



 Non-Imaging Domains > Linked External Data

Linked External Data

DOI: [10.15154/8873-zj65](https://doi.org/10.15154/8873-zj65) (Release 5.0)

PUBLISHED
June 30, 2023

List of Instruments

Name of Instrument	Subdomain	Table Name
<u>School</u>		
School Academic Achievement - Demographics- (SEDA)	School (Demographics)	led_l_seda_demo
School Academic Achievement - Reading (SEDA)	School (Reading)	led_l_sed_read
School Academic Achievement - Math (SEDA)	School (Math)	led_l_seda_math
School Academic Achievement - Math and Reading (SEDA)	School (Math & Reading)	led_l_seda_pool
<u>Environment</u>		
Residential Address Description Variables	Administration	led_l_admin
Parks (NaNDa)	Amenities & Services	led_l_parks
Social Services (NaNDa)	Amenities & Services	led_l_socsrv
Religious/Civic Organizations (NaNDa)	Amenities & Services	led_l_relciv

Name of Instrument	Subdomain	Table Name
Performing Arts and Sports Recreation Organizations (NanDa)	Amenities & Services	led_l_artsports
Neighborhood Socioeconomic Status and Demographics (NaNDa)	Amenities & Services	led_l_nbhsoc
Building Density (EPA)	Built Environment	led_l_densbld
Population Density (EPA)	Built Environment	led_l_denspop
Walkability (EPA)	Built Environment	led_l_walk
Road Proximity (Kalibrate)	Built Environment	led_l_roadprox
Urban/Rural Area (Census)	Built Environment	led_l_urban
Crime (ICPSR)	Built Environment	led_l_crime
Lead (Pb) Risk(Vox)	Built Environment	led_l_leadrisk
Vehicle Density (ACS)	Built Environment	led_l_densveh
Traffic Density (Kalibrate)	Built Environment	led_l_traffic
Census Return	Neighborhood Social Factors	led_l_censusret
Number of Jobs and Job Density (LODES)	Neighborhood Social Factors	led_l_lodes
Opportunity Zones and Investment Scores	Neighborhood Social Factors	led_l_oz
Rent and Mortgage Statistics	Neighborhood Social Factors	led_l_rentmort
Social Mobility (Opportunity Atlas)	Neighborhood Social Factors	led_l_socmob
Alcohol Outlet Density (Census 2016)	Neighborhood Social Factors	led_l_densalc
Satellite-based NO2 Measures (Air Quality Data for Health-Related Applications)	Air Pollution	led_l_no2
Satellite-based PM2.5 Measures (Air Quality Data for Health-Related	Air Pollution	led_l_pm25

Name of Instrument	Subdomain	Table Name
Applications		
Satellite-based O3 Measures (Air Quality Data for Health-Related Applications)	Air Pollution	led_l_o3
Satellite-based Particulate Measures	Air Pollution	led_l_particulat
Satellite-based Pollution Measures for Prenatal Addresses	Air Pollution	led_l_prenatal
Vapor Pressure Deficit (VPD) Estimates (PRISM)	Meteorology and Exposures	led_l_vpd
Temperature Estimates (PRISM)	Meteorology and Exposures	led_l_temp
Elevation of Address (Google API)	Meteorology and Exposures	led_l_elevation
Estimates of Environmental Noise (Harvard)	Meteorology and Exposures	led_l_noise
Selected Environmental Justice Screening and Mapping Tools (EJSCREEN) Measures	Meteorology and Exposures	led_l_ejscreen
Measure of Land Cover and Tree Canopy (NLCD)	Natural Space and Satellite	led_l_nlcd
Land-Use Measures (NLT)	Natural Space and Satellite	led_l_urbsat
Dissimilarity Index (ACS)	Residential Segregation	led_l_dissim
Exposure/Interaction Index	Residential Segregation	led_l_expint
Multi-group Entropy Index	Residential Segregation	led_l_entropy
Index of Concentration at the Extremes	Residential Segregation	led_l_ice
Gi* Statistics (ICPSR)	Residential Segregation	led_l_gi
Area Deprivation Index (ADI)	Neighborhood Composite Measures	led_l_adi
Social Vulnerability Index (SVI)	Neighborhood Composite Measures	led_l_svi
Child Opportunity Index (COI) 2.0	Neighborhood Composite Measures	led_l_coi

Name of Instrument	Subdomain	Table Name
Minority Health Social Vulnerability Index (MHSVI)	Neighborhood Composite Measures	led_l_mhsvi
Behavioral Health Measures (PLACES)	Community Health Burden	led_l_places
Policy-Related Variables		
Gender Bias Measures	Policy Vars	led_l_biasgender
Immigration Bias Measures	Policy Vars	led_l_biasimm
Race Bias Measures	Policy Vars	led_l_biasrace
Cannabis Legalizations Categories by State (NCSL and MPP)	Policy Vars	led_l_lawsmj
Affordable Care Act Medicaid Expansion Data	Policy Vars	led_l_aca
OPTIC-Vetted Co-prescribing Naloxone Policy Data (1990-2020)	Policy Vars	led_l_rxnalox
OPTIC-Vetted Good Samaritan Policy Data	Policy Vars	led_l_goodsam
OPTIC-Vetted Naloxone Policy Data	Policy Vars	led_l_polnalox
OPTIC Vetted Prescription Drug Monitoring Program (PDMP) Policy Data	Policy Vars	led_l_rxmonit
OPTIC Vetted Medical Marijuana Policy Data	Policy Vars	led_l_medmj
CDC Opioid Prescription Dispensing Data per 100,000 Residents	Policy Vars	led_l_rxopiod

General Information

An overview of the ABCD Study® can be found at abcdstudy.org and detailed descriptions of the assessment protocols are available at [ABCD Protocols](#). This page describes the contents of various instruments available for download. To understand the context of this information, refer to the release note [Start Page](#).

Detailed information about the instruments, the constructs they are intended to measure, and relevant citations for each measure are provided in the following:

Fan, C. C., Marshall, A., Smolker, H., Gonzalez, M. R., Tapert, S. F., Barch, D. M., Sowell, E., Dowling, G. J., Cardenas-Iniguez, C., Ross, J., Thompson, W. K., & Herting, M. M. (2021). Adolescent Brain Cognitive Development (ABCD) study Linked External Data (LED): Protocol and practices for geocoding and assignment of environmental data. *Developmental cognitive neuroscience*, 52, 101030. [Find here](#)

The COVID-19 geocoded data can be found [here](#).

Instrument Descriptions

School

Stanford Education Data Archive (SEDA)

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 ([Find here](#)) for school, district, county, metro, and commuting zone level data. For protection of participant identification and confidentiality, we do not release NCES school IDs or the name or address of the school.

In the 5.0 Release, we have subdivided the SEDA tables into logical subdivisions as compared to the 4.0 Release. Please note the table name changes.

Release 5.0 Data Table: led_l_seda_demo

Measure Description: - Measure includes SEDA 4.0 dataset demographic covariates. Sources of the SEDA covariate data include:

- 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);
- S. Department of Education's NCES [Common Core of Data \(CCD\)](#)- basic descriptive information on schools and school districts, including demographic characteristics; and
- S. Department of Education's [Civil Rights Data Collection \(CRDC\)](#) - school demographics.

See the SEDA documentation for more details [Find here](#).

ABCD Subdomain: School

Number of Variables: 112

Notes and special considerations: 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

Release 5.0 Data Table: led_l_seda_read

Measure Description: - Measure includes SEDA 4.0 dataset average test scores (reading/language arts)

The average test scores (reading/language arts) are included as a measure of the reading/language arts educational opportunities at each school at the following spatial resolutions: Geographic School District, County, Commuting Zone, Metropolitan Statistical Area.

ABCD Subdomain: School

Number of Variables: 420

Notes and special considerations: See led_l_seda_pool

Release 5.0 Data Table: led_l_seda_math

Measure Description: - Measure includes SEDA 4.0 dataset average test scores (math)

The average test scores (math) are included as a measure of the math educational opportunities at each school at the following spatial resolutions: Geographic School District, County, Commuting Zone, Metropolitan Statistical Area.

ABCD Subdomain: School

Number of Variables: 420

Notes and special considerations: See led_l_seda_pool

Release 5.0 Data Table: led_l_seda_pool

Measure Description: - Measure includes SEDA 4.0 dataset average test scores (reading/language arts and math)

The average test 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.

ABCD Subdomain: School

Number of Variables: 556

Notes and special considerations: 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. 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. 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; [Find here](#)).

Residential History-Derived Environmental Measures

Amenities and Services

Access to a number of parks, social services and amenities, and additional neighborhood demographics from the The National Neighborhood Data Archive (NaNDA) have been linked to participants' primary, secondary, and tertiary addresses. NaNDA is a publicly available data archive containing contextual measures for locations across the United States ([Find here](#)).

Code and source files used to derive these measures can be found here: Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

The following variables from NaNDA were linked:

Parks (NaNDA)

Release 5.0 Table: led_l_parks

Measure Description: - Measure includes: - Total number of parks (per census tract and per county) - Total number of parks per census tract (top coded at 10) - Proportion of open park land within census tract

ABCD Subdomain: Amenities & Services

Number of Variables: 12

Notes and special considerations: None

[Source](#)

Social Services (NaNDA)

Release 5.0 Data Table: led_l_socsrv

Measure Description: - Measure Includes senior centers, youth centers, food banks, job training programs, and day care centers:

- Total count of social services (per census tract and per county)
- Number of all social services per 1000 people (per census tract and per county)

ABCD Subdomain: Amenities & Services

Number of Variables: 12

Notes and special considerations: None

[Source](#)

Religious/Civic Organizations (NaNDA)

Release 5.0 Data Table: led_l_relciv

Measure Description: - Measure includes: - Total count of religious organizations (per census tract and per county) - Number of all religious organizations per 1000 people (per census tract and per county) - Total count of civic/social organizations(per census tract and per county) - Number of all civic/social organizations per 1000 people (per census tract and per county)

ABCD Subdomain: Amenities & Services

Number of Variables: 24

Notes and special considerations: None

Performing Arts and Sports Recreation Organizations (NaNa)

Release 5.0 Data Table: led_l_artsports

Measure Description: - Measure includes: - Total count of performing arts organizations (per census tract and per county) - Number of all performing arts organizations per 1000 people (per census tract and per county) - Total count of spectator sports organizations (per census tract and per county) - Number of all spectator sports organizations per 1000 people (per census tract and per county)

ABCD Subdomain: Amenities & Services

Number of Variables: 24

Notes and special considerations: None

[Source](#)

Neighborhood Socioeconomic Status and Demographics (NaNa)

Release 5.0 Data Table: led_l_nbhsoc

Measure Description: - The following were derived from the ACS 2013-2017 5-year estimates at the census tract level: - Proportion of people who are foreign born - Proportion of families with Income greater than 75K - Proportion of families with Income greater than \$100K - Proportion 16+ civ labor force unemployed

In addition, three factors associated with neighborhood sociodemographics and structural characteristics are included (based on Morenoff et al., 2007): - Neighborhood disadvantage score is characterized by high levels of poverty, unemployment, female-headed families, households receiving public assistance income, and a high proportion of African Americans in a census tract. - Neighborhood affluence score represents a mix of characteristics associated with neighborhood affluence (concentrations of adults with a college education; with incomes >75K; and employed in managerial and professional occupations). - Neighborhood ethnic/immigrant score represents ethnic and immigrant concentration. Higher values indicate more Hispanic and foreign born in the census tract.

ABCD Subdomain: Amenities & Services

Number of Variables: 24

Notes and special considerations: Population density was set to -1 for tracts with zero population in a given year. Population and area densities are missing in certain years for a few census tracts whose FIPS codes changed at some point between 2003 and 2017.

[Source](#)

Reference:

Morenoff, J.D., House, J.S., Hansen, B.B., Williams, D.R., Kaplan, G.A., & Hunte, H.E. (2007). Understanding social disparities in hypertension prevalence, awareness, treatment, and control: the role of neighborhood context [appendix A: supplemental material]. *Social Science & Medicine*, 65(9), 1853-1866. [Find here](#). Please refer to Appendix A: Supplemental Material.

Built Environment

Walkability (EPA)

Release 5.0 Data Table: led_l_walk

Measure Description: The description for the smart location database can be found [here](#) <https://www.epa.gov/smartgrowth/smart-location-mapping#walkability>. We used estimates based on 2010 census. The resulting variables include walkability index. The resolution of this database is at the census tract level

ABCD Subdomain: Built Environment

Number of Variables: 3

Notes and Special Considerations: None

Building Density (EPA)

Release 5.0 Data Table: led_l_densbld

Measure Description: The description for the smart location database can be found [here](#) <https://www.epa.gov/smartgrowth/smart-location-mapping#walkability>. We used estimates based on 2010 census. The resulting variables include gross residential density (i.e., building density). The resolution of this database is at the census tract level

Release 5.0 Data Table: led_l_densbld

ABCD Subdomain: Built Environment

Number of Variables: 3

Notes and Special Considerations: None

Population Density (EPA)

Release 5.0 Data Table: led_l_denspop

Measure Description: The estimation is based on the 2010 census tract while adjusted based on potential under-reporting across the world (UN adjusted). Data are from NASA Socioeconomic Data and Applications Center ([SEDAC](#))

ABCD Subdomain: Built Environment

Number of Variables: 3

Notes and Special Considerations: None

Traffic Density (Kalibrate)

Release 5.0 Data Table: led_l_traffic, led_l_roadprox

Measure Description: 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.

ABCD Subdomain: Built Environment

Number of Variables: 3

Notes and Special Considerations: None

Road Proximity (Kalibrate)

Release 5.0 Data Table: led_l_roadprox

Measure Description: Road proximity was 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 capture the number of meters participants' addresses were from major roads.

ABCD Subdomain: Built Environment

Number of Variables: 3

Notes and Special Considerations: None

Urban/Rural Area (Census)

Release 5.0 Data Table: led_l_urban

Measure Description: Urban area is a categorical variable indicating if participants' addresses were in census tracts considered to be urban (50,000 or more people), urban clusters (at least 2,500 and less than 50,000 people) or rural (less than 2,500 people per tract, and/or not included in urban areas or clusters). These measures were 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.

ABCD Subdomain: Built Environment

Number of Variables: 3

Notes and Special Considerations: The 4.0 version of the data dictionary contained errors in the labeling used for the urbanicity of participants' census tracts. The correct coding, below, was found in the 4.0 release notes and is now corrected in the 5.0 data dictionary.

Crime (ICPSR)

Release 5.0 Data Table: led_l_crime

Measure Description: The description for this database can be found [here](#). To maintain a stability on the crime estimates, we used three-year averages, 2010 to 2012. The resolution of this database is at the county level.

ABCD Subdomain: Built Environment

Number of Variables: 24

Notes and Special Considerations: None

Lead (Pb) Risk(Vox)

Release 5.0 Data Table: led_l_leadrisk

Measure Description: 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.

ABCD Subdomain: Built Environment

Number of Variables: 3

Notes and Special Considerations: None

Vehicle Density (ACS)

Release 5.0 Data Table: led_l_densveh

Measure Description: Vehicle density was calculated using data from the 2014-2018 American Community Survey 5-year estimates for primary, secondary, and tertiary addresses at the census tract level. Vehicle density was calculated in two ways: 1) As an area estimate (aggregate number of vehicles in a census tract per square mile of land area), and 2) as a population density (aggregate number of vehicles available in a census tract per individual). Vehicle density may be associated with neighborhood residents' levels of exposure to noxious chemicals and their vulnerability to vehicle-related injuries or fatalities.

Code and source files used to derive these measures can be found here: Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Built Environment

Number of Variables: 6

Notes and special considerations: None

Neighborhood Social Factors

Census Return

Release 5.0 Data Table: led_l_censusret

Measure Description: 2010 Decennial Census mail return rates and American Community Survey 2014-2018. 5-year self-response rates are available for participants' primary, secondary, and tertiary addresses at the census tract level. Census response/return rates have been previously associated with social capital and civic engagement, and anomie.

Code and source files used to derive these measures can be found here: Cardenas-Iniguez, C., Schachner, J., Ip, K. I., & Herting, M. (2023). Social and Environmental Context Variables for ABCD

Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Neighborhood Social Factors

Number of Variables: 6

Notes and special considerations: None

Number of Jobs and Job Density (LODES)

Release 5.0 Data Table: led_l_lodes

Measure Description: Number of jobs and job density (number of jobs per square mile of land area) are available at the census tract for participants' primary, secondary and tertiary addresses. Employment statistics are also available, broken down by race and ethnicity. Employment information was derived from the LEHD Origin-Destination Employment Statistics (LODES) dataset for the year 2016, and are available for download [here](#).

Code and source files used to derive these measures can be found here: Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Neighborhood Social Factors

Number of Variables: 54

Notes and special considerations: None

Opportunity Zones and Investment Scores

Release 5.0 Data Table: led_l_oz

Measure Description: Opportunity Zone (OZ) designations, compiled by the Urban Institute, are available for participants' primary, secondary, and tertiary addresses. The Tax Cuts and Jobs Act of 2017 included the designation of Opportunity Zones as a federal incentive to spur investment in low-income and undercapitalized communities. Census tracts were eligible for selection as an OZ if they considered "low-income communities" as determined by the New Market Tax Credit program, although official designations of OZ status were determined by state governors (and mayor of District of Columbia). In addition, Investment Scores, also calculated by the Urban Institute, are available as deciles scores and represent investment flow to census tracts over the years 2011-2015 based on four components: commercial lending, multifamily lending, single-family lending, and small business lending. Additional documentation for these variables can be found [here](#).

Code and source files used to derive these measures can be found here: Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Neighborhood Social Factors

Number of Variables: 6

Notes and special considerations: None

Rent and Mortgage Statistics

Release 5.0 Data Table: led_l_rentmort

Measure Description: In order to approximate cost of living that may be associated with housing, the following variables from the 2014-2018 American Community Survey (ACS) 5-year estimates have been linked to participants' primary, secondary, and tertiary addresses at the census tract level:

Percent home ownership, percent of households living with rent burden (rent is at least 30% of income), percent of households living with severe rent burden (rent is at least 50% of income), median monthly gross rent, and median monthly mortgage payments.

Code and source files used to derive these measures can be found here: Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Neighborhood Social Factors

Number of Variables: 15

Notes and special considerations: None

Social Mobility (Opportunity Atlas)

Release 5.0 Data Table: led_l_socmob

Measure Description: The Opportunity Atlas uses anonymous data based on 20 million 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 social 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/>.

ABCD Subdomain: Neighborhood Social Factors

Number of Variables: 40

Notes and Special Considerations: None

Alcohol Outlet Density (Census 2016)

Release 5.0 Data Table: led_l_densalc

Measure Description: A measure of alcohol outlet density was derived using Census 2016 data for the number of establishments with potential to serve or sell alcohol. This measure was linked to each participant's address at the resolution of ZIP codes.

- Data were obtained from the Census 2016 database containing ZIP Code Business Patterns (ZIP-BP) by Industry and Employment Size of Establishment (e.g., drinking places & full service restaurants)
- ZIP-BP has a High correlation with state/local data sources on alcohol outlet density (Matthews et al., 2011)
- The total number of establishments from the following North American Industry Classification System (NAICS) codes were summed:
 - 445110 (Supermarkets and grocery stores except convenience)
 - 445120 (Convenience stores)
 - 445310 (Beer, wine and liquor stores)
 - 447110 (Gasoline stations with convenience stores)
 - 722511 (Restaurants – full-service)
 - 722410 (Drinking places – alcoholic beverages)

Constructed Variables Contained in Analytic Files

Variable Name	Description
reshist_addr1_alcoh_out_dens	Estimate of alcohol outlet density (Number of Establishments where alcohol could be purchased and/or consumed based on North American Industry Classification System (NAICS) from ZIP Code Business Patterns from Census 2016)

ABCD Subdomain: Neighborhood Social Factors

Number of Variables: 3

Notes and special considerations: None

Reference:

Matthews, S. A., McCarthy, J. D., & Rafail, P. S. (2011). Using ZIP code business patterns data to measure alcohol outlet density. *Addictive behaviors*, 36(7), 777–780. [Find here](#)

Air Pollution

Satellite-based PM2.5 Measures (Air Quality Data for Health-Related Applications)

Measure Description: Measures of fine particulate air pollution (PM25) 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 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.

Release 5.0 Data Table: led_l_pm25

ABCD Subdomain: Air Pollution

Number of Variables: 12

Notes and Special Considerations: In quality control processes the 10×10 km estimates of PM2.5 were found to have errors and not recommended for use (variables names in releases prior to 5.0: reshist_addr1_pm25, reshist_addr2_pm25, reshist_addr3_pm25). These have been removed from the 5.0 release.

Satellite-based NO2 Measures (Air Quality Data for Health-Related Applications)

Measure Description: Measures of nitrous dioxide (NO2) 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 NO2 from 2016.

Release 5.0 Data Table: led_l_no2

ABCD Subdomain: Air Pollution

Number of Variables: 9

Notes and Special Considerations: In quality control processes the 10×10 km estimates of NO2 were found to have errors and not recommended for use (variables names in releases prior to 5.0: reshist_addr1_no2, reshist_addr2_no2, reshist_addr3_no2). These have been removed from the 5.0 release.

Satellite-based O3 Measures (Air Quality Data for Health-Related Applications)

Measure Description: Measures of 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 O3 from 2016.

Release 5.0 Data Table: led_l_o3

ABCD Subdomain: Air Pollution

Number of Variables: 9

Notes and Special Considerations: None

Satellite-based Pollution Measures for Prenatal Addresses

Release 5.0 Data Table: led_l_prenatal

Measure Description: A prenatal period of air pollution exposure was estimated from previous lifetime addresses for the child obtained at the 1-year follow-up, beginning with an approximate date of conception (calculated by subtracting 279 days from the reported birthdate) and continuing through the reported date of birth. Hybrid spatiotemporal models were used to create daily estimates for PM2.5, ozone (O3), and nitrogen dioxide (NO2) for the birth years [2004-2009] at a resolution of 1 km² [1-3]. The daily estimates of these air pollution exposures during the prenatal period of each participant were assigned to the corresponding reported birth year address(es), and daily estimates were averaged across the entire prenatal period to represent the prenatal exposure. For study participants with multiple self-reported addresses for the same calendar year, a weighted average of air pollution exposure was calculated in order to provide a single estimated exposure value for every participant. This was calculated by weighting the prenatal average exposure values for each residence by the reported percent of time spent at that residence; the sum of these weighted exposure averages was then divided by the sum of all reported percentages (regardless of total sum of time between addresses) (Eq 1).

$$\text{Weighted Average} = (\text{Air PollutionAddress}_1 * \text{Percent timeAddress}_1) + (\text{Air PollutionAddress}_2 * \text{Percent timeAddress}_2) / (\text{Percent timeAddress}_1 + \text{Percent time Address}_2)$$

The majority of participants (7,837 [91%]) had reported an address(es) corresponding to the calendar year congruent with the child's birth year as well as the reported percentages of time spent at each address summed to 100%. For a subset of participants (806 [9%]), errors occurred in the reporting of addresses or percentages of time spent at each address, including 386 [4%] participants who reported percentages of time at multiple addresses that did not sum to 100% and

420 [5%] participants who reported addresses with no corresponding percent of time spent living at that address. Given these reporting errors, it is unclear exactly how much time was spent at various addresses during the exposure period, thus introducing uncertainty in exposure classification.

In order to convey this uncertainty, a quality control (qc) variable was created in order to rank prenatal data according to how accurate we expect the air pollution exposure weighted average to be. This qc variable, 'prenatal_exposure_qc' is defined as follows:

prenatal_exposure_qc	Meaning	Number of Participants (Percent)
1	Reported percentages at all addresses sum to 100%	7837 (90.67%)
2	Reported percentages at all addresses sum to totals ranging from 90 to 110%	70 (0.81%)
3	Reported percentages at all addresses sum to totals less than 90% or greater than 110%	316 (3.66%)
4	Reported percentages at addresses were missing	420 (4.86%)

For more information on how residential histories were obtained see As reported in Fan et al. 2021 ([Find here](#)),

ABCD Subdomain: Air Pollution

Number of Variables: 4

Notes and Special Considerations: A number of assumptions and limitations of these data should be noted and highlighted in research and publications that use these derived ABCD data.

- First, ABCD did not specifically ask caregivers where they resided when pregnant or if addresses changed during pregnancy, but rather addresses reported during a given calendar year and coinciding with the participant's birth year were considered representative for the pregnancy period.
- Second, caregivers were only asked to provide move-in and move-out information in terms of calendar year, without month information provided. Thus, chronological sequence of addresses that overlap within a given year can not be deciphered and this error may factor into the weight sum scores computed above.
- The weighted sum equation was applied to all individuals regardless if their provided sum living in a location did or did not add to 100%. The error inherent to this approach can be reduced by only including individuals with a prenatal_exposure_qc variable equal to 1 in analyses using these data.
- The approach used above assigns a prenatal exposure based on daily exposures over the 9-months prior to a child's birthdate. This approach assumes all children were delivered full-term,

which is not necessarily true. It may be best to limit the sample of participants in analyses using these data that can be confirmed to have been in utero for full-term using caregiver report information provided about the child's medical history in other portions of the ABCD dataset.

References:

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M. B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L. J., & Schwartz, J. (2019). An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. *Environment international*, 130, 104909. [Find here](#)

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M. B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L. J., & Schwartz, J. (2020). Assessing NO2 Concentration and Model Uncertainty with High Spatiotemporal Resolution across the Contiguous United States Using Ensemble Model Averaging. *Environmental science & technology*, 54(3), 1372–1384. [Find here](#)

Requia, W. J., Di, Q., Silvern, R., Kelly, J. T., Koutrakis, P., Mickley, L. J., Sulprizio, M. P., Amini, H., Shi, L., & Schwartz, J. (2020). An Ensemble Learning Approach for Estimating High Spatiotemporal Resolution of Ground-Level Ozone in the Contiguous United States. *Environmental science & technology*, 54(18), 11037–11047. [Find here](#)

Satellite-based Particulate Measures

Release 5.0 Data Table: led_l_particulat

Measure Description: Annual estimates for 15 components were linked to baseline addresses for ABCD participants. Spatiotemporal models were used to create annual estimates for baseline years 2015-2018 at a resolution of 50m. Since the models are based on the calendar year, the number of days spent in each calendar year was calculated to determine an annual estimate for the year leading to the baseline date. For each calendar year (prior and visit year), the number of days spent in each year was divided by 365 with adjustments for leap year. These percentages were applied to each year's estimates and summed to represent each participant's estimated annual exposure.

Estimate = (ParticulatePrior Year * Percent timePrior Year) + (ParticulateVisit Year * Percent timeVisit Year)

Estimate = (ParticulatePrior Year * (DaysPrior Year 365 + ParticulateVisit Year) / (DaysVisit Year365)

Further details can be found [here](#) .

ABCD Subdomain: Air Pollution

Number of Variables: 45

Notes and special considerations: None

Reference:

Jin, T., Amini, H., Kosheleva, A., Danesh Yazdi, M., Wei, Y., Castro, E., Di, Q., Shi, L., & Schwartz, J. (2022). Associations between long-term exposures to airborne PM2.5 components and mortality in Massachusetts: mixture analysis exploration. *Environmental health : a global access science source*, 21(1), 96. [Find here](#)

Meteorology and Exposures

Temperature Estimates (PRISM)

Release 5.0 Data Table: led_l_temp

Measure Description: Daily maximum modeled temperature ("Tmax") for each of seven days preceding and including each participant's baseline interview date was linked to the participants' residential address. 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 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)

ABCD Subdomain: Meteorology and Exposures

Number of Variables: 21

Notes and Special Considerations: It is important to note that the days preceding and the day of baseline interview 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 data as part of 5.0 for baseline visit dates. It is highly recommended that users investigate any possible bias in who does and does not have maximum temperature if using these data.

Errors were identified in the 4.0 temperature linkages and should not be used given these issues. These errors were minimized for the provided data in the 5.0 release.

Vapor Pressure Deficit (VPD) Estimates (PRISM)

Release 5.0 Data Table: led_l_vpd

Measure Description: Daily maximum vapor pressure deficit ("VPDmax") for each of seven days preceding and including each participant's baseline interview date was linked to the participants' residential address. 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. 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) **ABCD Subdomain:** Meteorology and Exposures

Number of Variables: 21

Notes and Special Considerations: It is important to note that the days preceding and the day of baseline interview 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 vapor pressure deficit data as part of 5.0 for baseline visit dates. It is highly recommended that users investigate any possible bias in who does and does not have maximum vapor pressure deficit values if using these data.

Errors were identified in the 4.0 VPD linkages and should not be used given these issues. These errors were minimized for the provided data in the 5.0 release.

Elevation of Address (Google API)

Release 5.0 Data Table: led_l_elevation

Measure Description: 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.

ABCD Subdomain: Meteorology and Exposures

Number of Variables: 3

Notes and Special Considerations: None

Estimates of Environmental Noise (Harvard)

Release 5.0 Data Table: led_l_noise

Measure Description: The spatially varying noise data were obtained from a georeferenced noise model of expected environmental sound levels created by Dr. Peter James at Harvard. The model capitalized on acoustical data from 1.5 million hours of long-term measurements from 492 urban and rural sites located across the contiguous United States during 2000-2014. Geospatial sound models were developed based on these measurements to interpret and predict sound levels across the contiguous United States (Mennitt et al. 2014; D Mennitt and K Frstrup 2016). This geospatial sound model formulated relationships between sparsely measured acoustical metrics and nonacoustic environmental factors such as topography, climate, hydrology, and anthropogenic activity. The method utilized random forest, a tree-based machine learning algorithm, to perform the regression.

The resulting non-time-varying geospatial sound model enabled mapping of sound levels at 270m resolution.

All noise metrics are A-weighted. A-weighting is an adjustment that reflects how the human ear perceives sound across the frequency spectrum.

- The L10 is an exceedance metric that corresponds to the sound pressure level exceeded 10% of the time.
- The L50 is an exceedance metric that corresponds to the sound pressure level exceeded 50% of the time.
- The L90 is an exceedance metric that corresponds to the sound pressure level exceeded 90% of the time.
- Leq, or the average sound level in decibels equivalent to the total sound energy measured in a 24-hour period;
- LeqNight, or the average sound level in from the hours of 10p-7a; and
- Ldn, or the day-night average sound level, which is the average sound level over a 24-hour period where sound from 10pm-7am is upweighted by 10 dB.
 - Clarification: The day-night average sound level (Ldn or DNL) is the average noise level over a 24-hour period. The noise level measurements between the hours of 10pm and 7am are artificially increased by 10 [dB](#) before averaging. This noise is weighted to take into account the decrease in community background noise of 10 dB during this period. There is a similar metric called day-evening-night average sound level (DENL) commonly used in Europe or community noise exposure level (CNEL) used in California legislation; that is, the DNL with the addition of an evening period from 7 PM to 10 PM when noise level measurements are boosted 5 dB to account for the approximate decrease in background community noise by 5 dB during this period.
- Variables ending in _ant ant is anthropogenic, and _exi is total noise (anthropogenic and natural) sources. Natural sources were not included given the limited variability seen within largely urban locations in which most participants reside.

ABCD Subdomain: Meteorology and Exposures

Number of Variables: 27

Notes and Special Considerations: A number of assumptions and limitations of these data should be noted and highlighted in research and publications that use these derived ABCD data.

- Sources of the noise are not identifiable based on these models, since the source and sound level is assumed to be homogenous.
- All levels were projected for the summer season.

References:

D. J. Mennitt D.J., K. Fristrup, K., and L. Nelson, L. (2013) Mapping the extent of noise on a national scale using geospatial models. Presented at the 166th meeting of the Acoustical Society of America, San Francisco, USA. [Find here](#)

Mennitt, Daniel J., Frstrup, Kurt M. (2016). Influence factors and spatiotemporal patterns of environmental sound levels in the contiguous United States. *Noise Control Engineering Journal*, 64 (3), 342-353. [Find here](#)

Selected Environmental Justice Screening and Mapping Tools (EJSCREEN) Measures

Release 5.0 Data Table: led_l_ejscreen

Measure Description: Three measures from the EJSCREEN and National Air Toxics Assessment (NATA) are available for participants' primary, secondary, and tertiary addresses at the census tract level. These data were released in 2018, based on 2014 emissions data. Three measures include: 1) Diesel particulate matter level in the air, in g/m³; 2) lifetime cancer risk from inhalation of air toxics; and 3) ratio of exposure concentration to health-based reference concentration. Additional technical information about the EJSCREEN (including information about downloading the data) can be found [here](#).

Code and source files used to derive these measures can be found here:

Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Meteorology and Exposures

Number of Variables: 9

Notes and special considerations: None

Natural Space and Satellite

Measure of Land Cover and Tree Canopy (NLCD)

Release 5.0 Data Table: led_l_nlcd

Measure Description: Measures of land cover (e.g., low-, medium-, or high-density development, forest, wetland) and tree canopy derived from the National Land Cover Database (NLCD) are available for participants' primary, secondary, and tertiary addresses at the census tract and county level of aggregation. Land cover is measured as a percentage of all land within the tract/county. Tracts/counties only include the mainland US (i.e., Alaska and Hawaii are not included).

Land cover and tree canopy measures are derived from the National Land Cover Database [Find here](#). Measures provided here use the 2019 iteration of 2016 data. Shapefiles for census tracts were obtained from the US Census bureau TIGER/Line shapefiles for 2010 [Find here](#).

Code used to derive county and census tract levels are here:

Schertz, K., Kardan, O., Rosenberg, M. D., & Berman, M. (2021). National Land Cover Database at Census Tract and County Level (2016, mainland US). [Find here](#)

Number of Variables: 90

Notes and special considerations: None

ABCD Subdomain: Natural Space and Satellite

Land-Use Measures (NLT)

Release 5.0 Data Table: led__urbsat

ABCD Subdomain: Natural Space and Satellite

Measure Description: Here we include 11 variables comprising an “Urban-Satellite” dataset within ABCD which quantify the percentage or quantity of urbanicity measures such as nighttime lighting, population, vegetation, water, and others, within a 100m radius. These data provide information which can be used to link neuroimaging and public health data to satellite imagery measurements of macro environmental factors.

ABCD Subdomain: Natural Space and Satellite

Number of Variables: 33

Notes and special considerations: None

References: Xu, J., Liu, X., Li, Q., Goldblatt, R., Qin, W., Liu, F., Chu, C., Luo, Q., Ing, A., Guo, L., Liu, N., Liu, H., Huang, C., Cheng, J., Wang, M., Geng, Z., Zhu, W., Zhang, B., Liao, W., Qiu, S., ... IMAGEN Consortia (2022). Global urbanicity is associated with brain and behaviour in young people. *Nature human behaviour*, 6(2), 279–293. [Find here](#)

Residential Segregation

Dissimilarity Index (ACS)

Release 5.0 Data Table: led_l_dissim

Measure Description: The most common measure of **Evenness** is the **Dissimilarity Index**. This index measures the fraction of one group that would have to move to another area, in order to equalize the population distribution. While this measure is computed comparing the ratio of two racial/ethnic groups to each other within a census tract (derived from 2014-2018 ACS estimates), these values are aggregated for a given Metropolitan Service Area (MSA). Thus, we have linked Dissimilarity Indices for the MSA in which a participant may live, and not the census tract. The following contrasts were computed: Black vs. non-Hispanic white, Asian vs. non-Hispanic white, Hispanic vs. non-Hispanic white, nonwhite vs. non-Hispanic white.

Code and source files used to derive these measures can be found here:

Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Residential Segregation

Number of Variables: 12

Notes and special considerations: None

Exposure/Interaction Index

Release 5.0 Data Table: led_l_expint

Measure Description: The most commonly used measures of the **Exposure** dimension are the Indices of Isolation and Interaction. The **Exposure** dimension measures the likelihood of population subgroups interacting with one another using the Index of Interaction. On the other hand, the Index of Isolation measures how likely it is that one group is isolated, or only surrounded by other members of the same group. We have linked a version of the Exposure/Interaction Index that indicates the probability that one member of a group may interact with a member of a contrast group. While the measure is computed at the census tract (derived from 2014-2018 ACS estimates), these values are aggregated for a given Metropolitan Service Area (MSA). Thus, we have linked Exposure/Interaction Indices for the MSA in which a participant may live, and not the census tract. The following contrasts were computed: Black vs. non-Hispanic white, Asian vs. non-Hispanic white, Hispanic vs. non-Hispanic white, nonwhite vs. non-Hispanic white.

Code and source files used to derive these measures can be found here:

Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Residential Segregation

Number of Variables: 12

Notes and special considerations: None

Multi-group Entropy Index

Release 5.0 Data Table: led_l_entropy

Measure Description: Both the Index of Dissimilarity and Interaction Index can only measure the segregation of twogroups compared to each other. The Multi-Group Entropy Index measures the spatial distribution of multiple groups simultaneously (see [here](#)). This measure is dependent on the number of categories included in the computation, and for the linked measures. Multi-Group Entropy Indices at the census tract were derived using the following categories:

Three measures are available as of this release: Multi-Group Entropy Scores at the census tract and metro-level, and a Multi-Group Entropy Index that standardizes the sum of the census-tract scores within a metro by the standard deviation of the metro-level distribution.

Code and source files used to derive these measures can be found here:

Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Residential Segregation

Number of Variables: 9

Notes and special considerations: None

Index of Concentration at the Extremes

Release 5.0 Data Table: led_l_ice

Measure Description: Index of Concentration at the Extremes (ICE) quantifies how persons in a specified area are concentrated into the top vs bottom of a specified societal distribution. We've linked two versions of the ICE: one in which we compare the top income quintile to lowest income quintile (income ICE) as a measure of economic segregation, and one in which we compare the top income quintile for non-Hispanic white to lowest income quintile of non-Hispanic Black individuals in a census tract (income + BW ICE) as a measure of racialized economic segregation.

ABCD Subdomain: Residential Segregation

Code and source files used to derive these measures can be found here:

Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

Number of Variables: 6

Notes and special considerations: None

Reference: Krieger, N., Kim, R., Feldman, J., & Waterman, P. D. (2018). Using the Index of Concentration at the Extremes at multiple geographical levels to monitor health inequities in an era of growing spatial social polarization: Massachusetts, USA (2010-14). *International journal of epidemiology*, 47(3), 788–819. [Find here](#)

Gi* Statistics (ICPSR)

Release 5.0 Data Table: led_l_gi

Measure Description: *From ICPSR Documentation:*

Local Getis-Ord G_i^* statistics were calculated as a measure of residential racial segregation ([Find here](#)). Measures were calculated at the census tract level on the proportion of non-Hispanic white, non-Hispanic Black, non-Hispanic Asian and Pacific Islander for 1990 and 2000 census, and Hispanic persons per census tract. G_i^* statistics are Z-scores that compare the proportion of the population in the focal tract and its neighboring tracts, to the average proportion of a larger geographic unit. For the majority of tracts, the larger geographic unit was the Core-Based Statistical Area (CBSA) these tracts belonged to, and for the minority of tracts that fell outside the boundaries of a CBSA, the County was used as the larger unit.

Data for the measures were obtained from the IPUMS National Historical Geographic Information System (NHGIS) data finder. Data were downloaded for the 1990 and 2000 census, and the 2006-2009, 2010-2014, and 2015-2019 5-year American Community Survey (ACS) estimates. Geographically standardized time series tables were used for 1990 and 2000 census data. All other ACS data were standardized to 2010 census tract boundaries.

G_i^* statistics were calculated using both Rook and Queen conceptualization of spatial relationships. With Rook contiguity, neighbors are determined by those that share a common edge only, while Queen contiguity neighbors are those that share both an edge or a "corner" (common vertex). See detailed documentation for further details.

Code and source files used to derive these measures can be found here:

Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Residential Segregation

Number of Variables: 24

Notes and special considerations: None

Neighborhood Composite Measures

Area Deprivation Index (ADI)

Release 5.0 Data Table: led_l_adi

Measure Description: The Area Deprivation Index (ADI) was calculated based on a published study for the socioeconomic inequality impact on health [Find here](#). The database we queried 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.

ABCD Subdomain: Neighborhood Composite Measures

Number of Variables: 57

Notes and Special Considerations: None

Social Vulnerability Index (SVI)

Release 5.0 Data Table: led_l_svi

Measure Description: 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 and raw percentage scores) 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)

ABCD Subdomain: Neighborhood Composite Measures

Number of Variables: 108

Notes and Special Considerations: For the 5.0 release, SVI measures for ACS 2014-2018 have also been computed based on percentage scores for each census tract so as to increase compatibility with ADI and COI census-based measures.

Child Opportunity Index (COI) 2.0

Release 5.0 Data Table: led_l_coi

Measure Description: 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>)

ABCD Subdomain: Neighborhood Composite Measures

Number of Variables: 262

Notes and Special Considerations: None

Minority Health Social Vulnerability Index (MHSVI)

**** Release 5.0 Table:**** led_l_mhsvi

Measure Description: The Centers for Disease Control and Prevention (CDC) and U.S. Department of Health and Human Services (HHS) Office of Minority Health developed the Minority Health Social Vulnerability Index (MHSVI) to enhance existing resources to support the identification of racial and ethnic minority communities at greatest risk for disproportionate impact and adverse outcomes due to the COVID-19 pandemic. While many of the measures overlap with the SVI Index, additional measures of sociodemographic composition, health care infrastructure & access, and medical vulnerability have been linked to participants' primary, secondary, and tertiary addresses. More information about the MHSVI can be found [here](#).

The following variables from the MHSVI, derived from the 2014-2018 ACS 5-year estimates, were included:

- Percentage population estimates for white, American Indian and Alaska Native, Black or African American, Asian, Hispanic or Latino/a, Native Hawaiian/Pacific Islander (census tract level)
- Percentage population estimates of Spanish/Chinese/Vietnamese/Korean/Russian speakers who "speak English less than 'very well'" (county level)
- Hospitals per 100,000 people (county level)
- Urgent care clinics per 100,000 people (county level)
- Pharmacies per 100,000 (county level)
- Primary care physicians per 100,000 (care, non-federal; county level)
- Percentage population estimates for people without health insurance (census tract level)
- Percentage population estimates for people with no internet access (census tract level)
- Percentage population estimates for people without no computer access (census tract level)

ABCD Subdomain Classification: Neighborhood Composite Measures

Number of Variables: 54

Notes and special considerations: None

Community Health Burden

Behavioral Health Measures (PLACES)

Release 5.0 Data Table: led_l_places

Measure Description: Measures from the PLACES dataset are available for participants' primary, secondary, and tertiary addresses at the census tract level. The PLACES dataset is an expansion of the 500 Cities Project and is available from the Center for Disease and Control and Prevention (CDC), the Robert Wood Johnson Foundation (RWJF) and CDC Foundation. The data linked to ABCD are from the PLACES 2020 release, which reflect measures derived using the 2017/2018 Behavioral Risk Factor Surveillance System (BRFSS) data, and reflect years 2014-2018. Included are 27 measures for the entire United States: 5 chronic disease-related unhealthy behaviors, 13 health

outcomes, and 9 on use of preventive services. Additional technical information about PLACES (including information about downloading the data) can be found [here](#)

Code and source files used to derive these measures can be found here:

Cardenas-Iniguez, C., Schachner, J., Ip, K. I., Abad, S., & Herting, M. (2023). Social and Environmental Context Variables for ABCD Linked External Data Release 5.0. [Find here](#)

ABCD Subdomain: Community Health Burden

Number of Variables: 84

Notes and special considerations: None

Policy-Related Variables

Immigration Bias Measures

Release 5.0 Data Tables: led_l_biasimm

Measure Description: State-level measures of bias, with higher values reflecting more bias, are derived from data about implicit and explicit 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 Subdomain: Policy Vars

Number of Variables: 1 (led_l_biasimm)

Notes and Special Considerations: None

Gender Bias Measures

Release 5.0 Data Tables: led_l_biasgender

Measure Description: State-level measures of bias, with higher values reflecting more bias, are derived from data about implicit and explicit attitudes about gender. 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) to obtain 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.

ABCD Subdomain: Policy Vars

Number of Variables: 2

Notes and Special Considerations: None

Race Bias Measures

Measure Description: State-level measures of bias, with higher values reflecting more bias, are derived from data about implicit and explicit attitudes about race. 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) to compute a Racial Bias Measure.

Release 5.0 Data Tables: led_l_biasrace

ABCD Subdomain: Policy Vars

Number of Variables: 1

Notes and Special Considerations: None

Cannabis Legalizations Categories by State (NCSL and MPP)

Release 5.0 Data Table: led_l_lawsmj

Measure Description: 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
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/>.

ABCD Subdomain: Policy Vars

Number of Variables: 1

Notes and Special Considerations: None

Affordable Care Act Medicaid Expansion Data

Release 5.0 Data Table: led_l_aca

Measure Description: Information on state adoption of statewide Medicaid expansion covering the period 2010-2020 was obtained from two sources:

- KFF (2017) ([Find here](#)). This KFF webpage was updated in May 29, 2020 to reflect state Medicaid expansions through February 19, 2020. [Find here](#)
- Congressional Research Services (2014) The ACA Medicaid Expansion, December 30, 2014. [Find here](#)

From these two sources we constructed the following variable which indicated whether a state passed a law to expand Medicaid statewide (not just for selected subgroups) or if a Section 1115 waiver was obtained to expand Medicaid statewide prior to the 2014 federal allowance under the ACA. These data were mapped into the ABCD data based on (1) the date/month/year (DMY) of the Medicaid expansion policy adoption and (2) the interview date and state of residence of the ABCD resident.

Constructed Variables Contained in Analytic Files

Variable Name	Description
ACAexpand	Residential history derived -dichotomous indicator = 1 if the resident lives in a state at the time of the interview where Medicaid was expanded as a result of the ACA; 0 otherwise

ACA Policy last updated June 29, 2020

ABCD Subdomain: Policy Vars

Number of Variables: 3

Notes and special considerations: None

OPTIC-Vetted Co-prescribing Naloxone Policy Data (1990-2020)

Release 5.0 Data Table: led_l_rxnalox

Measure Description: Co-prescribing naloxone policies captured here represent state policies requiring physicians to co-prescribe naloxone with opioid and/or benzodiazepine prescriptions. Specific dimensions of Co-Prescribing Policy Data included in this public version of the data are based on a review of the legal statutes conducted by OPTIC investigators of laws in effect as of September 2019 and reported in Table 1 of Haffajee et al., 2020, as well as additional information contained in the original legal database.

Constructed Variables Contained in Analytic Files

Variable Name	Description
eff_nalcop	Residential history derived -dichotomous indicator = 1 if any state co-prescribing naloxone policy was effective at date of baseline visit (even if policy restricted to a small subset of patients); 0 otherwise.
eff_al	Residential history derived - dichotomous indicator = 1 if state co-prescribing naloxone policy was effective for all patients at date of baseline visit; 0 otherwise.

Last Updated: September 28, 2020

ABCD Subdomain: Policy Vars

Number of Variables: 6

Notes and special considerations: None

Reference: RAND-USC Schaeffer Opioid Policy Tools and Information Center. OPTIC-Vetted PDMP Policy Data. [Find here](#)

OPTIC-Vetted Good Samaritan Policy Data

Release 5.0 Data Table: led_l_goodsam

Measure Description: Good Samaritan laws are state laws that provide limited civil and/or criminal immunity to individuals who request assistance from authorities during an overdose. Specific dimensions of Good Samaritan policy data included in this public version of the data are based on a review of relevant protections granted through different variations of these laws as described in:

Davis, C. S., & Carr, D. (2015). Legal changes to increase access to naloxone for opioid overdose reversal in the United States. *Drug and alcohol dependence*, 157, 112–120. [Find here](#)

Davis, C., & Carr, D. (2017). State legal innovations to encourage naloxone dispensing. *Journal of the American Pharmacists Association : JAPhA*, 57(2S), S180–S184. [Find here](#)

Information reported as part of OPTIC vetted data that have been linked to the ABCD are:

Constructed Variables Contained in Analytic Files

**** Description****

Variable Name	Description
any_gsl	Residential history derived - dichotomous indicator = 1 if any type of Good Samaritan law was effective at date of baseline interview; 0 otherwise.

Variable Name	Description
gsl_arrest	Residential history derived - dichotomous indicator = 1 if Good Samaritan law that provides protection from arrest for controlled substance possession was effective at date of baseline interview; 0 otherwise.

These GSL variables can be found in the following data files on the RAND-USC Schaeffer OPTIC [\(here\)](#):

- Excel File: WEB_GSL.xlsx
- Stata File: WEB_GSL.dta

Last Updated: July 29, 2020

ABCD Subdomain: Policy Vars

Number of Variables: 6

Notes and special considerations: None

Reference: RAND-USC Schaeffer Opioid Policy Tools and Information Center. OPTIC-Vetted Good Samaritan Policy Data. [Find here](#)

OPTIC-Vetted Naloxone Policy Data

Release 5.0 Data Table: led_l_polnalox

Measure Description: Specific dimensions of naloxone policy data included in this public version of the data are based on a review of relevant protections granted through different variations of these laws as described in:

- Davis, C. S., & Carr, D. (2015). Legal changes to increase access to naloxone for opioid overdose reversal in the United States. *Drug and alcohol dependence*, 157 112-120. [Find here](#)
- Davis, C., & Carr, D. (2017). State legal innovations to encourage naloxone dispensing. *Journal of the American Pharmacists Association : JAPhA*, 57(2S), S180-184. [Find here](#)

Information was obtained from PDAPS (originally downloaded May 27, 2016, and redownloaded March 4, 2020), with a few modifications made in consultation with Corey Davis (who was also consulting with PDAPS to update their laws).

Constructed Variables Contained in Analytic Files

Variable Name	Description
any_nal	Residential history derived dichotomous indicator = 1 if any type of naloxone law was effective at date of baseline visit; 0 otherwise.
nal_protocol_standing	Residential history derived dichotomous indicator = 1 if a naloxone law allowing distribution through a standing or protocol order was effective at date of baseline visit; 0 otherwise.
nal_Rx_prescriptive_auth	Residential history derived dichotomous indicator = 1 if naloxone law allowing pharmacists prescriptive authority was effective at date of baseline visit; 0 otherwise

These naloxone variables can be found in the following data files on the RAND-USC Schaeffer OPTIC website ([Find here](#)):

- Data File for Excel: WEB_NAL.xlsx
- Data File for STATA: WEB_NAL.dta

Last Updated: July 29, 2020

ABCD Subdomain: Policy Vars

Number of Variables: 9

Notes and special considerations:

In Iowa, the legislature adopted two different bills (one House bill and one Senate bill) regarding this section, both with an effective date of May 27, 2016. However, one amended the section and made those amendments retroactive to April 6, 2016. However, given no action could be taken on the retroactive date, we assume the PDAPS effective date of May 27, 2016; thus, it is coded as June 2016.

In Jan 2008, California piloted naloxone programs in several counties (including the most populous LA and SF). However, this was not expanded statewide until January 2014. PDAPS uses the pilot date for the first law (2008), which is what we use here for "any NAL." However, one could make the argument that the 2014 date is preferable.

OPTIC Vetted Prescription Drug Monitoring Program (PDMP) Policy Data

Release 5.0 Data Table: led_l_rxmonit

Measure Description: Information for three variables included in the OPTIC Vetted prescription drug monitoring policy data set have been imported into the ABCD data. These data were based on the following information:

- PDAPS (policy data last reported on June 1, 2017, but last downloaded by us and checked in September 19, 2019). These data were used primarily for the construction of (1) AnyPDMP_partial / date_AnyPDMP for new first-time PDMP laws passed after 1/1/1998; (2) electronic_partial / date_electronic; and (3) Must Access_partial / date_prescriber_mustaccess.
- Use of Brandeis University PDMP Training and Technical Assistance Center (TTAC) from <http://www.pdmpassist.org/content/pdmp-legislation-operational-dates> to obtain dates on first PDMP policy (including paper-based systems) that were adopted prior to January 1, 1998, (for AnyPDMP_partial / date_AnyPDMP).
- Use of Horwitz, J., Davis, C. S., McClelland, L. S., Fordon, R. S., & Meara, E. (2018). The Problem of Data Quality in Analyses of Opioid Regulation: The Case of Prescription Drug Monitoring Programs (No. w24947). National Bureau of Economic Research for two variables: (1) Any_PMP_Horowitz_Partial / date_Any_PMP_Horowitz (source: Table 2, Column 1 of their table) and (2) Op_PDMP_partial / date_Op_PDMP (source: Table 2, Column 4 of their table).
- Consultation with Corey Davis who was also consulting with PDAPS to update their data.

Variable Name	Description
mustaccess	Residential history derived dichotomous indicator = 1 if legislation requiring prescribers to access PDMP before prescribing was effective at date of baseline visit; 0 otherwise.
AnyPDMP	Residential history derived dichotomous indicator = 1 if PDMP enabling legislation for any type of PDMP is in effect (including paper-based systems) at the time of baseline visit; 0 otherwise.
Op_PDMP	Residential history derived - indicates if a "modern PDMP system" became operational and users could access was enabled at date of baseline visit; 0 otherwise.

These PDMP variables can be found in the following data file on the RAND-USC Schaeffer OPTIC webpage ([Find here](#)):

- Excel File: WEB_PDMP.xls
- Stata File: WEB_PDMP.dta

Last Updated: December 10, 2020

ABCD Subdomain: Policy Vars

Number of Variables: 9

Notes and special considerations: Regarding "must access" provisions, various data sources use the term "mandatory use" to mean different things. Mandatory use, which differs from mandatory registration, may simply mean that prescribers must have an account on the system; that they must input their patient's information in the system, or that they must check the system before prescribing an opioid under different circumstances (see [here](#)). These are ultimately different levels of

specificity of the law, and in the spirit of Horowitz et al., 2018, which encourages greater specificity in how these laws are defined, it is appropriate to use the term “physician must access” only when prescribers are mandated to access the database (regardless of the circumstance, which could differ if it is new patient, existing patient, and so on).

OPTIC Vetted Medical Marijuana Policy Data

Release 5.0 Data Table: led_l_medmj

Measure Description: Specific dimensions of marijuana policy data included in this public version of the data are based on variables frequently considered in the marijuana policy research. The primary data come from legal review of laws conducted at RAND with funding from NIDA (PI: Rosalie Pacula) by Anne Boustead (JD, PhD), in consultation with Rosalie Pacula and Rosanna Smart. Updates since December 2017 and March 2019 drew on data reported by NCSL

(<http://www.ncsl.org/research/health/state-medical-marijuana-laws.aspx>) March 2019 with subsequent updates through December 2019 conducted by Jason Blanchard (JD, Boston University) as part of a Boston University-USC-RAND funded project by NIAAA. The early work is described in:

- Powell D, Pacula RL, and M Jacobson (2018). “Do Medical Marijuana Laws Reduce Addictions and Deaths Related to Pain Killers?” Journal of Health Economics. 25: 29-42. PMID: 29408153.
- Pacula RL, Powell D, Heaton P and E Sevigny (2015). “Assessing the Effects of Medical Marijuana Laws on Marijuana: The Devil is in the Details” Journal of Policy Analysis and Management, 34(1): 7-31. PMCID: PMC4315233.

Constructed Variables Contained in Analytic Files

Variable Name	Description
effMML	Residential history derived dichotomous indicator = 1 if the individual lives in a state with a medical marijuana law (MML) (i.e. patients have legal protection to possess MJ containing more than trace amounts of THC for medicinal purposes) that was effective on date of baseline visit; 0 otherwise.
effREC	Residential history derived dichotomous indicator = 1 if the individual lives in a state with a recreational adult-use cannabis law in effect at date of baseline visit; 0 otherwise.
active_medlegdisp	Residential history derived dichotomous indicator = 1 if the individual lives in a state where medical cannabis dispensaries were a legal source of supply as of the date of baseline visit; 0 otherwise.
active_dispREC	Residential history derived dichotomous indicator = 1 if the individual lives in a state where a rec store was opened in the state at date of baseline visit; 0 otherwise.

Variable Name	Description
decrimmmj_ind	Residential history derived dichotomous indicator = 1 if the individual lives in aa state that reduced the status offense for possessing up to one ounce of MJ to a non-criminal offense (civil offense / infraction / petty offense) as of the date of baseline visit; 0 otherwise.

Last Updated: July 7, 2021

Please cite RAND-USC Schaeffer Opioid Policy Tools and Information Center. [YEAR]. OPTIC-Vetted PDMP Policy Data. [Find here](#).

ABCD Subdomain: Policy Vars

Number of Variables: 18

Notes and special considerations: None

CDC Opioid Prescription Dispensing Data per 100,000 Residents

Release 5.0 Data Table: led_l_rxopiod

Measure Description: The CDC makes downloadable data on the rate of opioid prescriptions dispensed in each US state using data from IQVIA Xponent (<https://www.cdc.gov/drugoverdose/rxrate-maps/index.html>). IQVIA Xponent is based on a sample of approximately 49,900 retail (non-hospital) pharmacies, which dispense nearly 92% of all retail prescriptions in the United States. For this database, a prescription is a new or refill prescription dispensed at a retail pharmacy in the sample and paid for by commercial insurance, Medicaid, Medicare, cash or its equivalent, and other third-party coverage. This database does not include mail order prescriptions. For the calculation of dispensing rates, numerators are the projected total number of opioid prescriptions dispensed annually at the county level. Annual resident population denominators are from the U.S. Census Bureau.

Opioid prescriptions, including buprenorphine, codeine, fentanyl, hydrocodone, hydromorphone, methadone, morphine, oxycodone, oxymorphone, propoxyphene, tapentadol, and tramadol are included and identified using National Drug Codes. All prescriptions with days' supply between 1 and 365 days and a known strength description are included. Cough and cold formulations containing opioids and buprenorphine products typically used to treat opioid use disorder are not included. In addition, methadone dispensed through methadone treatment programs is not included in the IQVIA Xponent data.

Variable Names	Description
opioid_Rxper100K_0	Residential history derived measure of opioid prescriptions dispensed per 100K in the county in which the respondent resides in year of baseline visit
opioid_Rxper100K_1	Residential history derived measure of opioid prescription dispensed per 100K in the county in which the respondent resides one year prior to baseline visit
opioid_Rxper100K_2	Residential history derived measure of opioid prescription dispensed per 100K in the county in which the respondent resides two years prior to baseline visit
opioid_Rxper100K_3	Residential history derived measure of opioid prescription dispensed per 100K in the county in which the respondent resides three years prior to baseline visit
opioid_Rxper100K_4	Residential history derived measure of opioid prescription dispensed per 100K in the county in which the respondent resides four years prior to baseline visit
opioid_Rxper100K_5	Residential history derived measure of opioid prescription dispensed per 100K in the county in which the respondent resides five years prior to baseline visit

ABCD Subdomain: Policy Vars

Number of Variables:

Notes and special considerations: In January 2019, IQVIA changed the frame of measurement in their projected prescription services from reflecting prescription demand to be “number of prescriptions dispensed to bin” to reflecting total prescriptions actually “sold to the patient.” To do this, IQVIA eliminated the effects of voided and reversed prescriptions (prescriptions that were never received by the patient) beginning in 2017. This change in measurement frame resulted in a 1.9% downward shift in the measured opioid prescriptions dispensed. This enhancement was applied to data from 2017 and thereafter. Thus, caution should be exercised when examining trends during this time period. Furthermore, starting in 2019, prescriptions were based on the location of the prescriber, rather than the location of the pharmacy.

Residential Address Description Variables

Release 5.0 Data Table: led_l_admin

Measure Description: Included in this table are variables indicating the percentage of time spent at either the primary, secondary, or tertiary residential address, in addition to variables indicating the number of years of residence at a given address,

ABCD Subdomain: Administrative**Number of Variables:** 6

Notes and special considerations: The original address data collection processes in ABCD relied on a point-in-time capture of residential addresses rather than recording longitudinal residential history. As such the assumption has been made that addresses reflected participants' baseline addresses in all versions of the reshist releases to date (2.1, 3.0, 4.0, and 5.0). Users should note that for a given survey timepoint, the percent of time may not sum up to 100%.

To address this limitation, the LED Environment and Policy WG has been actively working to improve the collection of residential history data to improve the temporal and geographic accuracy of participants' reported addresses. These updates aim to improve the collection of retrospective and prospective data in order to provide more comprehensive and accurate residential histories that correspond to both the age of the child as well as each study visit. These data are expected to be available in future releases. Until then, users should be mindful of both the assumption and limitations of the currently available data.

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).