



POLIDATA® Political Data Analysis

DATABASE DEVELOPMENT, ANALYSIS AND PUBLICATION;
POLITICAL AND CENSUS DATA; LITIGATION SUPPORT

CLARK BENSEN

POLIDATA LLC · 1303 HAYWARD RD, P.O. BOX 530 · CORINTH, VT 05039

Tel: 703-690-4066 · Fax: 202-318-0793 · email: clark@polidata.org

PUBLISHER OF THE POLIDATA® DEMOGRAPHIC AND POLITICAL GUIDES AND ATLASES

Honorable Steven Dillingham, Director
U. S. Bureau of the Census
4600 Silver Hill Road
Washington, DC 20233

10 Apr 2020

Re: DAP2020

Director Dillingham,

This letter raises some concerns that I, as one who has been involved in districting projects since the 1980 Census, has about the Disclosure Avoidance Program (DAP). This is briefly described on a Bureau webpage entitled, "Statistical Safeguards":

Before we publish any statistic, we apply safeguards that help prevent someone from being able to trace that statistic back to a specific respondent.

We call these safeguards "disclosure avoidance," although these methods are also known as "statistical disclosure controls" or "statistical disclosure limitations."

Although it might appear that a published table shows information about a specific individual, the Census Bureau has taken steps to disguise the original data in such a way that the results are still useful. These steps include using statistical methods such as "data swapping" and "noise injection."

Before Census 2000 a similar issue faced the Bureau with regards to adjustment of the census counts. Congress enacted a statute¹ which addressed "Statistical Sampling or Adjustment" in the decennial. Important concerns of Congress expressed in the findings to PL105-119 are: "(5) the decennial enumeration of the population is one of the most critical constitutional functions our Federal Government performs; (6) it is essential that the decennial enumeration of the population be as accurate as possible, consistent with the Constitution and laws of the United States[;]."

The Supreme Court addressed that situation in an opinion announced on January 25, 1999², 14 months before Census Day 2000, "States use the population numbers generated by the federal decennial census for federal congressional redistricting. See *Karcher v. Daggett*, 462 U. S. 725, 738 (1983) ("[B]ecause the census count represents the 'best population data available,' . . . it is the only basis for good-faith attempts to achieve population equality" . . .).

While the Commerce case focused largely on sampling, the act is more expansive and another of its findings is: "(7) the use of statistical sampling or statistical adjustment in conjunction with an actual enumeration to carry out the census with respect to any segment of the population poses the risk of an inaccurate, invalid, and unconstitutional census[;]."

A review of the language in section (h) of the findings provides a definition of what the term 'statistical method' means. This definition includes "or any other statistical

¹ See Pub. L. 105-119; Sec. 209 (a) (5) [congressional findings] Statistical sampling or adjustment in decennial enumeration of population; <https://uscode.house.gov/statviewer.htm?volume=111&page=2480>

² See *Department of Commerce v. United States House of Representatives*, 525 US 316 (1999); (98-404); argued November 30, 1998; decided January 25, 1999.

procedure, including statistical adjustment, to add or subtract counts to or from the enumeration of the population as a result of statistical inference[;].”

My main concern is with respect to districting³ and is that if the Bureau implements the DAP as it is currently envisioned the thousands of entities across the nation that are responsible for revising current, or creating new district, boundaries for representative government at the state and local level will not have the “best population data available” and therefore will not be able to make good-faith attempts towards equality. I offer these comments with the understanding that many of the general concerns will be shared by numerous redistricting stakeholders once they know about DAP. Moreover, I believe there is general agreement regardless of political affiliation on this issue.

This is simply a question of process. The entities responsible for districting need to know, before the numbers are released in less than one year, that the numbers they receive will be sufficient to meet their critical need and that their own election calendars will not be disrupted by additional litigation over the numbers used to distribute political representation across their states or localities.

This is not a concern about the goals of the DAP to avoid inadvertent disclosure of personally identifiable information (PII). I believe there is substantial agreement that the privacy of certain individuals is a laudable aim in 2020⁴. However, it appears that the DAP presents a fundamental interference with the constitutional purposes of apportionment by reliance upon a statutory concern relating to privacy.

While a supplement to this letter will discuss some of the concerns shared by redistricting stakeholders, they will be listed below.

- 1) Adjusted numbers will not be “the best available population data”.
- 2) Stakeholders will be unable to “make good faith efforts” at equality.
- 3) Use of such a statistical method “poses the risk of an inaccurate, invalid, and unconstitutional census”.
- 4) Additional litigation over the numbers may result in distraction, delay, and costs to many districting entities.
- 5) The confidence amongst state and local governmental entities in the entire census process may be severely undermined.
- 6) While the Bureau is a national statistical agency, first and foremost it is the compiler of the “actual Enumeration” to fulfill the constitutional mandate.
- 7) Previous methods for disclosure avoidance were less pervasive. Because the previous methods were simpler techniques such as data swapping, rounding, top-coding, etc., the degree to which information was adjusted for protection was much less. On the other hand, the DAP for 2020 will affect every level of geography and the population counts.
- 8) Relative inaccuracy and bias in the DAP: “The new method allows us to precisely control the amount of uncertainty that we add according to privacy requirements.”

As discussed above, the implementation of the DAP is quite likely to affect redistricting stakeholders across the nation. It appears that there are several options available to the Bureau at this point.

- 1) **Continue with research but still implement DAP.** Of course, the Bureau could discount the concerns of the (currently) small group of stakeholders and local statistical entities and

³ However, given the feedback from the so-called Demonstration Data during 2019 there are other concerns, such as distribution of intergovernmental aid, that may motivate others to comment on the DAP.

⁴ Nevertheless, privacy was not an issue when the census was first taken. In fact, the first Census Act required the schedules to be posted for public review before they were submitted to the federal marshal. Specific requirements for privacy appear to have first been codified for the 1880 Census.

- proceed as currently planned. Nevertheless, based upon the most recent information from working groups it appears that while improvements may be made to the range of error introduced by noise injection, the counts will still not be available for most levels of geography.
- 2) **The Black Box Engine.** Some observers have suggested that districting entities could submit any plan of interest to a website whereby the unadjusted counts could be applied and thus the plan drafters could know expeditiously how far off their numbers were from equality. Aside from the obvious logistical issues for such a process it fails for the want of transparency.
 - 3) **Reduce the cross tabulations of data tables.** This could apply in a general sense to whatever cross tabulations that the Bureau provides. Such breakdowns appear to be largely developed by the Bureau for the use of federal, state, and local governments in their mission to fulfill their requirements for purposes other than apportionment.
 - 4) **Reduce the breakdowns of data tables into fewer cells.** The critical dataset for redistricting, the so-called PL94 dataset⁵ was, prior to Census 2000, a fairly simple dataset with a much smaller set of variables. With the addition of the multi-race response options in 2000 the number of data cells for the PL dataset expanded greatly. On its face this presents numerous privacy concerns even for areas that have a substantial number of persons because all six races are tallied for all multiple combinations. The level of detail in the PL94 dataset for each record is not needed by most districting entities and could be collapsed substantially and then DAP adjustments as previously done to the characteristic data could be undertaken.
 - 5) **Invariant Block Counts without Characteristic Information.** Another alternative would be to hold invariant the counts of population and housing⁶ and to simply provide no characteristic information at the block level. Choices for such an alternative could be a) include characteristic data only for areas at a specified geographic level or with counts above a threshold, as has been done with Special Tabulations previously, and/or b) have districting entities rely upon characteristic data from the American Community Survey (ACS).

Clearly, the perspective of districting stakeholders and local planning agencies is likely to something other than Option 1⁷. Because districting is done for so many types of entities there are varying degrees of resources and needs. Yet, considering the range of variations that are likely to be seen when a user compares the adjusted numbers to information they have independently collected over the decade, there are going to be a lot of queries. One would expect that local officials may find significant differences because they can spend the time to review the information, block by block. What does the Bureau propose for the Count Question Resolution process for Census 2020?

Other stakeholders may weigh in on this issue as well offering different options or perspectives. However, Options 4 and 5 at least appear to several stakeholders as being viable options. Option 4 could impose a burden on a relatively small number of entities but may not appease the concerns of the Bureau for privacy. Option 5 would affect substantially more entities but at least there is some alternative source of data that would provide less precision for the characteristic data and more statistical analysis for districting entities to comply with Voting Rights Act concerns. Nevertheless, even accepting Options 4 or 5 would be a substantial compromise for some stakeholders but if the only viable option for privacy is the DAP many stakeholders would likely choose one of the above or some other alternative not yet discussed.

⁵ See Pub. L. 94-171. <https://uscode.house.gov/statutes/pl/94/171.pdf>

⁶ Total Population and Voting Age Population, as well as the information on Housing Units and Group Quarters.

⁷ N.b., while there may not be much difference of opinion about the overall concern, there may well be with respect to options.

Concerns about Disclosure Avoidance Program
Polidata ® Political Data Analysis, Clark H. Bensen, Page 4

Respectfully yours,

/s/ Clark H. Bensen

Clark H. Bensen

Enclosures:

1) Supplement

[2020-0410a]

CC:

Honorable Wilbur Ross, Secretary
U.S. Department of Commerce
1401 Constitution Ave, NW
Washington, DC 20230

SUPPLEMENT

Introduction. For the sake of readers of this letter for whom Disclosure Avoidance is a new concept the following brief summary is provided. It is important to understand the widespread degree to which the counts from the ‘actual Enumeration’ are likely to be affected by the DAP.

In December of 2019 a conference was held that reviewed the results from the Bureau’s efforts of the application of the DAS to the 2010 Census data. Based upon information published by the Bureau during October 2019⁸ and additional material published subsequent to the December 2019 conference and recent meetings of the Expert Group (which now includes at least one for redistricting) it is still unclear exactly what the actual plan for the Bureau is or will be. Moreover, it appears that the current schedule is that final policy decisions will not be made, for the design of the DAS, until September 2020⁹.

Currently, the best information of the degree to which numbers eventually reported for the 2020 Census can only be gleaned from the information provided in the October 2019 memo which detailed the status of these numbers for the review of the 2010 Census data. In other words, the plan, at that point, was that some numbers would be ‘invariant’, that is, the reported number would be the enumeration counts and no alteration for privacy would be made, while others will be ‘variant’, that is, the numbers reported would be altered for privacy protection.

That proposal would treat only three types of counts as invariant: a) the state total population; b) the number of housing units in a census block; and c) the number and type of group quarters in a census block¹⁰. In other words, below the state, every number provided by the Bureau will not be a tabulation of the responses from an ‘actual Enumeration’ but the result of a statistical alteration. “Differential privacy allows us to inject a precisely calibrated amount of noise into the data to control the privacy risk of any calculation or statistic.”¹¹

Additionally, there is the question as to which metrics will be released with the adjusted numbers to allow users to assess the degree to which noise has been added. A recent March 2020 presentation¹² primarily addressed “making population counts more accurate” and reviewed numerous metrics that might “allow the public to see the improvements that are made” as the Bureau continues to test their DAS operations.

At this point it is an open question as to whether this will substantially change so that the block counts would be delivered as enumerated or adjusted. Regardless, what this indicates is that we are now less than one year away from releases of the numbers and the Bureau still does not know with any precision what method they will use or metrics they will provide. Notably, the implementation of disclosure avoidance will not be applied to the American Community Survey (ACS) until 2025¹³. Why is it that the purposes of apportionment will be the first real test case for such a statistical adjustment?

⁸ See Memorandum 2019.25: 2010 Demonstration Data Products – Design Parameters and Global Privacy-Loss Budget; https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/memo-series/2020-memo-2019_25.html

⁹ See Updates and DAS Development Schedule, March 18, 2020; <https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/disclosure-avoidance-system/2020-03-18-updates-das-development-schedule.pdf?#>

¹⁰ See the Bureau site: https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html

¹¹ See the Bureau site: <https://www2.census.gov/about/policies/2020-03-05-differential-privacy.pdf?#>

¹² See 2020 Census Disclosure Avoidance Improvement Metrics; <https://www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/disclosure-avoidance-system/2020-03-18-2020-census-da-improvement-metrics.pdf?#>

¹³ See <https://www.census.gov/newsroom/blogs/random-samplings/2019/07/boost-safeguards.html>

One issue that appears to have concerned the Bureau over the threat of what they term as reconstruction of the census appears to be the result of the extraordinary level of detail that is provided by two data products: a) the block-level data provided pursuant to PL94-171 and b) the numerous cross-tabulation tables that are provided by the Bureau of data at numerous levels of census geography.

The block-level data is the critical dataset for most redistricting stakeholders. Blocks have a huge range of population across any geographic area. Many have no population because they are industrial areas, or parks, or bodies of water, or highways, or mountains, or wide open range, or simply vacant housing. Many have a handful, and many have thousands, of persons. But, blocks are used as the lowest level for most districting datasets, generally because of a few factors that make them unusual amongst all the so-called 'summary levels' that the Bureau recognizes.

These characteristics of census blocks include:

- 1) they are the lowest level for which the counts have heretofore been tabulated and made available;
- 2) they cover the entire non-coastal geographic area of a state or locality;
- 3) pursuant to the Block Boundary Suggestion Project (BBSP) the states have the ability to designate the boundaries of the blocks;
- 4) tabulations generally account for how the block fits into higher levels of geography, such as Voting Districts (VTDs) the boundaries of which are designated by many states as Phase 2 of the BBSP;
- 5) the reported counts for every higher level of geography has been simply the sum of the information for all corresponding blocks;
- 6) redistricting stakeholders form one of the few groups that rely upon the block-level information as the critical data needed to fulfill their need, that is, the purposes of apportionment; equalizing population would be considerably more difficult if higher level information was the only level for which accurate data were available¹⁴.

Below are some notes on the concerns enumerated in the letter.

- 1) Adjusted numbers will not be "the best available population data".
 - a. This is the language used in the *Karcher* case which was quoted by the SCOTUS in the Commerce Department opinion in 1999 about adjustment.
 - b. The basic concern here is that the both phases of the apportionment process, i.e., the apportionment of seats to predetermined units (e.g., states) and the districting phase should rely upon the best available data.
 - c. The Bureau has indicated that the state-level counts would be held invariant; a position that changed after initial discussions with stakeholders.
- 2) Stakeholders will not be able to "make good faith efforts" at equality.
 - a. This language also refers to the *Karcher* case which basically requires a zero-tolerance for population amongst congressional districts.
 - b. Also of note are the *Larios v. Cox* case (out of Georgia) in 2004¹⁵ and the *Tennant v. Jefferson County Commission* case (out of West Virginia) in 2012¹⁶. Larios reiterated the focus of the reapportionment cases of the 1960s that the goal (therein for legislative districting) was to have equally populated districts.

¹⁴ Note also that blocks are numbered by the Bureau and thus Block Groups, the next higher level above Blocks, are simply agglomerations of adjacent Blocks for statistical purposes. Census Tracts, the next level up the main hierarchy (aka the Spine) are designed to be generally consistent over time but have, on average thousands of persons.

¹⁵ See *Cox v. Larios*, 542 US 947 (2004); no. 03-1413, decided June 30, 2004;

¹⁶ See *Tennant v. Jefferson County Commission*, 567 US 758 (2012); no. 11-1184; decided September 25, 2012.

- c. The West Virginia case muddied this up a bit (for congressional districting) allowing some leniency for population deviation based upon the competing interests of the lowest deviation and legitimate state objectives. In reality this opinion reminded stakeholders of the original perspective of the Court in *Karcher*.
- 3) Use of such a statistical method “poses the risk of an inaccurate, invalid, and unconstitutional census”.
 - a. In its findings, the Congress was apparently referring to the competing analyses of the proposed adjustment for undercount which adjustment was to be based upon a statistical method known as sampling.
 - b. The Commerce case hinged largely on the statutory interpretation of the Census Act in sections 141 and 195 and held that the statistical method known as sampling was not an available method for the numbers compiled for the purposes of apportionment.
- 4) Additional litigation over the numbers may result in distraction, delay, and costs to many districting entities.
 - a. National entities are frequently at the forefront of litigation over these types of issues and bear the cost of having the courts reach a generally applicable ruling. However, given the range of error that might be infused into the process by noise injection it is likely that numerous cases may occur because of a dispute over how to interpret the altered numbers. The burden and confusion in such cases may redound to localities that may not be able to afford litigation through the entire process.
- 5) The confidence amongst state and local governmental entities in the entire census process may be severely undermined.
 - a. Local officials will review the census results block-by-block and when they discover that the reported results are different, and frequently substantially so, they will be concerned.
 - b. In recent censuses there has been a Count Question Resolution Program (CQR) to review the counts upon request and correct them if and as needed. It is unclear how this can be implemented if DAP is used for 2020.
- 6) While the Bureau is a national statistical agency, first and foremost it is the compiler of the “actual Enumeration” to fulfill the constitutional mandate.
 - a. There appears to be a break in the internal firewall at the Bureau vis-à-vis fulfillment of the constitutional mandate and ongoing survey programs. Admittedly, the number of survey programs that are done for other agencies and those that present the demographics of the nation to the world are the everyday projects for much of the Bureau. Understandably, what is good enough for a statistical agency to present may fall short of the standard of care for the counts used for “the purposes of apportionment”.
 - b. Of course, there are some projects that focus on the high quality of the actual enumeration at the Bureau and Complete Count Committees, as well as NGOs, work diligently throughout the decade to make the decennial “the best population data available”. Implementing DAP may lessen that focus because the numbers that will be used for redistricting will not be from the enumeration but altered in the manner proposed by the data scientists and decided by the Disclosure Review Board.
- 7) Previous methods for disclosure avoidance were less pervasive.
 - a. The previous methods were simpler, and easily understandable, techniques such as data swapping, rounding, top-coding, etc. and the degree to which all census

information was adjusted for protection was much less. On the other hand, the DAP for 2020 will affect almost every level of geography and the population counts.

- b. The DAP really is a 'sea change' for redistricting and the census. Users of the special tabulations have accepted previous efforts at disclosure avoidance because those users are cognizant of the problems and the shortcomings in protected data for their specific purpose, which would rarely require the precision needed for the purposes of apportionment.
- 8) Relative inaccuracy and bias in the DAP.
- a. "The new method allows us to precisely control the amount of uncertainty that we add according to privacy requirements." Not only will the data scientists determine the best method to adjust the counts but there will inevitably be some loss of accuracy which will have some level of bias for or against some subgroup of the census universe.
 - b. It is still unclear exactly what this bias will be at this point but what is likely is that once a bias is anticipated or observed the question of using the DAP will no longer be simply one of process but a political fight of the disfavored groups against the favored groups.

###