year averages, 2010 to 2012. The resolution of this database is at the county level.

3. Area deprivation index, calculated from American Community Survey.

The area deprivation index (ADI) is calculated based on a published study for the socioeconomic inequality impact on health (<http://annals.org/aim/fullarticle/1983380/neighborhood-socioeconomic-disadvantage-30-day-rehospitalization-retrospective-cohort-study>). The database we used here for query is the 2011 - 2015 American Community Survey 5-year summary. Although the area deprivation index has 18 different sub-scores, the recommendation is to use the national percentiles. The resolution of this is at the census tract level.

4. Population density, UN adjusted, from SEDAC.

The estimation is based on the 2010 census tract while adjusted based on potential under-reporting across the world.

5. Satellite based pollution measures (PM 2.5 and NO2 levels).

Both were obtained from NASA SEDAC. The resolution is at 100 km². Both were 3 year average estimates, spanning from 2010 to 2012.

One-year annual averages of daily PM2.5 estimates from 2016 were also derived using spatio-temporal models at the 1 km² (<https://www.ncbi.nlm.nih.gov/pubmed/27023334>).

6. Elevation.

This is based on direct query to the Google map, which contains elevations given where participants live.

All variables are truncated to avoid potential identification while maintaining overall variations for analytical purposes.

7. Estimated risk of lead (Pb) exposure

All participants' primary, secondary, and tertiary addresses were, if valid, geocoded at the census tract level with respect to an estimated risk of lead (Pb) exposure. These data are from Vox.com (<https://www.vox.com/a/lead-exposure-risk-map>). These risk scores (deciles, 1-10, 10 being the most at risk) were calculated for each census tract and reflect a weighted sum between two summary measures of that census tract: the age of homes and the rate of poverty. The housing age and poverty rate subcomponents of the lead risk score generally reflect an estimated probability of lead exposure given the age of homes (i.e., older homes are more likely to contain lead hazards) and the proportion of individuals living in poverty (-125% of poverty level). The housing-age and poverty-rate scores for valid primary, second, and tertiary addresses are also included.

8. Air pollution exposure

Measures of fine particulate air pollution (PM25), nitrous dioxide (NO2), ozone (O3) were computed for all participants' primary, secondary, and tertiary addresses, when available. Air pollution estimates were using a hybrid spatio-temporal model that provides daily estimates of air pollution levels at 1km x 1km resolution (see <https://www.ncbi.nlm.nih.gov/pubmed/27023334> and <https://pubs.acs.org/doi/abs/10.1021/acs.est.9b03358>). Measures in the current release include mean, minimum, and maximum PM25, NO2, and O3 from 2016, as well as number of

days in 2016 that PM2.5 levels exceeded the National Ambient Air Quality Standards for safe levels of exposure.

9. Traffic count and road proximity

Traffic counts were derived for primary, secondary, and tertiary address at 1km x 1km resolution using data from Kalibrate (https://downloads.esri.com/esri_content_doc/dbl/us/Kalibrate_TrafficMetrixManual_Version140.pdf). These measures indicate the total volume of cars passing through a given area across the entirety of 2016. Road proximity measures capture the number of meters participants' addresses were from major roads.

10. Urban area

Urban area is a dichotomous variable indicating if a participants' addresses were in census tracts considered to be urban (greater than 2,500 people per tract) or rural (less than 2,500 people per tract). This measure was derived from publicly available census data from 2010 (<https://www.census.gov/programs-surveys/geography/about/faq/2010-urban-area-faq.html>).

List of values

- 1: Urbanized Areas (UAs) of 50,000 or more people;
- 2: Urban Clusters (UCs) of at least 2,500 and less than 50,000 people.
- 3: "Rural" encompasses all population, housing, and territory not included within an urban area.

11. Social Vulnerability Index (SVI)

The Social Vulnerability Index (SVI) refers to a 15-variable composite metric that quantifies how vulnerable different communities are to external stresses on human health, such as natural disasters, human-caused disasters, and disease outbreaks, made available from the Centers for Disease Control and Prevention (CDC; <https://www.atsdr.cdc.gov/placeandhealth/svi/index.html>). Higher values reflect more vulnerable communities, here, at the census-tract level. Both the composite and individual metrics (i.e., national percentiles) are available in the ABCD dataset. Data were derived from the 2014-2018 American Community Survey (ACS) 5-year summary (https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/SVI_documentation_2018.html).

12. Opportunity Atlas

The Opportunity Atlas uses anonymous data based on 20million Americans followed from childhood to their mid-30s to provide outcomes for adults who grew up in each census tract..

Percentile household incomes correspond to outcomes in adulthood of people who grew up in each census tract and were born between 1978 and 1983.

The variable corresponding to the 25th percentile is typically the main index of upward mobility for each census tract, as it captures the mean income rank in adulthood for children who grew up in low-income families in this census tract. See the website for more details: <https://www.opportunityatlas.org/>.

13. Child Opportunity Index (COI) 2.0

The COI 2.0 is a composite index derived at the census tract level that measures neighborhood resources and conditions relevant to children's healthy development. In addition to the COI index, there are 3 domain composite scores available: education, health and environment, and social and economic. There are also scores available for the 29 indicators comprising the composite scores. Raw indicator scores and z-scores are provided. The COI 2.0 data is derived from U.S. census tracts for 2010 and 2015.

See the COI 2.0 technical documentation for more details of variable sources and computations (<https://data.diversitydatakids.org/dataset/coi20-child-opportunity-index-2-0-database>)

14. Maximum Temperature and Maximum Vapor Pressure Deficit

Two meteorologic measures for each of seven days preceding and including each participant's baseline interview date was linked to the participants' residential address. The measures include (1) daily maximum modeled temperature ("Tmax") and (2) daily maximum vapor pressure deficit ("VPDmax"). Meteorologic data provided for linkage ranged from January 1, 2016 to December 31, 2018 and is modeled to 800m spatial resolution from the PRISM Climate Group at Oregon State University. Tmax data is in degrees C and the VPDmax data is in hPA. VPDmax data can be used to calculate daily relative humidity and maximum heat index. See, e.g., Spangler K, Weinberger K & Wellenius G. (2019). Suitability of gridded climate datasets for use in environmental epidemiology. *Journal of Exposure Science & Environmental Epidemiology*, 29 (777-89)

It is important to note that the days preceding and the day of were linked based on 'interview_date', which depending on the individual may or may not include the same tasks/scans across all participants. Users should confirm interview_date matches the date of any assessment (i.e. scan session day 1 or 2, or behavioral tests) at an individual level in order to accurately use these data.

It is important to note that not all participants currently have temperature and vapor pressure deficit data as part of 4.0 for baseline visit dates. It is highly recommended that users investigate any possible bias in who does and does not have maximum temperature and maximum vapor

pressure deficit values if using these data. If possible, Release 5.0 may include less missing data for these measures.

15. Cannabis legalization laws

Cannabis legalization categories were assigned to participants addresses based on their state of residence. The four cannabis legalization categories are: 1. Recreational – allows adults to use cannabis for recreational purposes, 2. Medical – allows adults to use cannabis for medical conditions, 3. Low THC/CBD – allows adults to use cannabis that is low in THC and high in CBD for medical conditions, and 4. No legal access to cannabis – forbids access to cannabis.

Information about states current cannabis laws were obtained from two websites:

<http://www.ncsl.org/research/health/state-medical-marijuana-laws.aspx> and

<https://www.mpp.org/states/>.

16. Stigma and Potential Biases

State-level measures of bias, with higher values reflecting more bias, are derived from data about implicit and explicit attitudes about each of these three identity groups aggregated at the state-level. Data sources for these included: Project Implicit (years 2003 to 2018), the General Social Survey (years 1974 to 2014), and the American National Election Survey (1992-2016). Second, for information on gender, they obtained state-level data of women's economic and political statuses (e.g., earning ratios, participation in the labor market and political office, business ownership, etc.) and information about reproductive policies, such as information about availability of abortion providers. Third, for information on attitudes towards Latinx individuals, they examined state-level policies on immigration, recognizing that many Latinx individuals are not immigrants but that such state-level policies likely influence the experience of all individuals in the community with that identity.

ABCD Stanford Education Data Archive (SEDA)

Note that due to NDA file size limitations, the SEDA are divided into five data files that together form the complete data set.

NCES school IDs corresponding to the schools ABCD participants reported attending at the baseline ABCD visit have been linked to the SEDA 4.0 dataset (<https://edopportunity.org/get-the-data/>) for school, district, county, metro, and commuting zone level data. The "CS" SEDA 4.0 files were used as recommended for research purposes in the SEDA documentation and include the Ordinary Least Squares and Empirical Bayes estimation methods. For protection of participant identification and confidentiality, we do not release NCES school IDs or the name or address of the school.

Test scores:

The SEDA 4.0 average tests scores (math + reading/language arts) are included as a broad and cumulative measure of the educational opportunities at each school. For the following spatial resolutions: Geographic School District, County, Commuting Zone, Metropolitan Statistical Area, we linked to two types of “pool” methods:

- “pooled overall” (or pool) files contain estimates that are averaged across grades, years, and subjects
- “pooled by subject” (or poolsub) files contain estimates that are averaged across grades and years within subjects

The linked “pool” variables have the following disaggregation: average test score mean (averaged across grades, years, and subjects), average “learning rate” across grades, average “trend” in the test scores across cohorts, and average difference between math and reading/language Arts (RLA) test scores, along with their standard errors. Estimates are reported for all students and by demographic subgroups.

The linked “poolsub” variables have the following disaggregation: average test score mean in math and in RLA (averaged across grades and years), average “learning rate” across grades in math and in RLA, and average “trend” in the test scores across cohorts in math and in RLA, along with their standard errors.

Please note, data disaggregated by race or ethnicity were intentionally not released and while there are variable names disaggregated by race and ethnicity, they do not contain any data. The following disaggregation categories are included for each geographic unit:

_all all
_fem female
_mal male
_mfg male minus female opportunity gap
_ecd economically disadvantaged
_nec not disadvantaged economically
_neg not disadvantaged economically minus economically disadvantaged gap

Please note that in ABCD we have linked to the cohort standardized (CS) scale scores. Following is the SEDA documentation on how to interpret CS scale scores:

"The CS scale is standardized within subject and grade, relative to the average of the four cohorts in our data who were in 4th grade in 2009, 2011, 2013, and 2015. We use the average of four cohorts as our reference group because they provide a stable baseline for comparison. This metric is interpretable as an effect size, relative to the grade-specific standard deviation of student-level scores in this common, average cohort. For example, a district mean of 0.5 on the

CS scale indicates that the average student scored approximately one half of a standard deviation higher than the average national reference cohort scored in that same grade. Means reported on the CS scale have an overall average near 0 as expected." (Stanford Education Data Archive Technical Documentation. Version 4.0. February 2021, pp 29-30)

Demographic Covariates:

We also provide the SEDA 4.0 dataset demographic covariates. Sources of the SEDA covariate data include:

- U.S. Census Bureau's [American Community Survey \(ACS\)](#) - demographic and socioeconomic characteristics of individuals and households residing in each unit (School district, County, Metro, Commute) (from [National Historical Geographic Information System](#) (NHGIS);
- U.S. Department of Education's NCES [Common Core of Data \(CCD\)](#) - basic descriptive information on schools and school districts, including demographic characteristics; and
- U.S. Department of Education's [Civil Rights Data Collection \(CRDC\)](#) - school demographics.

See the SEDA documentation for more details: <https://edopportunity.org/get-the-data/seda-archive-downloads/#documentation-4>).

Other Variable Notes:

The following categorical variable values were coded to numerical responses as follows.

Variable Key:

Variable	Value
led_sch_seda_s_type	Regular School = 1 Other/Alt School = 2
led_sch_seda_s_level	High = 3, Elementary = 1, Middle = 2, Other = 4
led_sch_seda_s_urbanicity	City = 1, Rural = 3, Suburb = 2
ledsch_seda_d_gslo in	Kindergarten = 0 Pre-Kindergarten = -1

ABCD Study[®], Teen Brains. Today's Science. Brighter Future[®]. and the ABCD Study Logo are registered marks of the U.S. Department of Health & Human Services (HHS). Adolescent Brain Cognitive DevelopmentSM Study is a service mark of the U.S. Department of Health & Human Services (HHS).