COI 2.0 ZIP CODE DATA

Technical documentation

01|08|22

Clemens Noelke, Robert W. Ressler, Nancy McArdle, Erin Hardy, and Dolores Acevedo-Garcia

diversitydatakids.org
data for a diverse and equitable future

Brandeis | THE HELLER SCHOOL FOR SOCIAL POLICY AND MANAGEMENT

DATA USE AGREEMENT

PREFERRED CITATION

Noelke, C., Ressler, R., McArdle, N., Hardy E., & Acevedo-Garcia D. (2021). Child Opportunity Index 2.0 ZIP Code Data Technical Documentation. Retrieved from https://www.diversitydatakids.org/research-library/research-brief/using-zip-code-level-data-coi.

COI 2.0 ZIP CODE DATA

The Child Opportunity Index (COI) 2.0 is a composite index of neighborhood opportunity capturing 29 neighborhood features relevant to children's healthy development measured at the census tract level.[1,2] To meet growing demand for ZIP code level COI 2.0 data, diversitydatakids.org has calculated and published ZIP code level estimates of COI 2.0 index and child population data. There are about 73 thousand census tracts in the US that intersect with about 39 thousand ZIP codes. ZIP codes, unlike census tracts, do not have precisely defined geographic boundaries. They are collections of addresses defined by the U.S. Postal Service (USPS) for efficient mail delivery. Our method for aggregating census tract to ZIP code data uses weights published by the U.S. Department for Housing and Urban Development. A ZIP code COI 2.0 estimate is then calculated as a weighted average of census tract data, where census tracts that are home to a greater proportion of residential addresses with a given ZIP code receive greater weight. The weights used to calculate ZIP code level COI 2.0 data change continually because ZIP code definitions are continually updated by the USPS (e.g., for example, to incorporate newly constructed buildings/addresses). As a result, the weights used to calculate ZIP code level COI 2.0 data also change continually. We, therefore, publish multiple ZIP code data files based on 2015 COI 2.0 census tract data that correspond to the ZIP code definitions in a given year. It is important to note that for all ZIP code COI 2.0 data files, neighborhood conditions were measured in and around 2015. The years in the data file titles DO NOT denote the years in which neighborhood conditions were measured but refer to the ZIP code definitions for a given year. COI 2.0 ZIP code data include Child Opportunity Levels and Scores as well as child population counts by race/ethnicity.

Users should be aware of the following issues and limitations.

- ZIP code definitions are regularly updated by the US Postal Service. We therefore created multiple single-year and multi-year files that exactly match ZIP code definitions for a given period. For example, users interested in studying hospital admissions in 2018 would select the file "COI 2.0 ZIP code data for 2018 ZIP codes", while users studying hospital admissions in 2020 would select the file "COI 2.0 ZIP code data for 2020 ZIP codes". We have published single-year files from 2015 to 2020, corresponding to ZIP code definitions of the respective year. A full list of available datasets, including multi-year files, and information on how to obtain them is available on page 12 ("List of Files").
- COI 2.0 ZIP Code data from different data files should not be compared. If values for a given ZIP code differ across ZIP code data files, this should **not** be interpreted as a change in opportunity. Neighborhood conditions are measured in or around 2015 for all the published data files. Any differences across annual files only reflect changes in the weights used to calculate ZIP code estimates from census tract data.
- Only nationally-normed ZIP code Child Opportunity Scores are available (that is, scores for which a ZIP code is ranked relative to all other ZIP codes in the U.S.) Metro and state-normed Child Opportunity Scores cannot be reliably calculated for all metro areas and states.
- 66% of ZIP code estimates are based on data from two or more census tracts. If opportunity across tracts differs substantially, the resulting ZIP code estimate will not reflect neighborhood opportunity of the constituent census tracts. To capture this variation, we have included ZIP code standard deviations for all ZIP code level COI 2.0 z-scores (overall, education, health & environment, social & economic). The standard deviations capture the variability of COI 2.0 z-scores across census tracts contributing to a given ZIP code estimate. Standard deviations are larger if the tracts underlying the ZIP code estimate have more dissimilar z-scores.
- Little is known about the reliability of ZIP code level vs. census tract level measures of residential context. We are currently investigating this issue[3] and will publish updates on our research when available. Our preliminary results

indicate that – whenever census tract data is not available – ZIP code COI data represent a viable alternative to study the association between residential environments and children's outcomes in a large national dataset.

The following sections outline our method to aggregate census tract data to ZIP codes and illustrate the limitations above. We first describe our aggregation method conceptually and discuss how we resolve some challenges involved in this process. The section "Technical Notes" describes in more detail the source datasets and formulas used to calculate ZIP code data. This section is followed by the data dictionary which lists variable names and definitions for the published data files.

AGGREGATING COI DATA FROM CENSUS TRACTS TO ZIP CODES

ZIP codes are collections of mail delivery routes drawn by the U.S. Postal Service (USPS). They are defined to facilitate efficient mail delivery to residential, business, and other addresses. Unlike census tracts that divide the country into distinct spatial areas, ZIP codes do not have exactly defined geographic boundaries. They are collections of mail delivery routes and not areal geographic units or geographic units covering a certain area. While there are about 39,000 (non-P.O. Box) ZIP codes, there are more than 73,000 census tracts in the 50 states plus D.C. Especially in more densely populated, urban areas, census tracts provide more granular information on children's residential environments. Moreover, they are precisely defined areal units, which facilitates spatial analysis and visualization. COI 2.0 data is available for about 72,200 census tracts.

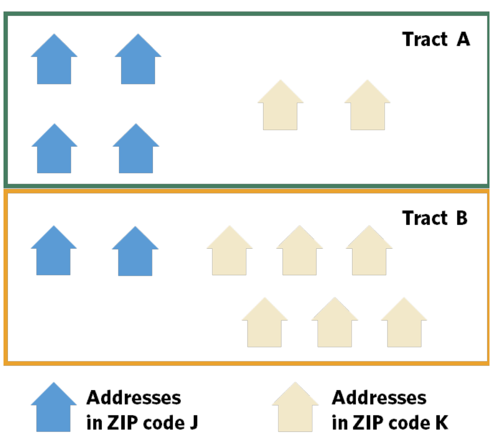
There are two main challenges when aggregating census tract to ZIP code data. First, a census tract may contain addresses belonging to multiple ZIP codes; and, addresses with the same ZIP code may be found in multiple census tracts: 66% of ZIP codes fall into two or more tracts, 25% fall into six or more tracts, 10% fall into 11 or more tracts, and 1% of ZIP codes fall into 20 or more tracts.[1] Second, ZIP codes are continuously updated: ZIP codes are sometimes retired or newly created, but most changes alter the delineation of ZIP codes, e.g., by adding or removing new addresses.

In the following section, we outline these issues and our aggregation method that addresses both of these challenges. The following sections also describe conceptually how we combine census tract COI data to obtain ZIP code estimates and how we deal with changing ZIP code delineations over time. We then discuss biases due to aggregation.

Combining census tract data to obtain ZIP code estimates using address weights

A census tract may contain addresses belonging to multiple ZIP codes, and the same ZIP code may be found within the area of multiple census tracts. Figure 1 illustrates this issue. It shows 6 residential addresses with ZIP code J and 8 residential addresses with ZIP code K. Census tract A contains 4 addresses belonging to J and 2 address belonging to K. Tract B contains 2 addresses belonging to J and 6 addresses belonging to K. ZIP code J has a total of 6 addresses spread over two tracts, and ZIP code K has a total of 8 addresses spread over two tracts. How do we aggregate or allocate data available for Tract A and B to ZIP codes J and K?

Figure 1. Two census tracts that contain addresses with two different ZIP codes.



Central to our approach are crosswalks provided by the U.S. Department of Housing and Urban Development (HUD) linking census tracts to ZIP codes. Each crosswalk contains one row per ZIP code-tract segment, or per combination of ZIP codes and the census tracts they fall into. For each segment, different crosswalk files provide the proportion of addresses within a ZIP code that fall into a given census tract (ZIP-Tract weights) and the proportion of addresses within a census tract that have a given zip code and (Tract-ZIP weights). We combine these crosswalks with census tract level COI and child population data, and then use the address weights to allocate census tract data to ZIP codes.

[1] Descriptive statistics based on the datafile used to calculate 2020 COI 2.0 ZIP code estimates, 2020 fourth quarter, which includes 39,061 ZIP codes linked to 72,190 census tracts with non-missing COI 2.0 data.

Table 1 illustrates these steps in greater detail using artificial data for COI 2.0 overall z-scores and child population counts. To aggregate census tract COI 2.0 z-scores, we take a weighted average of census tract z-scores, where the weights capture

TABLE 1. CALCULATING ZIP CODE COI 2.0 Z-SCORE AND POPULATION ESTIMATES AS WEIGHTED AVERAGES OF CENSUS TRACT LEVEL DATA.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ZIP Code	**Tract**	**Number of addresses**	**Weight (ZIP-Tract)**	**COI 2.0 z-scores**	**(4) x (5)**	**Weight (Tract-ZIP)**	**Population**	**(7) x (8)**
J	A	4	0.67	0.5	0.33	0.67	6	4
J	B	2	0.33	0.2	0.07	0.25	8	2
K	A	2	0.25	0.5	0.13	0.33	6	2
K	B	6	0.75	0.2	0.15	0.75	8	6
			ZIP Code J z-score		**0.40**	**ZIP Code J population**		**6**
			ZIP Code K z-score		**0.28**	**ZIP Code K population**		**8**

the proportion of a given ZIP code's addresses that fall into a given tract. For illustration, column 3 contains the number of residential addresses for each ZIP code and census tract segment. Column 4 contains a weight defined as the proportion of ZIP code addresses that fall into a given ZIP code-tract segment. These weights sum to one across ZIP code-tract segments within the same ZIP code. Column 5 contains COI 2.0 z-scores, which are identical for the segments belonging to the same census tract. To aggregate the ZIP code-tract segment data to the ZIP code level, we multiply the weight by the z-scores and then sum over the segments belonging to a given ZIP code. For example, the estimated ZIP code z-score for ZIP code J is equal to 0.33 + 0.07 = 0.40, which is closer to tract A's z-score (0.5) than to tract B's z-score (0.2), because tract A contains two thirds of the addresses in ZIP Code J, while tract B contains only one third of the addresses in ZIP code J.

To allocate census tract population counts to ZIP codes, we use a different set of weights. The weights ensure that the total number of children will be the same across ZIP codes and census tracts. To illustrate, column 7 contains the proportion of addresses in a census tract that fall into a ZIP code-tract segment. For example, one quarter of addresses in tract B have ZIP Code J addresses, while three quarters of addresses have ZIP code K. These weights sum to one across segments for a given census tract. Column 8 contains population counts, which are identical for the segments belonging to the same census tract. To aggregate tract data to the ZIP code level, we multiply the weight with the population data (column 9) and then sum over the segments belonging to a given ZIP code. For example, the estimated ZIP code population for ZIP code J is equal to 4+2 = 6. While the sum is a round number here, this is not generally the case.

Dealing with creation, retirement and change in ZIP codes

ZIP codes are sometimes retired or newly created. ZIP codes also change gradually, for example, due to new housing unit construction and the creation of new addresses. Because of the addition or removal of addresses, the weights used to allocate/aggregate census tract data to ZIP codes change. HUD therefore releases new crosswalks at the end of each quarter to capture these changes. We use all available crosswalks to generate annual and multi-year COI 2.0 ZIP code estimates that account for temporal change in ZIP codes. The ZIP code estimates for any given year or period therefore include all ZIP codes that existed in that period, i.e., it excludes ZIP codes that were retired in preceding years or newly created in subsequent years. While the creation and retirement of ZIP codes is easily dealt with by selecting only tracts that existed within a given time span, the changing address distribution across zip codes requires an adjustment to the calculations described in the preceding section.

Table 2 illustrates the problem of changing ZIP codes due to new construction and how we resolve it using hypothetical data. ZIP code L is observed in tract C and D. L, C, and D are observed in two consecutive quarters, Q1 and Q2. In the second

TABLE 2. CALCULATING ZIP CODE COI 2.0 Z-SCORE ACROSS TWO QUARTERS AS WEIGHTED AVERAGES OF CENSUS TRACT LEVEL DATA.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ZIP Code	Tract	Quarter	Number of addresses	Raw Weight (ZIP-Tract)	Rescaled Weight	COI 2.0 z-scores	(4) x (5)
L	C	Q1	5	0.625	0.313	0.5	0.156
L	C	Q2	5	0.500	0.250	0.5	0.125
L	D	Q1	3	0.375	0.188	0.2	0.038
L	D	Q2	5	0.500	0.250	0.2	0.050
					ZIP Code L z-score		**0.369**

quarter, two addresses have been added that fall into the area covered by tract D, changing the corresponding weights. In Q1, the segment specific address proportions in column 5 are 0.625 and 0.375 for tracts C and D, respectively. In Q2, the segment specific proportions both are 0.5. Within each quarter, the proportions sum up to one across the different ZIP code-tract segments. To calculate a single ZIP code estimate spanning both quarters, we rescale these proportions to sum up to one across all segments linked to a specific ZIP code. We do so by dividing the proportion for each segment by the number of observed quarters per segment. In this case, each segment is observed twice and we therefore divide the raw proportions in column 5 by 2 in order to obtain rescaled weights (column 6). The rescaled weights now sum up to one across all segments and quarters for ZIP code L. We then multiply rescaled weights with tract-level COI 2.0 z-scores and sum the resulting products to obtain ZIP code J's two-quarter COI 2.0 z-score estimate. The process for allocating population data to the ZIP code level is analogous, but uses Tract-ZIP rather than ZIP-tract weights.

We thus account for changing ZIP code address distributions by averaging over the address distribution for multiple quarters and obtaining a ZIP code estimate that is constant across them. In Table 2, we could have calculated quarterly ZIP code estimates (rather than two-quarter estimates), with each quarter differing from the other because of different underlying weights. This would be misleading: changes between quarters could be interpreted as changes in opportunity. However, opportunity measured at the census tract level did not change, only the underlying weights do. Relatedly, ZIP code level data from different periods should not be compared, because differences reflect, at least in part, changes in the weights underlying the ZIP code aggregates.

Standard deviations for ZIP code level z-scores

When aggregating z-scores from the census tract to ZIP code level, we also calculated weighted standard deviations for each ZIP code z-score. Standard deviations are set to zero if the ZIP code estimate is based on a single census tract.

The standard deviations capture the variability across the census tract z-score values underlying a ZIP code estimate. Standard deviations are larger if the tracts underlying the ZIP code estimate have dissimilar z-scores, and they are smaller if the underlying tracts are more similar. Standard deviations may be downward biased for ZIP code estimates based on very few census tracts.

TECHNICAL NOTES

Data used

We used all HUD-provided, quarterly ZIP-TRACT (used for estimating ZIP code z-scores) and TRACT-ZIP (used for estimating ZIP code population counts) crosswalks, starting with the first quarter of 2012.[4,5] The current release spans the period from 2012q1 to 2020q4. The crosswalks include one row per ZIP code-tract segment, and for each segment contain weights that are defined as the proportion of addresses linked to a specific ZIP code and census tract. ZIP-TRACT crosswalks contain address proportions that sum up to one across ZIP code-tract segments for every ZIP code – like the weights in column 4 of Table 1 – and are used for estimating ZIP code COI 2.0 z-scores. TRACT-ZIP crosswalks contain address proportions that sum up to one across ZIP code-tract segments for every census tract – like the weights in column 7 of Table 1 – and are used for estimating ZIP code population counts. The crosswalks exclude ZIP Codes that are PO Boxes only. We also excluded ZIP codes outside of the 50US states plus D.C. We retained the proportion of total addresses in a given ZIP code-tract segment to be used as weights. To generate multi-quarter and multi-year estimates, we then stacked all quarterly versions of the two crosswalk types (TRACT-ZIP, ZIP-TRACT) into two long data files.

Both HUD crosswalks and COI 2.0 data files use 2010 census tract boundaries. Census tract level COI 2.0 domain (education, health and environment, social and economic) and overall z-scores as well as age 0-17 population counts (based on American Community Survey Summary Files) are taken from data.diversitydatakids.org/dataset/coi20-child-opportunity-index-2-0-database.

Quarterly interpolation of COI 2.0 z-score and population data

Census tract COI 2.0 z-scores and population data are available for two time points, 2010 and 2015. HUD crosswalks are available quarterly. To merge both datasets using census tract FIPS codes, we inter- and extrapolated COI 2.0 z-scores and population data on a quarterly time scale from 2010 to 2020. Between 2010q2 and 2014q4, we calculated weighted averages of the 2010 and 2015 data, where the weight applied to 2010 data declines linearly with time, and the weight assigned to 2015 data increases linearly with time. ZIP code estimates for 2015 and later are fully based on 2015 COI 2.0 data.

Specifically, quarterly weights for 2010 data were defined as

$$(1) \quad w_{2010,t} = 1.05 - 0.05 \times q_t$$

where t indices quarters from 2010q1 and 2014q4 and q is an integer variable running from 1 to 20, corresponding to the 20 quarters between 2010q1 and 2014q4 (inclusive). $w_{2010,t}$ is therefore equal to 1 in 2010q1, 0.95 in 2010q2, …, 0.05 in 2014q4, and 0 in 2015q1.

Quarterly weights for 2015 data were defined as

$$(2) \quad w_{2015,t} = 1 - w_{2010,t}$$

$w_{2015,t}$ is therefore equal to 0 in 2010q1 and 1 in 2015q1 and thereafter. 2010 and 2015 data weights sum up to one in each quarter.

For each census tract level variable y, i.e., COI 2.0 z-scores and population data, and quarter t, quarterly interpolated estimates y_t were calculated as

$$(3) \quad y_{ct} = \left(w_{2010,t} \times y_{c,2010}\right) + \left(w_{2015,t} \times y_{c,2015}\right)$$

where c indicates census tracts, $y_{c,2010}$ is the 2010 value and $y_{c,2015}$ is the 2015 value for census tract c. Therefore, quarterly estimates between 2010q1 and 2014q4 are weighted averages of 2010 and 2015 data. Estimates are equal to 2015 census tract values thereafter.

Finally, we dropped quarters before 2012q1, because 2012q1 is the first year that quarterly HUD crosswalks are available.

Estimation of ZIP code COI 2.0 z-scores

We combined the stacked ZIP-TRACT crosswalk with the quarterly COI 2.0 estimates described in the preceding section. Each row in the resulting data file is a ZIP code-tract segment for a given quarter. We drop segments that belong to tracts with missing COI 2.0 z-score data. In the resulting dataset, most ZIP code-tract segments exist across every crosswalk. Some appear over time, some disappear, and some are observed intermittently. For each segment, the address proportions used as weights may change from quarter to quarter, either because ZIP code definitions change or because the number of addresses in a given ZIP code-tract segment changes.

We then calculate annual and multi-year ZIP code estimates for all single years and combinations of consecutive years for the period from 2012 to 2020. For a full list of files, see page 11 below. For each year or set of years, we first subset the data file to all quarters for the given year (or years), e.g., 2012q1 through 2013q4.

We then calculate the weight for a given segment and quarter as

$$(4) \qquad w_{kct} = p_{kct}/\mathrm{T_k},$$

In a given quarter, each segment s is uniquely identified by the combination of ZIP code k and census tract c. p_{kct} is the proportion of addresses in ZIP code k that are located in census tract c in quarter t. T_k is the number of quarters ZIP code k is observed for. Note that in this example (2012q1 through 2013q4), the denominator of equation 4 would equal 8 for most ZIP codes, because – by construction – the segment-specific weights for ZIP code k sum to one within each quarter, and since eight quarters are combined in this example, the denominator is equal to 8.[2] However, because of missing COI 2.0 data for about 1% of census tracts, we drop all segments linked to tracts with missing COI data. To ensure that w_{kct} sum to 1 within a given ZIP code, we therefore normalize w_{kct} by multiplying them with the sum of w_{kct} for a given ZIP codes, i.e., $\sum_{\mathrm{k=1}}^{\mathrm{T}_k} \sum_{\mathrm{c=1}}^{S_c} w_{kct}$, where S_c is the number of census tracts with non-missing COI data linked to ZIP code k and T_k is the total number of quarters ZIP code k is observed for.

We then calculate the segment-specific weighted z-score as

$$(5) \qquad y_{kct} = w_{kct} \times y_{ct}$$

where y_{ct} is the quarterly interpolated census tract estimate as derived in equation 3 above.

The zip code level z-score $\bar{y}_k$ is then calculated by summing across all y_{kct} for a given ZIP code, or

$$(6) \qquad \bar{y}_k = \sum_{k=1}^{T_k} \sum_{c=1}^{S_c} y_{kct}$$

The ZIP code level standard deviation across z-scores is calculated as

$$(7) \qquad sd_k = \sqrt{\frac{1}{N_k - 1} \sum_{k=1}^{T_k} \sum_{c=1}^{S_c} w_{kct} \times (y_{kct} - \bar{y}_k)},$$

where N_k is the total number of segments across all quarters for a given ZIP code. The ZIP code standard deviation captures both variation in weights w_{kst} and variation in y_{kst} (for pre-2015q1 data). The inclusion of weights also implies that we weigh segments containing many addresses more strongly when calculating the standard deviation, which we also do when calculating the ZIP code mean in equation 5 and 6. Weighting reduces standard deviations as segments with more addresses tend to be closer to the ZIP code mean $\bar{y}_k$ and therefore assigned a greater weight.

---

[2] We drop ZIP codes for which $\sum_{\mathrm{k=1}}^{\mathrm{T}_k} \sum_{\mathrm{c=1}}^{S_c} w_{kct} < 0.5$, where S_c is the number of census tracts with non-missing COI data linked to ZIP code k and T_k is the total number of quarters ZIP code k is observed for. That is, we only retain ZIP codes for which at least 50% of addresses are in tracts with non-missing COI 2.0 data.

Estimation of ZIP code population counts

The approach taken to estimating ZIP code population counts is identical in key respects to the approach taken to calculate ZIP code COI 2.0 estimates. The main difference is that a different set of address proportions, taken from the stacked TRACT-ZIP crosswalk described above, are used to ensure that the total number of children across ZIP codes and across census tracts is identical. While the weights derived in equation 4 above sum up to one within ZIP codes, the weights used here sum up to one within census tracts. We combined the stacked TRACT-ZIP crosswalk with quarterly population estimates (see page 7, for details). Each row in the resulting data file is a ZIP code-tract segment for a given quarter. For each segment, the address proportions used as weights may change from quarter to quarter.

We calculate the weight for a given segment and quarter as

$$(8) \quad w_{ckt} = p_{ckt}/\mathrm{T_c},$$

In a given quarter, each segment is uniquely identified by the combination of census tract c and ZIP code k. p_{ckt} is the proportion of addresses in census tract c that have ZIP code k in quarter t. T_c is the number of quarters a census tract is observed. To adjust for the deletion of segments linked to census tracts with missing data and to ensure that w_{ckt} sum up to 1 within census tracts, we normalize by multiplying the weights with with $\sum_{\mathrm{c=1}}^{\mathrm{T}_c}\sum_{\mathrm{k=1}}^{S_k} w_{ckt}$, where S_k is the number of census tracts with non-missing population data linked to ZIP code k and T_c is the total number of quarters census tract c is observed for.

We then calculate the segment-specific weighted population as

$$(9) \quad n_{ckt} = w_{ckt} \times p_{ct}$$

where p_{ct} is the quarterly interpolated census tract population value for segment s as derived in equation 3 above.

The ZIP code level population estimate $\bar{n}_k$ is then calculated by summing across all $\mathrm{n_{ckt}}$ for ZIP code k, or

$$(10) \quad \bar{n}_k = \sum_{c=1}^{T_\mathrm{c}}\sum_{k=1}^{\mathrm{S_k}} n_{\mathrm{ckt}}.$$

Mapping ZIP Codes to 2015 Metropolitan Areas

In order to calculate metro-normed COI 2.0 metrics, we first need to assign ZIP codes to metro areas. While none of the ZIP codes in our data crossed state boundaries, many ZIP Codes extend across metro area boundaries. We therefore need to first assign ZIP codes to metro areas. To do this, we first map census tracts to 2015 metro areas (100 largest metro areas only), because census tracts are perfectly nested within metro areas, i.e., are either fully inside or outside metro area boundaries. We then link these metro-area mapped census tracts with the HUD ZIP-TRACT crosswalks to estimate the portion of residential addresses inside and outside a given metro area boundary for each ZIP code. For metro-normed COI metrics (those that compare ZIP Code areas only to other ZIP Code areas within the same metro area), we only include zip-codes that have at least 50% of their area inside a given metro area.

Constructing ZIP code level COI 2.0 metrics

Using the ZIP code level COI 2.0 overall and domain-specific z-scores, we then construct nationally-, state-, and metro-normed COI 2.0 Child Opportunity Levels and nationally-normed Child Opportunity Scores. Metro and state-normed Child Opportunity Scores cannot be reliably calculated for all metro areas and states. We essentially follow the same approach outlined on pages 17 and 18 of the COI 2.0 technical documentation[1], except that instead of census tract level data, we are now working with ZIP code level data. For example, nationally-normed Child Opportunity Levels are constructed by ranking all neighborhoods/ZIP codes in the U.S. in terms of the overall COI 2.0 z-score and then dividing neighborhoods/ZIP codes into five ordered groups containing 20% of the child population each, labeled very low-, low-, moderate-, high-, and very high-opportunity. State- and metro-normed Child Opportunity Levels are constructed similarly, by ranking neighborhoods/ZIP codes within a given state or metro area and then assigning neighborhoods to groups containing 20% of the child population in that state or metro. While Child Opportunity Levels divide neighborhoods into five ordered groups, Child Opportunity Scores are constructed by grouping neighborhoods into 100 ordered groups. Each group contains 1% of the child population and is assigned a numerical score from 1 to 100. Because several states and metro areas have fewer (sometimes substantially fewer) than 100 ZIP codes, we only calculate nationally-normed Child Opportunity Scores for ZIP codes.

LIST OF FILES

Only a subset of files is currently published at data.diversitydatakids.org. If you would like to request a file that is currently not publicly available, please submit a request at diversitydatakids.org/contact-us. The specific year or year range is encoded in the file name as either <year>.csv for a single year (aggregate of 4 quarters) or <first_year>_<last_year>.csv for a year range. Years refer to ZIP code definitions, not years in which opportunity data was measured. For example, 2014_2015.csv uses ZIP code definitions from the 2014q1 through 2015q4. See the section “Technical Notes” above for further details. Each file contains COI 2.0 metrics and population data.

File name	Years covered	Number of ZIP codes	Publicly available
2012.csv	2012	38,633	yes
2012_2013.csv	2012-2013	38,657	
2012_2014.csv	2012-2014	38,744	
2012_2015.csv	2012-2015	39,170	
2012_2016.csv	2012-2016	39,194	
2012_2017.csv	2012-2017	39,232	
2012_2018.csv	2012-2018	39,244	
2012_2019.csv	2012-2019	39,276	
2012_2020.csv	2012-2020	39,291	
2013.csv	2013	38,612	yes
2013_2014.csv	2013-2014	38,701	
2013_2015.csv	2013-2015	39,130	
2013_2016.csv	2013-2016	39,160	
2013_2017.csv	2013-2017	39,192	
2013_2018.csv	2013-2018	39,204	
2013_2019.csv	2013-2019	39,236	
2013_2020.csv	2013-2020	39,251	
2014.csv	2014	38,668	yes
2014_2015.csv	2014-2015	39,108	
2014_2016.csv	2014-2016	39,131	
2014_2017.csv	2014-2017	39,163	
2014_2018.csv	2014-2018	39,174	
2014_2019.csv	2014-2019	39,207	
2014_2020.csv	2014-2020	39,222	
2015.csv	2015	39,079	Yes
2015_2016.csv	2015-2016	39,103	
2015_2017.csv	2015-2017	39,134	
2015_2018.csv	2015-2018	39,146	
2015_2019.csv	2015-2019	39,178	
2015_2020.csv	2015-2020	39,194	
2016.csv	2016	39,066	Yes
2016_2017.csv	2016-2017	39,101	
2016_2018.csv	2016-2018	39,113	
2016_2019.csv	2016-2019	39,151	
2016_2020.csv	2016-2020	39,166	
2017.csv	2017	39,050	Yes
2017_2018.csv	2017-2018	39,063	
2017_2019.csv	2017-2019	39,101	
2017_2020.csv	2017-2020	39,118	
2018.csv	2018	39,026	Yes
2018_2019.csv	2018-2019	39,078	
2018_2020.csv	2018-2020	39,095	
2019.csv	2019	39,067	Yes
2019_2020.csv	2019-2020	39,084	
2020.csv	2020	39,069	Yes

REFERENCES

1. Noelke C, McArdle N, Baek M, et al. *Child Opportunity Index 2.0 Technical Documentation*. Brandeis University, The Heller School for Social Policy and Management; 2020. diversitydatakids.org/researchlibrary/research-brief/how-we-built-it

2. Acevedo-Garcia D, Noelke C, McArdle N, et al. Racial and ethnic inequities in children's neighborhoods: Evidence from the new Child Opportunity Index 2.0. *Health Affairs*. 2020;39(10):1693-1701. doi:10.1377/hlthaff.2020.00735

3. Ressler RW, Noelke, C, Acevedo-Garcia D. Does spatial aggregation affect racial/ethnic inequities in access to neighborhood opportunity? In: *American Community Survey User Conference*. ; 2021.

4. U.S. Department for Housing and Urban Development, Office of Policy Development and Research. HUD USPS ZIP code crosswalk files. Published 2020. Accessed January 14, 2020. https://www.huduser.gov/portal/datasets/usps_crosswalk.html

5. Din A, Wilson R. Crosswalking zip codes to census geographies: geoprocessing the U.S. Department of Housing & Urban Development's ZIP code crosswalk files. *Cityscape*. 2020;22(1):293-314.