



Child Care Data Center (CCDC) Methodology

Data Sources

The following documentation outlines the various data sources that were included in the Child Care Data Center (CCDC). Following a general description of each data source there is a more detailed description of which data elements were included in each section of CCDC.

State Data Submissions

In February 2020, six pilot states submitted child care data at the provider level and answered a set of survey questions. In July 2020, we requested the same provider-level data that reflected the number of child care providers that were still open. Four states were able to provide this follow-up dataset.

Provider-Level Data Variables

- Provider ID
- Business Name
- Street Address
- City
- State
- Zip Code
- County
- Latitude
- Longitude
- Type of Setting (e.g. center-based, family/home-based)
- License Type (e.g. licensed, license-exempt, unlicensed)
- Total capacity (number of slots)
 - Capacity by age group (e.g. number of slots for infants, toddlers, etc)
- Price by Age Group (infant, toddler, preschool, school-age)
- Languages accommodated by provider
- Hours of operation (opening and closing time)
- Days of operation
- QRIS participation (yes/no)
 - QRIS level (if applicable)
- Accreditation (yes/no)
- Accepts child care subsidy (yes/no)

Survey Questions

- Total number of child care referrals received statewide in 2019.
- Total number of requests statewide for child care during nonstandard hours in 2019.
- QRIS level definitions in the state (number of levels/ratings and what is considered the highest level/ rating)

Third-Party Data

Demographic Data from the U.S. Census Bureau

The following data elements were obtained through the 2018 5-year American Community Survey (ACS) Public Use Microdata Sample (PUMS):

- Population of children under age six
- Children under age six with all parents in the workforce
- Race/ethnicity of householders
- Median income of parents with children under age six
- Housing age

The U.S. Census Bureau and American Community Survey produce the public use microdata set (PUMS). Our partners at NORC pulled the PUMS data and then used the small area microdata set (SAMS) method to properly weight the data based on the a sample of parents and families with children under age 6 and, in some cases, a sample of families with all parents in the workforce with children under 6.

NORC created a weight using proprietary statistical methods -- which is a number. That number is indicative of *how representative* the ACS participant is of the population (in this case all parents of children under 6 or all children under 6 or children under 6 with all parents in the workforce). The weight is then multiplied by the response provided by that participant. Then those estimates are “rolled up” to the geographic level of interest (state, county, census tract). The numbers displayed on the CCDC or used to calculate the numbers on the CCDC are simply *estimates* of the population we are studying in each data story.

The raw data that was used to make these calculations can be found here: <https://www.census.gov/programs-surveys/acs/data/pums.html>

Some data were suppressed for privacy protection in locations with smaller populations.

Toxic Release Inventory Programs

<https://www.epa.gov/toxics-release-inventory-tri-program/tri-data-and-tools>

The Toxics Release Inventory (TRI) Program tracks the industrial management of toxic chemicals that may cause harm to human health and the environment. TRI data are reported by certain industrial and federal facilities.

EPA Superfund Sites

<https://www.epa.gov/superfund/superfund-data-and-reports>

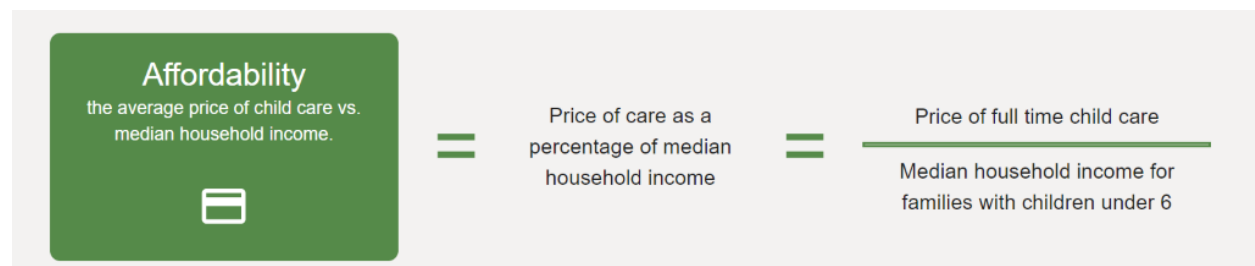
This dataset includes the locations of superfund sites. These are areas contaminated by hazardous waste and can include landfills, manufacturing facilities, mining sites and processing plants.

CDC's State Surveillance Data

<https://www.cdc.gov/nceh/lead/data/state.htm>

This dataset contains records about 418 U.S. counties from selected states measuring lead blood level for children under 72 months of age.

AFFORDABILITY STORY



Full-time child care price data were retrieved from the state submissions through provider-level datasets. These data were reported for different time periods (hourly, daily, weekly, monthly).

We calculate the yearly rate for full-time infant, toddler and preschool providers as follows:

- Monthly price x 12 months
- Weekly price x 52 weeks
- Daily price x 260 days
- Hourly price x 2,080 hours

Note that when more than one time period is listed for a provider, we prioritize longer time over shorter time. For example, if a provider reports a weekly and an hourly rate, we use the weekly rate.

Median income for households with children under age six was obtained from the ACS PUMS dataset. We also used this dataset to calculate median income for married couples with children under age six and single parents with children under age six.

Users can use the tools in this section to estimate how much of a family's budget could be spent on child care at various income levels. For example, they can explore how much a family earning double the median income could pay towards child care.

Nativity of parents data were derived from the SAMS dataset developed by NORC, which included calculations of median income for parents of young children born inside and outside the United States.

Race/Ethnicity of householder data were obtained from the SAMS dataset developed by NORC, which included calculations of the median income for different racial and ethnic groups headed by married couples or by single parents.

Federal poverty level (FPL) is based the following

- An annual income of \$20,578 or less for a family of 3 (2 adults and 1 child)
- An annual income of \$17,622 for a family of 2 (1 adult and 1 child)

Various poverty levels are calculated based on the numbers and family size above. For example, a family of 3 earning \$41,156 would be at 200% FPL (\$20,578 x 2).

ACCESS STORY

Child care slots data were obtained from the provider-level datasets submitted by the pilot states.

- These numbers reflect the number of slots in December 2019, with the exception of Missouri, whose data reflects February 2020.
- The number of slots were also broken down by:
 - Setting type (center-based or family/home-based)
 - Age group. Please note that if a state was not able to provide data at this level, then the message “not reported” will display or the user will be unable to filter down to this level in the county map.
 - QRIS level (quality rating)

Number of children under age six whose parents work was obtained from the ACS PUMS dataset. We calculated the number of child care slots available for children under six whose parents work.

Number of licensed child care providers accepting subsidies was also obtained from the provider-level dataset submitted by the six pilot states.

- This is a Yes/No variable which indicates whether a provider accepts CCDF subsidy vouchers from the state.
- We divided the number who responded ‘yes’ by the total number of licensed providers in the state to get a rate of subsidy acceptance per 100 providers.
- We also broke down the number of providers that accept child care subsidies by setting type and by quality rating.

Number of licensed child care providers who operate during non-standard hours (NSH) was obtained from the provider-level dataset obtained from the six pilot states.

- Non-standard hour care is defined as a child care providers who is open outside of the traditional 6:00am to 6:00pm, Monday to Friday business hours.
- The provider-level dataset included hours and days of operation. Providers open during any time outside of the above criteria are considered NSH providers.
- We divided the number of NSH providers by the total number of licensed providers to get a rate of NSH providers per 100.
- We also present the number of NSH providers by setting type and by quality rating.

Percent of requests for child care during NSH was obtained from the survey responses submitted by the six pilot states.

- We divided the number of requests for NSH care statewide by the total number of referrals statewide in 2019 to get the percentage.

LEAD EXPOSURE STORY

Child Care Slots in High-Risk Locations

To calculate lead risk among child care providers, we calculated the total number of slots for providers who reside in certain areas that are at a higher risk for lead exposure (within 1 mile of a Toxic Release Inventory site and/or in the same zip code as a superfund site).

We also broke down the number of licensed child care slots in these areas by setting type and by child care subsidy acceptance.

Please note that these numbers are simply an estimate of the number of young children who *may* attend child care in areas that are at a higher risk for lead exposure. It does not reflect the actual number of children.

Percent of Old Homes by County

To calculate the percentage of children at risk for lead exposure in their own homes, we used SAMS data to examine the number of children under age six who live in homes built before 1979.

Age of Housing and Blood Lead Level Testing Rates, by County

We used CDC State Surveillance Data to calculate the percentage of children under age six who have been tested for lead levels. These data were compared to the percent of homes in each county that were built before 1979 (indicating an increased risk for lead exposure).

COVID-19 STORY

In July 2020, CCAoA asked the six pilot states to submit the same provider-level variables that they submitted in February 2020. This dataset would reflect the number of child care providers that were open as of mid-July 2020. Our goal was to analyze the loss of providers since COVID-19 began. Four states were able to provide this second dataset: Missouri, Oregon, Washington and Wisconsin.

Individual Neighborhoods Impacted by COVID-19

The first map breaks down the loss of child care providers since COVID-19 began. We compared the number of providers in each tract from the first dataset to the number in the second dataset to calculate the percent lost. We made the same calculations with the number of slots (capacity).

- Users can also break down the losses by:
 - provider type (center-based and family/home-based)
 - program quality rating (QRIS participation and rating)
 - subsidy acceptance (providers who accept child care subsidies)

Furthermore, we compared the percentages of various races and ethnicities within each census tract to determine if communities of color were more likely to be impacted by losses in child care supply.

- We used SAMS data at the census tract level to input the percentage of the following racial/ethnic groups in each tract:
 - White
 - Black

- Hispanic
- Asian
- Alaska Native/Native American (AINA)
- Native Hawaiian/Pacific Islander (NAPI)
- Multiracial
- Other

How COVID-19 Has Affected the Overall Supply of Child Care

Using the February 2020 and July 2020 datasets, we compared the change in child care supply (providers and slots) using the following categories:

- setting type (center-based, family/home-based)
- providers accepting child care subsidies (number of slots not available)
- quality rating (if applicable)