**SDoH4CWS**

**(Social Determinants of Health For Child Welfare Services)**

CPS Report Rates and Racial Disproportionality Adjusted For Social Determinants of Health:

Data And Programs.

Brett Drake PhD & Dylan Jones PhD, 2025

EARLY BETA TEST VERSION ON LDBase – Summer, 2025

NOTE – ZIPCODE OPTIONS ARE NOT CURRENTLY AVAILABLE ON LDBase, but will be later in 2025

*This project funded as part of DataSMART, under the Center for Innovation in Child Maltreatment Policy, Research, and Training (CICM). NICHD (1P50HD096719)*

Child Protective Services (CPS) Agencies routinely generate and review data showing report rates and racial disproportionality statistics. SDoH4CWS supports states and counties who want to take *the next step* and understand local reporting rates and racial disproportionality while also considering Social Determinants of Health (SDoH).

**SDoH4CWS has two parts.** We provide the first part. Each state also provides one part from their own internal data.

**SDoH4CWS Part 1: Programs and Datasets We Provide**

**The first two parts** are SDoH data and programs we provide in the “SDoH4CWS” download package.

 (1) Data files with Social Determinant of Health Measures (total and “same-race”) at the County and Zipcode levels for 2022. The SDoH measures are derived from the 2017-2022 American Community Survey 5 year estimates. These files are “**County2022SDoH.sas7bdat**” and “**Zipcode2022SDoH.sas7bdat**”. More files for different years and tract level files are forthcoming.

 (2) Programs in SAS and R which allow states or counties to better understand their CPS reporting rates and racial disproportionality while taking SDoH into consideration. The following programs are available and are described in detail below:

* **SDoHCounty2022PredVsActual.sas**
* **SDoHCounty2022Scatterplot.sas**
* **SDoHCounty2022Disp.sas**
* **SDoHZip2022PredVsActual.sas**
* **SDoHZip2022Scatterplot.sas**
* **SDoHZip2022Disp.sas**

**SDoH4CWS Part 2: Zipcode and County .csv Files You Provide**

*(Important note: A version of “County.csv” is already included. This is for testing and troubleshooting only. These are simulated data and NOT REAL, they do not include estimates for any groups but White, Black and Hispanic – you will need to replace it with your own county.csv file as described below)*

**The third part** of SDoH4CWS is an input file which must be generated by the state or county. That file includes the following ***numeric*** variables at either the zipcode or county levels. *Variables are counts of the number of persons of that race (and total) who were reported to CPS during the year in question.* If you want more stable counts, you can make them five year (or whatever) totals, and divide by the number of years. There is more detailed information about how to create these files in the documentation within the above “SDoH4CWS” programs.

**“Zipcode.CSV” file structure: All variables numeric. You create and provide this file.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ZIPCODE | TotWh | TotBl |  | TotAs | TotNH | TotAI | TotHi | Tot |
| 63005 | 14 | 0 |  | 2 | 0 | 0 | 0 | 16 |
| 63010 | 455 | 2 |  | 5 | 0 | 0 | 11 | 473 |
| 63011 | 68 | 1 |  | 4 | 0 | 0 | 2 | 75 |
| 63012 | 47 | 0 |  | 0 | 0 | 0 | 1 | 48 |

**“County.CSV” file structure: All variables numeric. You create and provide this file.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FIPS | Tot | TotBl | TotWh | TotAs | TotNH | TotAI | TotHi |
| 6001 | 7929 | 2404 | 963 | 707 | 47 | 20 | 2752 |
| 6007 | 1230 | 44 | 672 | 16 | 0 | 40 | 151 |
| 6013 | 6421 | 1563 | 1597 | 385 | 39 | 18 | 2223 |
| 6017 | 1457 | 18 | 616 | 8 | 0 | 15 | 222 |
| 6019 | 19051 | 2242 | 2370 | 700 | 11 | 85 | 11441 |

“FIPS” is the standard county level FIPS code for any county. This is usually a 5 digit value, e.g. “29510” is Missouri’s (29) St. Louis City (510). The above example is only 4 digits as the FIPS code for California is a single digit “6”. If the state or county only wants to check report rates and not assess racial disproportionality (in other words, only use the “…PredVsActual” programs), only the “Zipcode / FIPS” and “Tot” columns are needed.

For your convenience, we provide example county data “County.CSV” for Ohio based on NCANDS and we also produce fabricated (entirely fake) data at the zipcode level for Missouri in “Zipcode.CSV”.

**Running the Programs:**

The following steps will allow you to run the programs:

1. To run the county programs, you need to create “County.CSV” as described above. Zipcode programs require the “Zipcode.csv” file to be created.
2. Open the file program you want to run (example: SDoHCounty2022PredVsActual). Use the search function to identify each time “TODO” comes up. These are places you must change the path to the directory you are using. We assume you will put everything in one directory. Some programs have “TWEAK” noted in the program – these are things you can change. The most important one is setting the population threshold for each area. Using very small child population areas (say, 30 children in a given zipcode) will produce very unreliable results. Setting this threshold allows you to look at larger (more stable) areas at a threshold you choose. Default is 500 children.
3. Run the program. The output will pop up in your directory. Output is a .CSV file for everything except the scatterplots and a “.png” file for scatterplots.

**The Programs:**

The below text explains what each program does. You can look a the “OUTPUTEXAMPLE…” files provided to see what the output looks like. The outputs come to you in .CSV files, but the scatterplot comes in a .PNG file. For the “OUTPUTEXAMPLE…” files we include “.XLSX” files with the bar charts added manually. To do this, open the .CSV file in Excel. Delete cell “A1”. Highlight the data. Then hit the Insert and bar graph icon tabs as shown. It’s that easy! If it takes you more than a minute, there’s probably an issue. You can add titles or labels and clean up the font size and graph colors if you want.

**SDoHCounty2022PredVsActual.sas** or **SDOHZip2022PredVsActual.sas** produce data output in a csv file with three variables, the area (county name or zipcode) the report rate per 1000 children and the predicted rate using a regression model (PROC GENMOD in SAS) controlling for a range of variables (For SDoH variable sourcing and description see provided “ACS Codebook”). The bar chart does not pop up automatically but can me made as follows: (1) delete the data in cell “A1”. (2) highlight A1-last “C” column with data. (3) hit “Insert” on top bar and then the first bar chart Icon. (4) resize and add title. It should take under a minute to add the bar chart once you try it a time or two. The purpose of this chart is to show both what the actual rate of maltreatment is in each county, but also to see what the computer “expects” the rate to be for each county once 14 SDoH variables are considered (Urban/Rural, Births to Married Families, Commute Time, Food Stamps, Health Insurance, Housing Value, Housing Units Occupied, Owner Occupied Units, Unemployment, High School Noncompletion, Families in Poverty, Percentage Married, Median Household Income, Persons Working).



**SDoHCounty2022Scatterplot.sas** or **SDOHZip2022Scatterplot.sas** produce data output in a .png file. Just open that file to look at it. NOTE- your computer may be configured to show it under the SAS results tab, not export it as a .png file. The output should look like this This file produces TWO points on the scatterplot for each area. One plot is the White report rate for the area vs. the White income for that area (e.g. the number of White children in 63130 against the White Median Family Income in 63130). The other point is for the Black report rate vs. Black Median Family Income. This allows you to see an “apples vs. apples” comparison of racial disproportionality, with each race’s report rates shown against that race’s income level for each data point (again, two data points for each race). The output looks like this (Ohio data):



**SDoHCounty2022Disp.sas** or **SDOHZip2022Disp.sas** compute disparities by race/ethnicity (Black or other nonwhite race against White race). Two disparites are presented – the actual disparity (minority rate/White rate) and the SDoH predicted disparity rate using the 14 SDoH variables. This allows you to see both the current racial disparity as it exists, and what the model thinks that the disparity “should be” when SDoH factors are considered. Why “should” disparity ever be more than 1:1? Because different races face very different SDoH contexts. This program takes one data point for each area from each race, combines them into a single file, and compares the rates of maltreatment reporting to the same-race SDoH variables (e.g. white models use SDoH variables specific to Whites- White unemployment rate, for example). This results in a predicted value for each data point (e.g. a predicted value for Black people in San Diego, one for LA, and similar predicted White data points). These predicted data points are used to create an SDoH-predicted disparity rate. So in very coarse terms, the “actual” disparity rate is the degree to which a minority group is over (>1) or under (<1) reported compared to Whites using actual report data and no SDoH controls. The SDoH “predicted” disparity is what the computer would expect the disparity to be given the different SDoH conditions experience by each race. Example output:



So that’s what the programs do. Have fun!

**SDoHCountyDispComparisons.sas**: This program calculates Black/White and Hispanic/White disparities for CPS reports and for various non-CPS outcomes *when those outcomes are available from County Health Rankings at the county level*. Those outcomes include disparities in Infant Mortality, Child Mortality and Teen Birth rates. This program allows you to look at racial differences in CPS reporting and see if they are different from differences in other outcomes at the county level. For example, in the graph below, Franklin county’s CPS disparity is virtually identical to disparities in Child Mortality and Teen Birth Rate, but is lower than the disparity in Infant Mortality. The “blank” bars are counties with missing data in the County Health Rankings database.



**Quick Guide to SDoH variables**

All variables are from the ACS 5 year estimates as accessed through Social Explorer. Complete codes for all variables used are available in the attached ACS Codebook Excel file. If you want to see how the variables are created, see the subfolder with original data creation programs. The variables represent the following:

SDoHPop ="Persons <18yo";

SDoHMarried ="% fams w kids (are married)";

SDoHFamPov ="Pct of all fams in poverty";

SDoHMedFamInc ="Median Family Income";

SdOHEdDrop ="Pct HS Dropout (=>25yo)";

SDoHWorking ="%16-64yo in Labor Force";

SDoHTrans ="% w own car/pooling to work";

SDOHHIns ="Persons <18y w Health Insur";

SDoHUnemp ="% 16-64yo who are Unemployed";

SDOHFStamps ="Persons w food stamps";

SDOHOwnerOcc ="Pct owner occupied units";

SDOHHouseValue ="Median Home Value";

SDOHOccupied ="Pct of housing units occupied";

SDOHCommTime ="Commute time in minutes";

SDOHBirthM ="% births to married moms";

Name ="County Name";

IMPORTANT NOTE: These variables are also available for each race. For example, the Hispanic variables are:

SDOHPopSRHi ="Hispa Persons <18yo";

SDoHMarriedSRHi ="Hispa Pct of all families w kids who are married couples";

SDoHFamPovSRHi ="Hispa Pct of all families in poverty";

SDoHMedFamIncSRHi="Hispa Median Family Income";

SDOHEdDropSRHi ="Hispa Pct HS Dropout (=>25yo)";

SDoHWorkingSRHi ="Hispa Pct of all 16-64yo who are working in Labor Force";

SDOHHInsSRHi ="Hispa Persons <18y w Health Insur";

SDOHTransSRHi ="Hispa Persons driving to work in car";

SDoHUnempSRHi ="Hispa Pct of all 16-64yo who are Unemployed";

SDOHFStampsSRHi ="Hispa Persons recieving food stamps";

SDOHBirthMSRHi ="Hispa Pct of births last yr to married moms";

THESE VARIABLES ARE AVIALABLE FOR ALL RACES. The “SR” means “Same Race”. “SRHi” means “SRHispanic”. Other races are Bl/Black, Wh/White, As/Asian, AI/American Indian or Alaskan Native, NH/Native Hawaiian or Pacific Islander. Other and multiple race are not included due to inconsistencies between how Census and state CWS assign these values,

**Extra: SDoH Indexes Are Available:** The following pre-existing and publicly available SDoH indexes are provided in case you are interested: ADI, BRIC, COI, CRE, SDI, SVI, SoVI, YI. In our view, the ADI and the YI are the best for child welfare purposes (show strongest correlations). These are provided as variables in the County2022SDoH file. Not all these values are from the same year, see SDOHIndexDocumentation for full details. In addition, the County2022SDoH file includes descriptive variables for counties for such things as rates of low birth weight and homicide rates. These are used in the County2022DispComparison program and may be of interest to you in any case. These come from the County Health Rankings website as noted in the documentation file above.

Questions can be addressed to Brett Drake at brettd.wustl.edu or 314 805 8422.

---------------------

*Note – zip county crosswalk to 2013 data linkages from 2013 HUD USPS Crosswalk files at* [*https://www.huduser.gov/portal/datasets/usps\_crosswalk.html*](https://www.huduser.gov/portal/datasets/usps_crosswalk.html) *Zipcodes were associated with the county containing the largest part of the zipcode (usually 100%).*