

# What the Supplemental Demographic and Housing Characteristics File from the 2020 Census Tells Us About Future Statistics on Children from the Census Bureau

By

William P. O'Hare <sup>1,2</sup>

October

In mid-September 2024, the Census Bureau released the Supplemental Demographic and Housing Characteristics (S-DHC) file which is the last of the 2020 Census datafiles. It is the first and only 2020 Census file which uses a variant of Differential Privacy (DP) that keeps the connection between children and their parents in the data processing step. The S-DHC file uses a variant of differential privacy called PHSafe which is different than the one used for most of the other datafiles from the 2020 Census. This has important implications for data availability.

The S-DHC file is described by the Census Bureau as:

“•Combined characteristics about households and the people living in them. For example, the count of people living in certain types of households and housing units.”

---

<sup>1</sup> Consultant to the Count All Kids Campaign

<sup>2</sup> This research was funded by the Census Equity Initiative, but it is not responsible for the content of the publication.

- Data on population in households, including average household size by age and tenure, average family size, household/family type of people under 18 years and total population in households by tenure.

These tables are also known as the person-household join tables. “

This last bullet point is very important because it has huge implications for the amount of data that is likely to be released from a census or survey and the implications are particularly important for the child population.

As noted above, person-household join tables are important because they provide two demographic characteristics of a person compared to only one characteristic as in the tables in the DHC, DHC-D and DHC-B files. According to the Census Bureau (2024, page 2) related to join tables, “Combining the data increases the risk of disclosing information about individuals because information for each person in the household (especially the householder) is linked to the information for everyone else in the household. This interrelationship makes it much harder to obscure the effect of one person’s record has on the others, which in turn make it hard to guarantee that they are protected.” For example, one cell of Table PH6

indicates children in married-couple families thus providing two characteristics (age and family structure) for persons in that cell.

With respect to the S-DHC file, the Census Bureau (2024, page 1) says, “Combining details about the structure of household and the people living in them required stronger disclosure avoidance techniques.”

Stronger disclosure avoidance techniques seem to translate into less data provided.

It makes sense that tables that provide more characteristics for an individual need a higher level of disclosure protection, but I was a little surprised by how much this reduces the data provided.

Probably the most notable thing about the S-DHC file is the dearth of detailed data in contrast to the DHC, DHC-A and DHC-B files. It appears that use of a different variant of DP in the S-DHC file greatly limits the data that can be released compared to DHC, DHC-A and DHC-B.

The S-DHC file only has 8 tables and only one of those tables provides data for young children. Table 1 below provides detailed information about the eight tables in the S-DHC file.

2020 S-DHC Table Number	Title	2020 Geographies
PH1*	AVERAGE HOUSEHOLD SIZE BY AGE	Nation and State
PH2	HOUSEHOLD TYPE FOR THE POPULATION IN HOUSEHOLDS	Nation and State
PH3*	18 YEARS	Nation and State
PH4*	POPULATION IN FAMILIES BY AGE	Nation and State
PH5*	AVERAGE FAMILY SIZE BY AGE	Nation and State
PH6	FAMILY TYPE AND AGE FOR OWN CHILDREN UNDER 18 YEARS	Nation and State
PH7*	TOTAL POPULATION IN OCCUPIED HOUSING UNITS BY TENURE	Nation and State
PH8*	AVERAGE HOUSEHOLD SIZE OF OCCUPIED HOUSING UNITS BY	Nation and State
TABLES WITH AN ASTERISK (*) ARE REPEATED FOR NINE MAJOR RACE/ETHNIC GROUPS		

Here are a few examples showing how detailed data is restricted in the S-DHC file compared to other DHC files. The DHC, DHC-A and DHC-B files provide data for Nation, States, Counties, Census Tracts, Places, and American Indian/Alaskan Native/Native Hawaiian areas while the S-DHC only provides data for the nation and states.

The DHC-B file provides data for approximately 1,500 detailed race/ethnic groups, while the S-DHC only provides data for nine main race/ethnic groups.

Extremely limited age details are available from the S-DHC file and age detail is an especially important consideration for children. There are enormous differences between preschoolers and teenagers.

Only one of the eight tables in the S-DHC file provides data for young children while DHC, DHC-A and DHC-B provide much more age detailed for children. However, that table (PH6) is one of only two tables in the S-DHC file that is not repeated for race and ethnic groups. Oddly, the age breaks used in this table do not show age data for 0 to 4 separately, they show ages 0 to 3. This is at least partly related to a desire to limit potential re-identification.

The point is the S-DHC file provides much less data than other DHC files. Compared to the other DHC files the data from the S-DHC files are restricted in terms of geographic levels, race/ethnic groups, and age breaks.

This is likely due to the tighter requirements for privacy protection needed for the join tables. If the use of a PHSafe-like variant of differential privacy means the information that can be released will be greatly curtailed, that has important implications for the use of differential privacy in Census Bureau surveys such as the American Community Survey (ACS), the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP), which contain a lot of join tables.

The fact that the S-DHC file is the only 2020 Census file to use the PHSafe DP variant and there is little detail made available in the S-DHC suggests implementation of a PHSafe-like variant may limit data in other Census Bureau data products like the ACS.

### DP Breaks the Link Between Child and Parents

One of the biggest differences between the S-DHC file and other DHC files is how DP processes cells connecting children and their parents. In the 2020 Census data files prior to S-DHC being released, differential privacy was applied to cells of tables independently. For example, if the table for a census block showed 2 people over age 18 and 2 people under age 18, noise was applied to each cell independently. DP is administered to children and parents independently, so it may eliminate the adults in a household that has children by randomly subtracting data from the number of adults. Thus, if DP subtracted 2 people from the cell for age 18 plus and added two people to the number of people under age 18, it would result in a census block with four children (under age 18) and no adults (over age 18).

In processing millions of census blocks (and other small geographic units) the application of noise produced a large number of impossible or

improbable occurrences. One example of the kinds of problems caused by not keeping children linked to their parents is the production of many blocks where there are children, but no adults may be related to the link between children and adults in a household that is broken when 2020 Disclosure Avoidance System (DAS) with DP was applied to the DHC file. If the processing retained the link between young children and their parents in a household, it is doubtful that there would be such a high number of blocks with children and no adults. In producing a demonstration file using the 2010 Census, implementation of DP resulted in over 163,000 census blocks with children but no adults, compared to just 82 such block prior to application of DP (O'Hare 2022).

Keeping the link between children and their parents in a dataset has important consequences for child researchers and child advocates who use statistical measures of child wellbeing. Many of the most important measures of child well-being require a linkage between a child and his or her parents. For example, the poverty status of a child, which is measured in the American Community Survey, is linked to the income of older household members.

This statistical disconnection of children and parents based on DP is an ongoing concern and is likely to have important impacts in other Census

products which have historically had more detailed data on young children. For example, the connection between children and parents is critical for a lot of data from the American Community Survey. A table that involved poverty status by race by age by family structure provides a lot of important information for child advocates and child researchers, but it involve a combination on 4 dimensions .The importance of the ACS for child well-being measures is underscored by the fact that the KIDS COUNT project at the Annie E. Casey Foundation (The Annie E. Casey Foundation 2024) uses 16 indictors to rank states on overall child wellbeing and 9 of them are from ACS. Moreover, some of the indicators used in the KIDS COUNT Data Book are relatively complex meaning they reveal many characteristics of an individual simultaneously and therefore may be eliminated or reduced in scope in the context of privacy protection. A list of the 9 ACS-derived indicators used in the KIDS COUNT Data Book is provided below.



Child Well-Being Indicators Used in the KIDS COUNT Data Book
Child poverty rate
Parents lack secure employment
Household has high housing costs
Teens not in school and not working
Young child not in preschool
Child without health insurance
Children in single-parent families
Children in families where household head lacks a high school diploma
Children living in high-poverty areas

A recent survey of child advocates also indicates the ACS is widely used (O’Hare 2022). The summary of that study (page 16) says, “The data analyzed here underscores the extent to which state and child advocacy organizations make extensive use of the data from the ACS in a variety of ways I doubt there is any other single source of data as important as the ACS for state and local child advocates or researchers.” Other prominent child-focused groups such as Zero-to-Three, members of the Partnership for America’s Children, and First Focus also rely extensively on data from the ACS

With regard to uses of DP in ACS the Census Bureau (2022, page 1) has said, “Our current assessment is that the science does not yet exist to comprehensively implement a formally private solution for the ACS. We expect a multiyear development period, including data user review and

feedback that will extend beyond 2025.” As the Census Bureau pursues this multi-year research effort, I hope it includes factors such as availability of data and timeliness of data as well as the tradeoff between privacy protection and data accuracy in its decision-making

The Census Bureau is wise to delay implementation of a new DAS in the ACS until it has thoroughly tested it. Many data users feel the Census Bureau rushed the DP implementation in the 2020 Census. According to the National Academy of Sciences study assessing the 2020 Census (National Academy of Sciences, page 17), “In short, the new DAS was not ready for use in the 2020 Census production, degrading the value of the 2020 Census data products in terms of both quality and timeliness.” It is critically important that the Census Bureau not make this mistake again with the ACS

Despite the assurances from the Census Bureau that they will undertake a long and thorough research effort in determining how to apply any new privacy/confidentiality approach in the ACS, the results of the 2020 Census use of PHSafe DP variant for disclosure avoidance related to join tables leaves me very concerned about losing important data on the well-being of children from Census Bureau surveys.

## References

O'Hare, W. P. (2022). "USE OF THE AMERICAN COMMUNITY SURVEY DATA BY STATE CHILD ADVOCACY ORGANIZATIONS," *March 18, 2022*  
Posted on Count All Kids website <https://countallkids.org/resources/use-of-the-american-community-survey-data-by-state-child-advocacy-organizations/>

O'Hare, W.P. (2022). Analysis of Census Bureau's March 2022 Differential Privacy Demonstration Product: Implications for Data on Young Children, Posted on the Count All Kids website,  
<https://countallkids.org/resources/analysis-of-census-bureaus-march-2022-differential-privacy-demonstration-product-implications-for-data-on-young-children/>

The Annie E. Casey Foundation 2024). The 2024 KIDS COUNT Data Book, The Annie E. Casey Foundation, Baltimore, MD  
[https://www.aecf.org/resources/2024-kids-count-data-book?gad\\_source=1&gclid=CjwKCAjwgfm3BhBeEiwAFfxrG\\_8afLLAgVBhnN0zGef00pGJNbPvo9Y\\_R-RXwOb5mGMnzfmZsYeEfxoC\\_MAQAvD\\_BwE](https://www.aecf.org/resources/2024-kids-count-data-book?gad_source=1&gclid=CjwKCAjwgfm3BhBeEiwAFfxrG_8afLLAgVBhnN0zGef00pGJNbPvo9Y_R-RXwOb5mGMnzfmZsYeEfxoC_MAQAvD_BwE)

U.S. Census Bureau (no date ) <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/release/about-2020-data-products.html>

U.S. Census Bureau (2022). Disclosure Avoidance Protections for the American Community Survey, Donna Daily, December 14  
<https://www.census.gov/newsroom/blogs/random-samplings/2022/12/disclosure-avoidance-protections-ac.html>

U.S. Census Bureau (2024a). "2020 Census Supplemental Demographic and Housing Characteristics File (S-DHC)"  
<https://www.census.gov/data/tables/2024/dec/2020-census-s-dhc.html#about>

U.S. Census Bureau (2024b) Understanding the Supplemental Demographic and Housing Characteristics File, Stacy Vidal, Evan Brassell, Michael Hawes, Sept 16

