Northwestern | INSTITUTE FOR POLICY RESEARCH

Working Paper Series

WP-18-10

Balancing 2020 Census Cost and Accuracy: Consequences for Congressional Apportionment and Fund Allocations

Zachary Seeskin
Statistician
NORC at the University of Chicago

Bruce Spencer
Professor of Statistics
IPR Fellow
Northwestern University

Version: May 11, 2018

DRAFT

ABSTRACT

The researchers question how accurate the 2020 census needs to be, given that accuracy is expensive but inaccuracy distorts distributions of congressional seats and federal funds. Although the 2010 census had small measured errors for states, 0.6% on average (as measured by root-mean-square error, RMS), the researchers project that Texas loses and Minnesota gains a seat if the 2020 census has the same errors. Projections further show that if 2020 census error for state populations increases to 0.7% RMS, an additional seat is lost by Florida and gained by Ohio, and if error increases to 1.7% RMS, Texas loses a second seat, to the benefit of Rhode Island. The researchers find expected distortions in fund allocations increase about $9–$13 billion for each 0.5% increase in average error.

Correspondence to: bspencer@northwestern.edu.

* This work was supported by the US National Science Foundation (grant SES-1129475 to Northwestern University). We are grateful to Connie Citro, Mike Cohen, Jamie Druckman, Chuck Manski, Diane Schanzenbach, and John Thompson for comments. They should not be held responsible for the views expressed. Any errors are the responsibility of the authors alone.

Data and software code used in the analyses are at http://bit.ly/seeskin-spencer-data.

The U.S. Constitution requires that the population be enumerated decennially, for purposes of allocating Representatives among the states.

> Representatives shall be apportioned among the several States according to their respective numbers, counting the whole number of persons in each State, excluding Indians not taxed. The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct. (Art. I, Sec. 2, as amended)

The Constitution requires a census but does not say how accurate the census should be. Accuracy and cost are closely related. Perfect accuracy is unattainable at any cost. As demographer Nathan Keyfitz noted, “Asking why the census cannot [accurately] count 100 percent of the population in a free society is like asking why books contain typographical errors, why manufactured products often have defects, or why the police cannot catch all criminals.” (*1*, 46) Accuracy can be increased through investment of more resources in the census.

Understanding the cost-accuracy tradeoff is critically important for choosing and evaluating a census design. Associated with any design is a cost-accuracy curve (“cost curve”) that specifies the cost of attaining a given profile of accuracy. The cost curve is determined by census technology and social behavior, including the cooperation of the public with providing information requested. *Figure 1* shows an illustrative example of the cost curve. Empirical determination of the curve is challenging, and indeed is a reason for testing and development activities at the Census Bureau.

Our study analyzes the effects of alternative levels of 2020 census accuracy on apportionment of the House of Representatives and on allocation of billions of dollars of federal funds. We argue that paying attention to census cost alone, without concern for accuracy, leads to large and perhaps counter-intuitive shifts in allocations and apportionment.

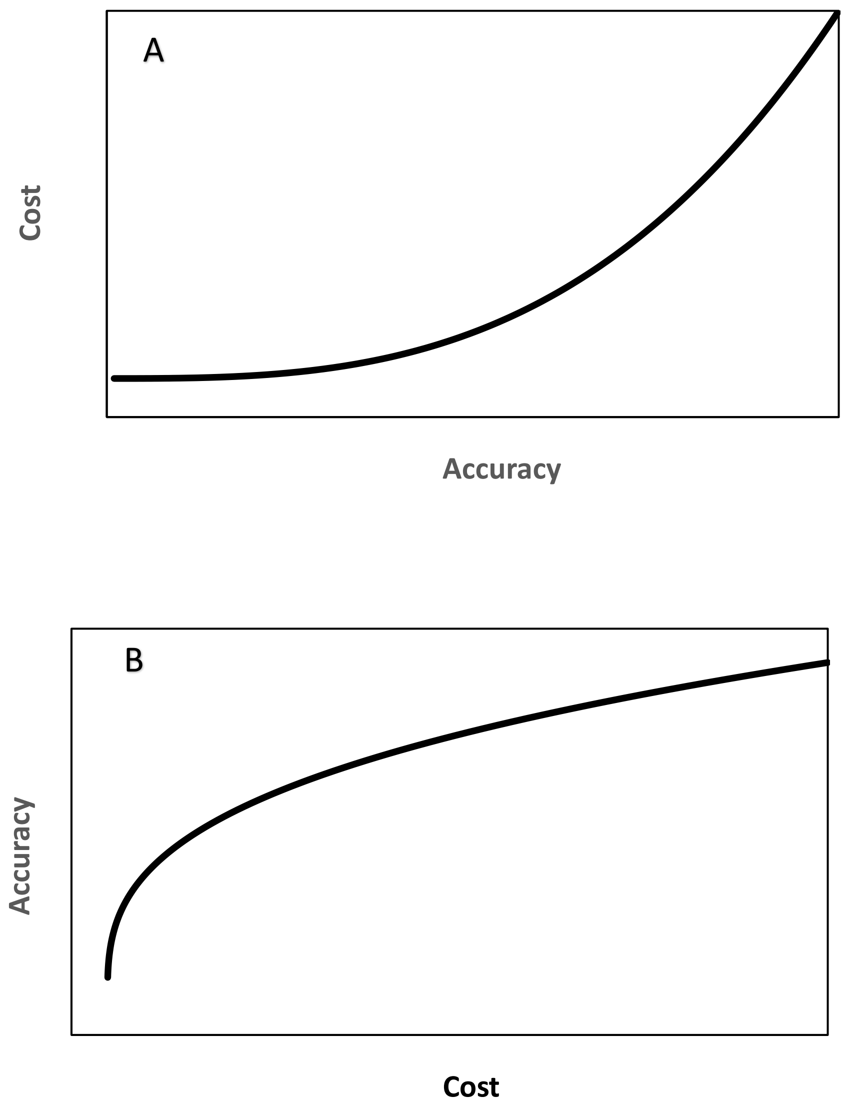

Figure 1. **The cost-accuracy curve shows the cost of attaining accuracy and the accuracy attainable at given cost. (A**) Accuracy typically is attained at increasing marginal cost and (**B**) additional spending yields decreasing returns in accuracy.

For at least the last five censuses, high accuracy was sought and spending was adjusted to try to attain it. This is evinced by the successful requests by the Census Bureau for additional funds in the years just prior to those censuses.

By contrast, for the 2020 census, Congress adopted a cost target instead of an accuracy target, and the Census Bureau is held responsible to achieve acceptable accuracy at that cost. The target was set so that the 2020 cost per housing unit remains at the same (inflation-adjusted) level as attained in 2010, or about $12.5 billion in 2020 dollars (*2*, Recommendation 3). This is almost 30% below the projected cost of repeating the 2010 census methods, and is attainable only with successful innovations, notably use of internet as the main venue for census reporting, use of modern geospatial imaging to update mailing addresses, use of mobile devices by census takers to collect data from households not completing a census form, and use of administrative data to remove vacant housing units and compensate for lack of data from non-respondents. Such innovations are still under development and require testing under realistic conditions (*3*, *4*).

The underfunding of requested census testing and development in the years leading up to the 2020 census demonstrates lack of concern for accuracy relative to cost (*4*, *5*). Indeed, although the accuracy attainable for that cost is uncertain at this point, the concerns outside the Census Bureau have focused almost exclusively on cost (*6*-*10*). The present dominating focus on cost leaves open the possibility that the accuracy attained by the census may be unsatisfactory for society's needs (just as a dominating focus on accuracy would run the risk of excessive spending to obtain inconsequential improvements in accuracy).

Statistical decision theory is a framework that jointly considers both costs and benefits of census accuracy and quantifies the tradeoff. This prevents excessive emphasis on either cost or accuracy. The benefits of the census arise from how its products are used. Reductions in census cost necessitate reductions in census accuracy, and reductions in accuracy lead to distortions in census uses. In certain situations, the benefit of a good can be reflected by its value in the market. However, the market does not properly value data, as data are a public good and will not be adequately provisioned by the free market (*11*). The most visible uses of the census results include intergovernmental allocation of funds by formulas using population statistics, apportionment of the U.S. House of Representatives, and redrawing of Congressional district boundaries. When the census population numbers contain errors, the fund allocations, Congressional apportionment, and district sizes are different from what they would be if the census numbers had no error.

Historically, census counts understated true size of population, and census error was quantified by net undercount rate, which equals the difference, true minus census, divided by true. Although the estimated net undercount rate for 1990 was 1.61%, the censuses in 2000 and 2010 were estimated to exceed true population size nationally, with net undercount rates estimated at –0.49% and –0.01%, respectively (12). For census uses that involve dividing a fixed total, including apportionment of the House of Representatives ("House") and programs that use statistical formulas to allocate fixed amounts of fund total among states, what matters are the states' differential undercount rates, defined as the net undercount rate for the state minus the rate for all states combined. Differential net undercount rates are defined analogously for demographic groups, with estimates shown in *Table 1*. The differential rates are fairly consistent across the three censuses, with non-Hispanic Whites overcounted relative to the nation as a whole, and Hispanics and non-Hispanic Blacks undercounted.

Inaccuracy in the census can distort the reapportionment of the House, where states can gain or lose a seat after only small changes in population (*1*). The distribution of House seats depends on the states' shares of population and is calculated by the "equal proportions" method (*13-16*). Projections of House reapportionment following the 2020 census can be calculated from projections of 2020 state population shares (*17*). To illustrate effects of census inaccuracy on apportionment, we modify the projections of 2020 state population by allowing for census errors.

Table 1. **Estimated differential net undercount rates for demographic groups in last 3 censuses.** Source: (*12*)

	Estimated Differential Net Undercount (%)		
Group	1990 Census	2000 Census	2010 Census
Non-Hispanic White	–0.9	–0.6	–0.8
Non-Hispanic Asian	0.8	–0.3	0.1
Hispanic	3.4	1.2	1.6
Non-Hispanic Black	3.0	2.3	2.1
Non-Hispanic native Hawaiian or other Pacific Islander	0.8	2.6	1.4
American Indian on reservation	10.6	–0.4	4.9
American Indian off reservation	n.a.	1.1	–1.9

Table 2 shows illustrative projections of winners and losers under three alternative levels of census error. The first column shows the effect on apportionment if errors in 2020 census state population shares equal errors measured for the 2010 census (*18*) – Texas loses a House seat to Minnesota. The last two columns show shifts in House seats if the patterns of error in the 2020 census resemble those measured for states in the 2010 census, but the overall error in population shares is exaggerated in 2020 due to underfunding. If the sizes of errors in 2020 are 20% larger than for 2010 (RMS size 0.71 versus 0.59), Florida also loses a seat and Ohio gains one; if the

RMS sizes of the errors in 2020 is 1.67, Texas is projected to lose a second seat, to the benefit of Rhode Island. In relying on 2010 census error estimates, these projections may be conservative due to changing demographics. For example, Hispanics comprise a larger proportion of Florida's population now than in 2010, and Hispanics tend to be undercounted relative to non-Hispanic Whites.

Table 2. **Projected gains and losses of House seats at different levels of 2020 census error.**

	RMS[1] relative error in state 2020 population shares		
State	0.59[2]	0.71	1.67
Florida	–	lose 1	lose 1
Minnesota	gain 1	gain 1	gain 1
Ohio	–	gain 1	gain 1
Rhode Island	–	–	gain 1
Texas	lose 1	lose 1	lose 2
Every other state	–	–	–
Seats shifted	2	4	6

– indicates no change. [1] RMS relative error is root-mean-square relative error. [2] The measured errors for states the 2010 census had RMS size 0.59.

As indicated in *Figure 2*, the expected number of changes in House seats due to error in the 2020 census tends to increase by about 2.5 – 3.5 when the root-mean-square (RMS) size of state errors increases by 1%. The RMS size of state errors is the square root of the mean of the states' squared undercount rates; columns 1, 2, and 3 in *Table 2* correspond to 2020 census error RMS sizes of 0.59%, 0.71%, and 1.67% respectively. We considered a variety of parametric error models, including state undercount rates multivariate normally distributed with zero mean, equal variance, and constant correlation, as well as other models (*20*). The right-hand axis of *Figure 2* shows the expected number of shifts in House seats for the models with correlation 0 and 0.5 as well as the error distributions used in *Table 1*, which were patterned on the measured errors for the 2010 census. When the errors are random, the actual number of malapportioned seats can be less than or appreciably greater than the expected number; e.g., in the model with uncorrelated errors, the actual number of malapportioned seats has about a 1 in 7 chance of being at least 20 with RMSE at 4%, at least 16 at 3%, and at least 10 at 2% (*20*).

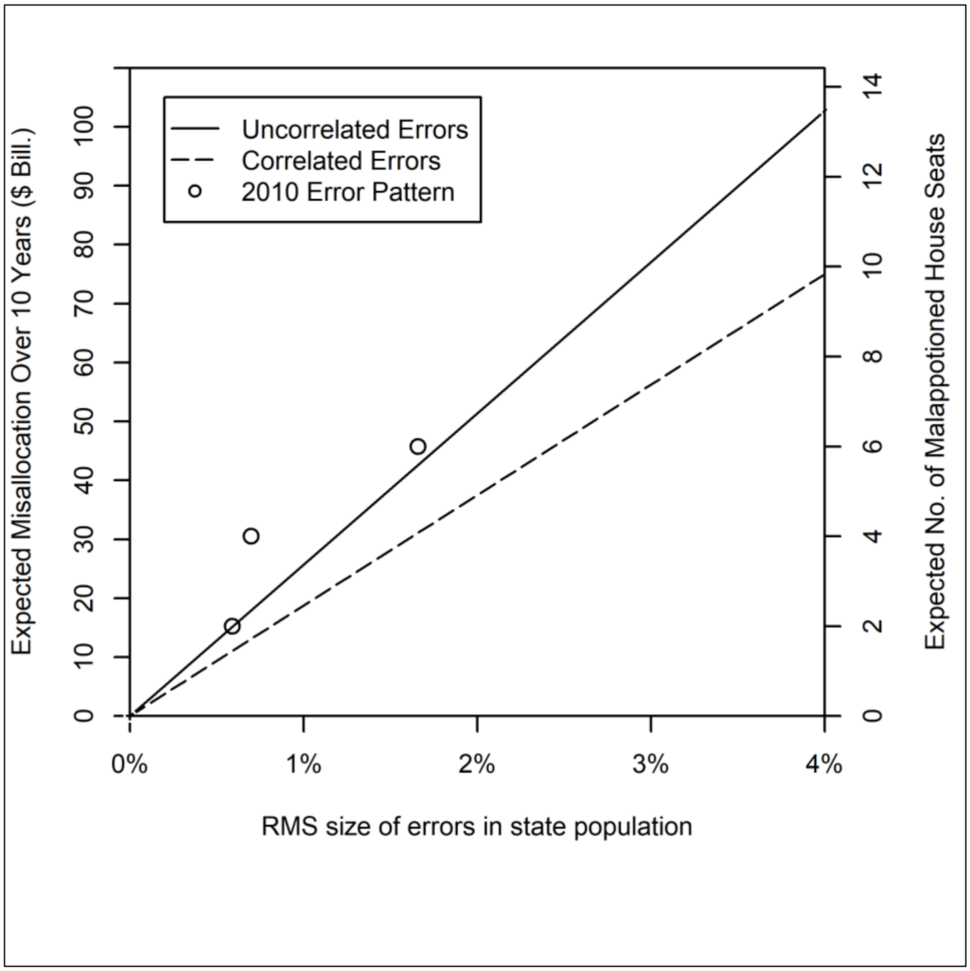

Figure 2. **Expected funds misallocations and malapportioned House seats.** (FY2015 dollars**)**

Census data affect the distribution of many billions of dollars of funds – more than $675 billion in allocations from 132 programs in FY 2015 according to a recent Census Bureau study (*21*). In fact, the cost-benefit analyses that have been carried out to date have focused on uses of census data for allocation of funds (*22*-*27*). With so many programs, it is not feasible to study the effects of census error on each program, and we selected a disproportionate stratified sample of 18 programs that accounted for 80% of the total obligations in FY 2007 (*28*). Sample weighting estimates were used to obtain unbiased estimates reflecting all allocation programs listed in both (*21*) and (*28*), and sampling variances were relatively small (c.v. < 4%). The expected amount of misallocated funds due to census error (if the same programs are in place at the same funding level for the decade following the 2020 census) is estimated at $80 billion for the decade if the RMS size of the census errors is as large as 4%. As seen in *Figure 2* (left hand axis), the expected amount misallocated increases linearly with the RMS size. Actual misallocations can be higher or lower than expected amounts.

Apportionment and allocations of funds, along with redistricting following each census, are highly visible uses of census data, but they are not the only important uses. It is noteworthy that some of the most important uses of the census may be the least visible, including research in social, economic, behavioral, medical, and policy areas and applications of that research. The role of census data in policy development and decision-making by the Congress and the White House, by state and local governments, and by businesses and other organizations has not received sufficient study. For example, surveys are widely used sources of information, and almost all national population surveys – whether government or private sector, whether by internet, mail, phone, or in-person – directly or indirectly use decennial census numbers for adjusting their results. Public health impacts of census error are discussed in (*29*).

In conclusion, inaccuracy in the 2020 census can cause quite large – and counterintuitive – distortions in distributions of federal funds to states and local governments. If the average root-mean-square error of state populations is 2%, the expected shifts in fund allocations is on the order of $40 – $50 billion over ten years and the expected shifts in House apportionment is around 6 seats; if the average RMS error is as large as 4%, the expected shifts double in size. The actual shifts could be smaller or even greater than the expected values. We hope the average error is much smaller than 2% or 4%, as appears to be the case for previous censuses (*30*), but the reality will strongly depend on the level of census funding.

References

1. N. Keyfitz, Information and allocation: two uses of the 1980 census. *Am. Stat.* **33**, 45-50 (1979).
2. National Research Council, *Change and the 2020 Census: Not Whether But How* (National Academy of Sciences, Washington, DC, 2011).
3. U.S. Census Bureau, *2020 Census Operational Plan: A New Design for the 21st Century, Version 2.0* (Department of Commerce, Washington, DC, 2016).
4. B. Tarran, Interview with John Thompson. *Significance* **14**(4), 6-7 (2017).
5. T. Bahrampour. Census watchers warn of a crisis if funding for 2020 count is not increased. *The Washington Post* (April 18, 2017; http://wapo.st/2pLLTQc?tid=ss_mail&utm_term=.ca49246f1f93).
6. U.S. Government Accountability Office, (GAO), *High-risk series: key actions to make progress addressing high-risk issues* (Report GAO-16-480R, GAO, 2017; gao.gov/assets/680/676800.pdf).
7. U.S. Government Accountability Office, (GAO), *2020 census: Census Bureau needs to improve its life-cycle cost estimating process.* (Report GAO-16-628, GAO, 2016; gao.gov/assets/680/678117.pdf).
8. U.S. Government Accountability Office, (GAO), *2020 census: additional actions would help the bureau realize potential administrative records cost savings.* (Report GAO-16-48, GAO, 2015; gao.gov/assets/680/673204.pdf).
9. U.S. Government Accountability Office, (GAO), *2020 census: additional steps are needed to build on early planning*. (Report GAO-12-626, GAO, 2016; gao.gov/assets/600/590939.pdf).
10. U.S. Government Accountability Office, (GAO), *2020 census: sustaining current reform*

efforts will be key to a more cost-effective enumeration. (Report GAO-12-905T, GAO, 2012; gao.gov/assets/600/592601.pdf).

11. C. R. Sims, Can we measure the benefits of data programs? *Proc. Am. Stat. Assoc., Soc. Stat. Sec.*, 60-67 (1984).
12. H. Hogan, P. J. Cantwell, J. Devine, V. T. Mule Jr, V. Velkoff, Quality and the 2010 census. *Popul. Res. Policy. Rev.* **32**, 637–662 (2013).
13. M. L. Balinski, H. P. Young, *Fair Representation: Meeting the Ideal of One Man, One Vote*. (Yale University Press, New Haven, CT, 1982).
14. Census Bureau, https://www.census.gov/population/apportionment/about/computing.html (2013).
15. L. Gilford, B. Causey, "The effect of undercount adjustment on the census." (Census Bureau, 1981).
16. B. D. Spencer, Statistical aspects of equitable apportionment. *J. Am. Stat. Assoc.* **80**, 815-822 (1985).
17. K. W. Brace, Some change in apportionment allocations with new 2017 census estimates; but greater change likely by 2020 (Election Data Services, Manassas, VA, Dec. 26, 2017; electiondataservices.com/wp-content/uploads/2017/12/NR_Appor17c3wTablesMapsC2.pdf).
18. T. Mule, "2010 Census coverage measurement estimation report: summary of estimates of coverage for persons in the United States" (DSSD 2010 census coverage measurement memorandum series #2010-G-01, Census Bureau, 2012; census.gov/coverage_measurement/pdfs/g01.pdf).
19. Z. Seeskin, "*Topics on Official Statistics and Statistical Policy,"* thesis, Northwestern University (2016; https://search.proquest.com/docview/1826016819).

20. Supporting Information (appended to this document).

21. M. Hotchkiss, J. Phelan, *Uses of Census Bureau data in federal funds distribution, version 1.0.* (Census Bureau, Washington, DC, 2017; https://www2.census.gov/programs-surveys/decennial/2020/program-management/working-papers/Uses-of-Census-Bureau-Data-in-Federal-Funds-Distribution.pdf)

22. P. Redfern, The different roles of population censuses and interview surveys, particularly in the UK context. *Int. Stat. Rev.* **42**(2), 131-146 (1974).

23. B. D. Spencer, *Benefit-Cost Analysis of Data Used to Allocate Funds: General-Revenue Sharing.* (Springer, New York, 1980).

24. J. Aldrige, 2011 census business case (General Register Office for Scotland, 2006; whatdotheyknow.com/request/8345/response/20302/attach/3/business%20case.pdf).

25. Merits of Statutory Instruments Committee, Draft census (England and Wales) order 2009 etc. (Parliament of the United Kingdom, 2009; publications.parliament.uk/pa/ld200809/ldselect/ldmerit/176/17606.htm).

26. C. Bakker, Valuing the census (Statistics New Zealand, Wellington, 2014; archive.stats.govt.nz/methods/research-papers/topss/valuing-census.aspx).

27. B. D. Spencer, J. May, S. Kenyon, Z. Seeskin, Cost-benefit analysis for a quinquennial census: the 2016 population census of South Africa. *J. Off. Stat.* **33**, 249-274 (2017).

28. L. M. Blumerman, P. M. Vidal, P. M., *Uses of population and income statistics in federal funds distribution – with a focus on Census Bureau data* (Governments Division Report Series, Research Report #2009-1, Census Bureau, Washington, DC 2009; https://www.census.gov/prod/2009pubs/govsrr2009-1.pdf).

29. R. T. Wilson, S. H. Hasanali, M. Sheikh, S. Cramer, G. Weinberg, A. Firth, S. H. Weiss, C.

L. Soskolne, Challenges to the census: international trends and a need to consider public health benefits. *Public Health* **151**, 87-97 (2017).

30. National Research Council, *Envisioning the 2020 census*. (National Academy of Sciences, Washington, DC, 2010).
31. Data and software code are available for download from http://bit.ly/seeskin-spencer-data.
32. Census Bureau, Annual estimates of the resident population for the United States, regions, states, and Puerto Rico: April 1, 2010 to July 1, 2017 (NST-EST2017-01) (Population Division, Census Bureau, December 2017; https://www2.census.gov/programs-surveys/popest/tables/2010-2017/state/totals/nst-est2017-01.xlsx).
33. Census Bureau, Methodology for the United States population estimates: vintage 2017 nation, states, counties, and Puerto Rico – April 1, 2010 to July 1, 2017 (2017; https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2010-2017/2017-natstcopr-meth.pdf).
34. B. D. Spencer, Models for error in postcensal population estimates. Pp. 217-228 in National Research Council, *Estimating population and income of small areas.* (National Academy of Sciences, Washington, DC, 1980).
35. Census Bureau, Congressional apportionment (2010 Census Brief, 2011; https://www.census.gov/prod/cen2010/briefs/c2010br-08.pdf).
36. Census Bureau, Table 2-rev, comparison of population estimates and census counts for the United States, regions, states, and Puerto Rico: April 1, 2010 (March 2012; https://www2.census.gov/programs-surveys/popest/datasets/2010/2010-eval-estimates/table-2-rev.xls).

37. Weldon Cooper Center for Public Service, *National and State Population Projections* (Demographic Research Group, University of Virginia, Charlottesville, 2013).

38. Social Security Administration, Table, CPI for Urban Wage Earners and Clerical Workers (2018; https://www.ssa.gov/oact/STATS/cpiw.html).

39. Catalog of Federal Domestic Assistance (CFDA) (Office of Management and Budget, Washington, DC, 2015; https://cfda.symplicity.com).

40. United States Code Office of the Law Revision Council of the U.S. House of Representatives, Washington, DC; http://law.justia.com/codes/us).

41. National Research Council, *Statistical issues in allocating funds by formula*. (National Academy of Sciences, Washington, DC, 2003).

Supporting Information

S1. Introduction

This material provides additional details about estimates of the distributions of distortions in allocations of representation and funding among states that arise at alternative profiles of accuracy in the 2020 census. The apportionment algorithm as well as funding formulas and total funding amounts as of FY 2007 are treated as fixed. Allocations (of funds or representation) that would occur with error-free statistics are treated as *true values* for the allocations, in contrast to *empirical* or *estimated* allocations based on inaccurate statistics. The difference, estimated minus true allocation, is the *error in allocation* or, more simply, the *misallocation*; the absolute value of the difference is called the *absolute misallocation*. Discussion and results for measures of discrepancy other than sum (across states) of absolute values, including sum of square errors, mean absolute percentage error, maximum absolute error, and maximum absolute percentage error are in (*19*). The term "error" is standard usage in statistics and does not imply that someone made a mistake. Relative error is defined as the error divided by the quantity being estimated.

The calculations of errors in apportionment and in fund allocation involve joint specification of the true population and the census population numbers for states, or equivalently the true population numbers and the census errors. (For fund allocations, we include Washington D.C. as a state.) Different specifications were used for errors in 2020 apportionment in *Table 2* and in *Figure 2*, and for errors in fund allocation in *Figure 2*. (Note: Tables and figures are identified as *Table 1*, *Figure 1*, etc. when they appear in the main text and as *Table S1*, *Figure S2*, etc. when they appear in this Supplementary Information.) The following material discusses the methods and data for the results in the main text, and provides supplementary results. For additional data and software code, see (*31*).

The organization of the Supplementary Information is as follows. Methods, data, and results are discussed in Sections S2–S4 for apportionment and in Section S5–S6 for fund allocations.

i. Section S2 discusses the data and models used to project individual states' errors in apportionment, as shown in *Table 2.* The true 2020 population numbers were projected by short-term linear extrapolation of postcensal estimates from 2017, and 2020 census errors were modeled by scaling the measured errors in the 2010 census (*18*).

ii. Section S3 discusses an alternate specification for true 2020 population numbers and census errors, which was used for errors in apportionment reported in *Figure 2.* The vector of 2020 true state population sizes was considered to be random, with mean vector equal to state population projections based on the 2010 census and constant relative variances based on empirical differences between 2010 census numbers and projections for April 1, 2010 (*19*). A variety of alternative parametric models were developed for 2020 census errors conditional on the true 2020 population.

iii. Section S4 provides supplementary results.

iv. For errors in fund allocation as displayed in *Figure 2*, we used a different approach, which is discussed in Section S5. Unlike apportionment, which depends only on state population sizes in 2020, formula-based allocations of funds depend on a wide variety of population statistics and other statistics. Rather than jointly forecast the values of all such statistics ahead to 2020, which would involve complexity and uncertainty of forecasts, we obtained the latest values available of the statistics used to calculate allocations for the 18 programs studied, and we treated those as error-free. Thus, the true state population numbers used in our analysis of allocation of funds are based directly or indirectly on the 2010 census, but not on projections or forecasts of the 2020 population sizes.

S2. Projected Gains and Losses of House Seats for Individual States Shown in *Table 2*

First, we created a projection of the state population sizes for apportionment after the 2020 census. Second, we adjusted the projections accounting for three alternative levels of 2020 census error. Third, we compared the apportionments based on the populations regarding the projections as true with the apportionments based on populations incorporating alternative error specifications for 2020 census error.

S2.1. Projection of true 2020 apportionment population sizes

The projection of 2020 apportionment populations is developed in two steps. The first step took the Census Bureau's postcensal estimates x for 7/1/2016 and y for 7/1/2017 and linearly extrapolated (projected) forward 33 months (2.75 years) to 4/1/2020 as $z = y + 2.75(y - x)$. The Census Bureau develops postcensal estimates by accounting for change since the previous census due to births, deaths, and net movement in and out of the state. The Census Bureau's estimates are available in (*32*) and the underlying methodology is described in (*33*). Although undercount in the prior census does affect postcensal estimates (*34*), for the purposes of this analysis we are not modifying the projections to account for undercount, as such modification would be both complex and uncertain.

The second step involved modifying the projection, z, for differences between the census population and the apportionment population. The modification for state i involves multiplication of the projected population z_i by the ratio r_i of the 2010 apportionment population (*35*) to the 2010 census population (*36*). The projected 2020 true apportionment population size for state i is $v_i^{2020} = r_i z_i$. Denote the sum of v_i^{2020} across the 50 states by v_+^{2020}.

S2.2. State-level differential undercount in 2010 apportionment populations

Three steps were followed to use the estimated net undercount rates for the 50 states in the 2010 census to calculate differential net undercount rates for the states. First, we calculate undercount-adjusted population sizes. Second, we use those to calculate the undercount rate for all 50 states combined. Finally, we calculate the differential net undercount rate.

i. For state i, denote the undercount rate in the 2010 census by u_i, the 2010 census apportionment population size by v_i^{2010} and the true 2010 apportionment population size by t_i^{2010}. We assume state apportionment populations have the same undercount rates as the state census populations. This implies $u_i = \left(t_i^{2010} - v_i^{2010}\right) / t_i^{2010}$, or $t_i^{2010} = v_i^{2010} / \left(1 - u_i\right)$.

ii. Denote the sum of v_i^{2010} and t_i^{2010} across the 50 states by v_+^{2010} and t_+^{2010}, respectively, and define $u_+ = \left(t_+^{2010} - v_+^{2010}\right) / t_+^{2010}$. We may rewrite this as $u_+ = \sum_i t_i^{2010} u_i / \sum_i t_i^{2010}$.

iii. The differential undercount rate for state i is defined as $d_i = u_i - u_+$. The differential undercount is a linear approximation to the relative error of the state i share of the apportionment population. Substituting the estimated undercount rate $\hat{u}_i$ and $\hat{u}_+$ (*18*) for u_i and u_+, we estimate the 2010 differential undercount rate for state i by $\hat{d}_i = \hat{u}_i - \hat{u}_+$.

S2.3. Modeling state-level undercount in the 2020 census from measured 2010 undercount

The projected 2020 apportionment population size v_i^{2020} of state i is adjusted for illustrative profiles of net undercount in the 2020 census. To do this, we introduce a multiplier λ to apply to the differential undercount as in the 2010 census. This leads to projected apportionment enumerations a_i^{2020} for 2020. The formula for this is $a_i^{2010} = v_i^{2020}\left(1-\hat{u}_+ - \lambda\hat{d}_i\right)$. One can interpret $\lambda > 1$ as less accuracy (larger state differential undercounts) than 2010, $\lambda = 1$ as the same accuracy as 2010, and $0 \le \lambda < 1$ as more accuracy. If $\lambda = 0$ then there is no error in the state i share of apportionment population, $a_i^{2020} / a_+^{2020} = v_i^{2020} / v_+^{2020}$. It may be noted that $\hat{u}_+$ was so close to zero, at -0.017%, that similar results are found if λ is applied to undercount rather than differential undercount.

v. Notice that the RMS sizes of the differential undercount also scale by λ. Choices of λ equal to 1, 1.2, and 2.385 correspond to RMS sizes of 0.59, 0.71, and 1.67 as shown in *Table 2*. Finally, the apportionments are then calculated using the Equal Proportions apportionment method with the a_i^{2020} values as the population sizes of the states. For 1 $\le \lambda$ < 1.2 there were 2 House seats misallocated, for 1.2 $\le \lambda$ < 2.385 there were 4 seats misallocated, and for λ = 2.385 there were 6 seats misallocated; see *Table 2*.

S3. Joint Distribution of 2020 Population True Values and Estimates as Applied to Apportionment in *Figure 2*

S3.1. Probability distribution for 2020 population true values

For analysis of apportionment as reported in *Figure 2,* true population sizes of states were taken to be multivariate normal with means equal to projections for 2020. We used projections made by the University of Virginia's Weldon Cooper Center for Public Service based on the 2010 census results (*37*) because the Census Bureau stopped producing state projections. We chose a diagonal covariance matrix with variances consistent with empirical errors in past ten-year projections for 2010, as discussed below. Apportionments are integers, and it is theoretically possible that a change in population of 1 person can cause a state to gain or lose a seat (*1*). Specifying a variance for the true values prevents our estimates from being sensitive to true population sizes being necessarily near or far from values that would change apportionments. Simulations showed the variance around the means to have little if any effect on the estimates of malapportionment arising from census inaccuracy. No adjustment was made for differences between state population and state apportionment population. The numerical values are shown in *Table S1.*

***Table S1.* Specification of moments of state populations in 2020**.

State	Mean	Coeff. of Variation (%)	State	Mean	Coeff. of Variation (%)
Alabama	5,066,866	2.6	Montana	1,055,292	4.1
Alaska	811,718	4.4	Nebraska	1,908,775	3.6
Arizona	7,604,382	2.5	Nevada	3,328,548	3.1
Arkansas	3,120,724	3.1	New Hampshire	1,446,097	3.9
California	41,715,522	2.0	New Jersey	9,252,696	2.3
Colorado	5,733,049	2.5	New Mexico	2,307,561	3.5
Connecticut	3,723,612	2.8	New York	19,952,674	1.8
Delaware	997,528	4.2	N. Carolina	10,736,114	2.3
Florida	21,784,582	1.8	North Dakota	678,125	4.5
Georgia	11,078,010	2.3	Ohio	11,763,865	2.2
Hawaii	1,489,774	3.8	Oklahoma	3,986,956	2.7
Idaho	1,772,613	3.7	Oregon	4,223,601	2.7
Illinois	13,277,307	2.1	Pennsylvania	12,961,019	2.1
Indiana	6,804,046	2.5	Rhode Island	1,085,957	4.1
Iowa	3,085,572	3.0	S. Carolina	5,118,310	2.7
Kansas	3,011,419	3.1	South Dakota	853,943	4.3
Kentucky	4,558,229	2.7	Tennessee	6,919,966	2.5
Louisiana	4,635,071	2.7	Texas	28,738,112	1.8
Maine	1,394,018	3.9	Utah	3,193,030	3.1
Maryland	6,282,303	2.5	Vermont	662,770	4.7
Massachusetts	6,806,874	2.3	Virginia	8,871,484	2.3
Michigan	10,074,617	2.2	Washington	7,576,478	2.3
Minnesota	5,704,065	2.5	W. Virginia	1,817,852	3.5
Mississippi	3,111,177	3.2	Wisconsin	6,004,398	2.5
Missouri	6,336,145	2.5	Wyoming	594,027	4.8

The variances of the 2020 population sizes were specified to be consistent with the observed levels of error in state population projections prepared a decade earlier by the Census Bureau. Specifically, in 2005 the Census Bureau used 2000 census results to project state populations for July 1, 2010. The error in those projections was estimated by the difference between the projection, Y, and the Census Bureau's population estimates for July 1, 2010, X, which are equal to the 2010 census enumeration adjusted for births, deaths, and net migration over the 3 month interval from April 1 to July 1. The relative error was computed as the error $Y - X$ divided by X, or equivalently $Y / X - 1$. The relative errors were observed to be approximately normally distributed about zero, and the relative errors tended to be closer to zero for the larger states than the smaller states. To model the squared relative error as a function of the true population size, a lowess fit of $(Y / X - 1)^2$ against X was conducted in Stata® 11 using a bandwidth of 0.8 and preserving the mean. The lowess fitted values were used as estimates of both the relative variances of the population projections for 2020 and the relative variances of the future 2020 state population sizes.

The assumption of independence for the distribution of true population sizes of states was motivated by the following considerations. State population projections typically are controlled to sum to national forecasts, which account for births, deaths, and net immigration since the last census. The latter likely induce a source of positive covariance among state population projections (if the projections are treated as random variables). However, the dominant source of error in forecasts of 10 years or shorter will be uncertainty about interstate migration. Since the interstate migration flows must sum to zero, the covariances cannot all be positive, but will have a more complex pattern. For simplicity, the 2020 population sizes are taken to be independent, knowing that only the population shares matter for apportionment, and that the shares implicitly include some negative covariances because the sum of shares is always 1.

S3.2. Conditional distribution of 2020 census errors given true population sizes

S3.2.1. Uncorrelated errors model and correlated errors model

Various parametric error models were examined to explore the sensitivity of findings to alternative error distributions. Two such models were used to construct *Figure 2*. Both models assume relative errors had a multivariate normal distribution with zero mean, common standard deviation, and constant correlation. In the *uncorrelated errors* model the correlation was zero, and in the *correlated errors* model, the common correlation was 0.5.

S3.2.2. Differential bias model and accurate small states model

Two additional error models are the differential bias model and the accurate small states model. The *differential bias* model is like the uncorrelated errors model except that biases are present, with one sign for the 25 most populous states ("large states"), and opposite sign for all others including Washington, D.C. ("small states"), and equal magnitudes of relative biases for all states; relative standard deviations of errors for all states were equal to each other and to the absolute value of the relative biases. The *accurate small states* model is like the uncorrelated errors model except that errors for small states were identically zero (zero means and standard deviations). For each of these models, specification of the average root-mean-square-error (RMSE) was sufficient to completely specify the model.

S3.2.3. More general error models

We also considered more general models. In these models, each state's relative error was assumed to be distributed as a linear function of a Student's *t* random variable with the same degrees of freedom. The error distributions were characterized by six parameters: ρ the common correlation of the errors for each pair of states, σ_L the common standard deviation of the relative error for large states, σ_S the common standard deviation of the relative error for small states, μ_L the common mean of the relative error for large states, and μ_S the common mean of the relative error for small states, and degrees of freedom, δ. The square of the RMSE of any state's relative error equals $\sigma^2 + \mu^2$, and so the average RMSE can be derived directly. Similar to the previously discussed error models, specification of the average root-mean-square-error (RMSE) was sufficient to completely specify the model.

S3.3. Simulating from the joint distribution

To conduct simulations, we first selected a vector of population sizes from the distribution described in Section S3.1 and then selected a vector of relative errors from the distribution described in Section S3.2. This joint selection specifies a pair consisting of the true population vector and the vector of errors. For each pair, House apportionment by the Equal Proportions method was computed twice, once for the true populations and once for the population numbers incorporating the errors, and the differences in apportionment for each state were recorded. The process was repeated, independently, 5,000 times.

S4. Number of Malapportioned Seats in the House of Representatives under Alternative of 2020 Census Error Models

S4.1. Expected number of malapportioned seats

Figure S1 and *Table S2* show the expected number of malapportioned House seats under the alternative joint distributions of population and census error presented in Sections S3.1–S3.2.2. The numbers are derived from the simulations described in Section S3.3. Standard errors for all estimates of malapportionment in *Figure S1* and *Table S2* are less than 0.05 House seats (*19*, 37).

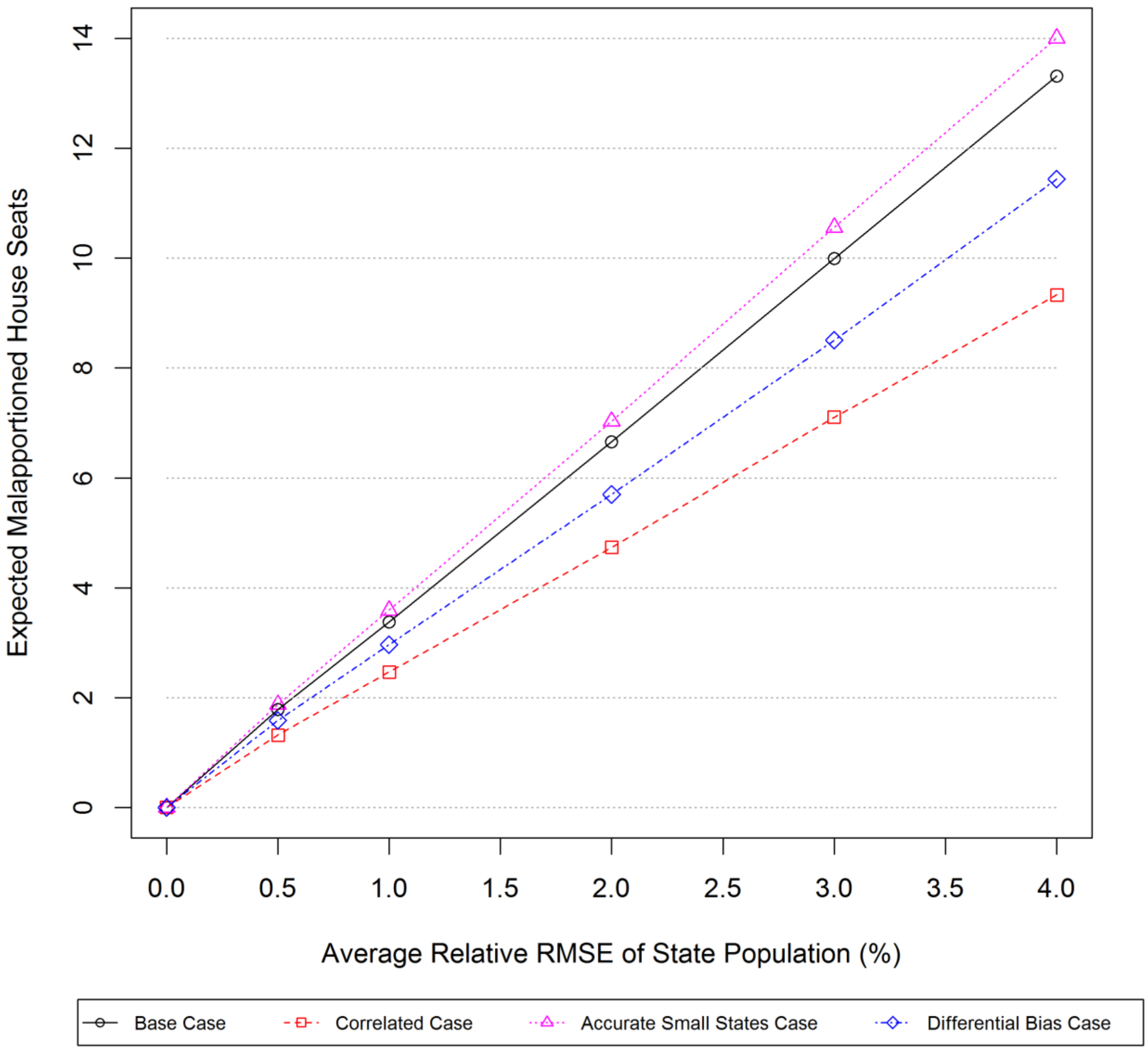

Figure S1. **Estimated expected number of malapportioned seats under alternative 2020 census error distributions.**

***Table S2*. Estimated expected number of malapportioned seats in the U.S. House, with alternative error models.**

Expected Number of Malapportioned House Seats					
	Average Relative RMSE of State Population Numbers				
Error model	0.5%	1.0%	2.0%	3.0%	4.0%
Uncorrelated Errors	1.79	3.38	6.66	10.00	13.32
Correlated Errors ($\rho = 0.5$)	1.32	2.46	4.74	7.11	9.33
Accurate Small States	1.88	3.59	7.03	10.56	14.01
Differential Bias	1.59	2.96	5.70	8.51	11.44

Estimated standard errors for all numbers do not exceed 0.05.

Estimates of the expected number of malapportioned House seats under the more general census error models of Section S3.2.3 can be readily computed using linear regression models that we fitted. The coefficients of the equations are shown in the first row of *Table S3*. To obtain the coefficients, we fitted the regression models to simulation-based estimates of sums of expected absolute deviations for 973 different possible combinations of the six parameters defined in Section S3.2.3: δ ranging between 4 and 60, ρ between 0.0 and 0.8, σ_L and σ_S between 0.2% and 5.0%, and μ_L and μ_S between -3.0% and +3.0%. For each combination of parameters, the sum of expected absolute deviations was estimated by the average of the sum of absolute errors over 2,500 simulations. To avoid extrapolation outside the range of the parameter values used to fit the regression, the regression models should only be used to approximate expected absolute loss within the above ranges of parameters. If one wishes to study normally distributed census statistics, using $\delta = 60$ is recommended. For the regression fit, $R^2 = 0.986$.

The nominal *p*-values (assuming normality) for all regression coefficients were below 0.001. Further details are in (*19*).

Table S3. **Coefficients for linear regression predictions of expected numbers of malapportioned House seats and sums of misallocated funds ($ bill.).**

	const.	δ'	ρ	ρ^2	σ'_L	σ'_S	μ'_L	μ'^2_L	μ'_S	μ'^2_S	$\rho\sigma'_L$	$\mu'_L\mu'_S$
Seats	3.457	.071	-18.767	-4.302	2.688	.295	-7.413	10.587	7.083	10.293	-1.942	-.169
Funds	2.858	.059	-13.145	-3.239	1.940	.309	2.209	11.643	-2.240	10.026	-1.287	-.135

Note: $\delta' = (\delta - 60)/10;\ \sigma' = 100\sigma - 1;\ \mu' = 100\mu.$ Regressor values should be used only in the following ranges: $.0 \le \rho \le .8;\ .2 \le \sigma' \le 5;\ -3 \le \mu' \le 3;\ 4 \le \delta \le 60,$ with 60 used for normal distribution.

The following results are implied by the regression model.

(a) A census error distribution with greater kurtosis than normal ($4 \le \delta \le 60$) leads to smaller absolute errors for constant variance. With each increase of 10 degrees of freedom, malapportioned House seats increase on average by 0.07.

(b) The predicted sums of absolute errors are sensitive to the constant correlation ρ between state census number relative errors, decreasing by about by 1.60 House seats as ρ increases from 0 to 0.8.

(c) The sum of expected absolute errors I apportionment is sensitive to the coefficient of variation of the state population numbers, increasing by about 2.7 with each 1.0% increase in the c.v. for large states and by 0.3 with each 1.0% increase in the c.v. for small states.

(d) The effect of the coefficient of variation for state census numbers on expected sums of absolute errors decreases as the constant correlation between the state census relative errors increases. For each 0.1 increase in the correlation, the effect of a 1.0% change in the coefficient of variation for large states decreases by 0.19 for House seats. Although negative correlations are possible, which would increase the effect of coefficient of variation, the negative correlations cannot be too large in magnitude because the correlation matrix is non-negative definite. For example, the minimum possible constant correlation for the census numbers of the 50 states and D.C. is -0.02.

(e) The sum of expected absolute errors in apportionment is sensitive to the relative biases of state census numbers, although less than to the coefficient of variation. As μ_L and μ_S vary between -3.0% and +3.0%, expected House malapportionment varies by about 1.5 House seats up or down. The relationship is convex, reflecting increased malapportionment with the magnitude of census bias.

S4.2. Probability distributions of number of malapportioned seats

The number of malapportioned seats is random and can be much greater than the expected number. *Table S4–Table S7* display the estimated probability distributions for the number of malapportioned seats under the alternative error models and alternative levels of relative RMSE of census numbers for states.

Table S4. Estimated probability distribution of number of House seats misallocated, uncorrelated errors accuracy profile.

Relative RMSE of census numbers	Probability that number of misallocated seats equals or exceeds *k*					
	k = 2	*k* = 4	*k* = 6	*k* = 8	*k* = 10	*k* = 12
0.5%	.714 (.006)	.166 (.005)	.013 (.002)	.000 (.000)	-- (--)	-- (--)
1.0%	.934 (.004)	.561 (.007)	.168 (.005)	.025 (.002)	.002 (.001)	-- (--)
2.0%	.998 (.001)	.956 (.003)	.761 (.006)	.428 (.007)	.147 (.005)	.035 (.003)
3.0%	1.000 (--)	.998 (.001)	.975 (.002)	.867 (.005)	.640 (.007)	.340 (.007)
4.0%	1.000 (--)	1.000 (--)	.998 (.001)	.983 (.002)	.914 (.004)	.752 (.006)

Relative RMSE of census numbers	Probability that number of misallocated seats equals or exceeds *k*					
	k = 14	*k* = 16	*k* = 18	*k* = 20	*k* = 22	*k* = 24
0.5%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
1.0%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
2.0%	.004 (.001)	.001 (.000)	-- (--)	-- (--)	-- (--)	-- (--)
3.0%	.132 (.005)	.037 (.003)	.009 (.001)	.001 (.001)	-- (--)	-- (--)
4.0%	.520 (.007)	.296 (.006)	.133 (.005)	.048 (.003)	.012 (.002)	.002 (.001)

-- signifies number < 0.02%. Number in parentheses is estimated standard error of probability.

***Table S5.* Probability distribution of number of House seats misallocated, correlated errors accuracy profile.**

Relative RMSE of census numbers	Probability that number of misallocated seats equals or exceeds *k*					
	k = 2	*k* = 4	*k* = 6	*k* = 8	*k* = 10	*k* = 12
0.5%	.578 (.007)	.077 (.004)	.004 (.001)	-- (--)	-- (--)	-- (--)
1.0%	.846 (.005)	.330 (.007)	.052 (.003)	.003 (.001)	.000 (.000)	-- (--)
2.0%	.984 (.002)	.812 (.006)	.427 (.007)	.124 (.005)	.019 (.002)	.001 (.001)
3.0%	.999 (.000)	.968 (.002)	.816 (.005)	.500 (.007)	.206 (.006)	.055 (.003)
4.0%	1.000 (.000)	.996 (.001)	.958 (.003)	.807 (.006)	.539 (.007)	.257 (.006)

Relative RMSE of census numbers	Probability that number of misallocated seats equals or exceeds *k*					
	k = 14	*k* = 16	*k* = 18	*k* = 20	*k* = 22	*k* = 24
0.5%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
1.0%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
2.0%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
3.0%	.008 (.001)	.001 (.000)	-- (--)	-- (--)	-- (--)	-- (--)
4.0%	.084 (.004)	.020 (.002)	.003 (.001)	.001 (.000)	-- (--)	-- (--)

-- signifies number < 0.02%. Number in parentheses is estimated standard error of probability.

***Table S6.* Estimated probability distribution of number of House seats misallocated, accurate small states case accuracy profile.**

Relative RMSE of census numbers	Probability that number of misallocated seats equals or exceeds *k*					
	k = 2	*k* = 4	*k* = 6	*k* = 8	*k* = 10	*k* = 12
0.5%	.744 (.006)	.179 (.005)	.015 (.002)	.000 (.000)	-- (--)	-- (--)
1.0%	.955 (.003)	.611 (.007)	.196 (.006)	.031 (.002)	.003 (.001)	.000 (.000)
2.0%	.999 (.000)	.970 (.002)	.813 (.006)	.490 (.007)	.190 (.006)	.045 (.003)
3.0%	1.000 (--)	.998 (.001)	.982 (.002)	.900 (.004)	.699 (.006)	.424 (.007)
4.0%	1.000 (--)	1.000 (--)	.998 (.001)	.988 (.002)	.934 (.004)	.804 (.006)

Relative RMSE of census numbers	Probability that number of misallocated seats equals or exceeds *k*					
	k = 14	*k* = 16	*k* = 18	*k* = 20	*k* = 22	*k* = 24
0.5%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
1.0%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
2.0%	.007 (.001)	.001 (.000)	-- (--)	-- (--)	-- (--)	-- (--)
3.0%	.196 (.006)	.062 (.003)	.014 (.002)	.003 (.001)	.002 (.001)	.000 (.000)
4.0%	.607 (.007)	.377 (.007)	.185 (.005)	.074 (.004)	.029 (.002)	.008 (.001)

-- signifies number < 0.02%. Number in parentheses is estimated standard error of probability.

***Table S7.* Estimated probability distribution of number of House seats misallocated, differential bias accuracy profile.**

Relative RMSE of census numbers	Probability that number of misallocated seats equals or exceeds *k*					
	k = 2	*k* = 4	*k* = 6	*k* = 8	*k* = 10	*k* = 12
0.5%	.662 (.005)	.123 (.003)	.008 (.001)	.001 (.000)	-- (--)	-- (--)
1.0%	.902 (.003)	.462 (.005)	.106 (.003)	.011 (.001)	.001 (.000)	-- (--)
2.0%	.995 (.001)	.911 (.003)	.610 (.005)	.256 (.004)	.066 (.002)	.011 (.001)
3.0%	1.000 (.000)	.990 (.001)	.919 (.003)	.707 (.005)	.409 (.005)	.169 (.004)
4.0%	1.000 (--)	.999 (.000)	.988 (.001)	.936 (.002)	.788 (.004)	.544 (.005)

Relative RMSE of census numbers	Probability that number of misallocated seats equals or exceeds *k*					
	k = 14	*k* = 16	*k* = 18	*k* = 20	*k* = 22	*k* = 24
0.5%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
1.0%	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)	-- (--)
2.0%	.001 (.000)	.000 (.000)	-- (--)	-- (--)	-- (--)	-- (--)
3.0%	.050 (.002)	.009 (.001)	.001 (.000)	.000 (.000)	-- (--)	-- (--)
4.0%	.294 (.005)	.124 (.003)	.039 (.002)	.009 (.001)	.001 (.000)	.000 (.000)

-- signifies number < 0.02%. Number in parentheses is estimated standard error of probability.

S5. Expected Sums of Errors in Fund Allocations Due to Census Error

A recent study (*21*) found census data affect the distribution of hundreds of billions of dollars in allocations from more than 100 different programs (132 programs allocated more than $675 billion in FY 2015). This updated an earlier study's finding (*28*) that 140 federal grant and direct assistance programs distributed approximately $450 billion in FY 2007 at least partly on the basis of population and income data. Analyzing the effect of census error on fund allocations is more complicated than for apportionment. There are many allocation programs and they typically are complex, involving statistics other than just state population numbers and using census numbers in different ways. To model the accuracy of the other statistics in even a single program can itself be a major undertaking even for a retrospective analysis (*23*). The various statistics in the allocation formulas change over time, and if we were to use 2020 population numbers for simulating fund allocations, we should also use future values of the other statistics in computing the allocations. We had no confidence that we could forecast the future values of the other statistics accurately even if we had the resources to carry out the forecasting, and so we used the latest available numerical values of all of the statistics the government used to compute the allocations as if they were true numbers. To analyze the effect of census error, we used the models described in Section S3.2. This approach of conditioning on observed statistics as if they were true and adding error to the census population numbers may, depending on the extent of biases in the other statistics, lead to overstatement of the effect of census error (*19*).

A stratified simple random sample of 18 formula-based fund allocation programs was selected from the 140 listed in (*28*) as using Census Bureau population or income data to determine the allocations. We selected with certainty the 8 largest programs, which accounted for 4/5 of the total FY 2007 obligations, and we selected a disproportionate stratified sample of

10 of the remaining 132 programs. The sample design and selected programs are shown in *Table S8*. The sampling inclusion probability for a program in stratum h is equal to n_h / N_h, with n_h the sample size and N_h the population size in the stratum. Sampling weights were set equal to N_h / n_h, the reciprocal of the inclusion probability.

For each selected program, we analyzed the effect of census error on allocations, as described in Section (*19*, 31-37). For any given parametric model of census error, the sum of misallocations for the selected program was simulated, just as described for apportionment in Section S3.3 except that the true value for the population was held fixed. The average across simulations was calculated for each program. The average was then multiplied by the ratio of the FY2015 obligation from (*21*) to the total amount allocated for the year for which the data were available and analyzed. The ratio-adjusted amount provides an estimate of the sum of FY 2015 misallocations due to census error for the selected program. Finally, results were multiplied by ten to reflect estimates of the effect of the decennial census on the sum of misallocations over a decade in 2015 dollars.

The weighted sum of the latter (ratio-adjusted amounts) was calculated, using sampling weights equal to N_h / n_h, with h denoting the stratum to which the program belonged. The weighted sum estimates the sum of the expected values of misallocations for all 140 allocation programs in (*28*) if their allocated amounts were equal to the FY2015 obligations in (*21*). However, the population sampled (a) excludes 7 programs that came into being between FY2007 and FY2015, and whose FY2015 obligations totaled $93.9 billion, and (b) includes 15 FY2007 programs that did not exist in FY2015, totaling $2.3 billion in 2015 dollars (*19*). (FY2007 dollars were converted to FY2015 dollars according to the Consumer Price Index for Urban Wage Earners and Clerical Workers (*38*), yielding an adjustment factor of 231.810/202.767 = 1.143.)

Sampling errors (reflecting finite number of simulations as well as sampling of the allocation programs) for FY2009 amounts were moderate, with margins of error (two standard errors) of about 15% of the total being estimated (*19*, 27). The sampling errors were not calculated for the FY2015 amounts, but the use of ratio-adjustment suggests that they will be similar in percentage terms.

The resulting weighted estimates are shown in *Figure 1* for the uncorrelated errors model and the correlated errors model.

Estimates of the expected sum of misallocations for other error models may be obtained from the regression model indicated in the second row of *Table S2*, according to the directions provided in Section S4.1. The regression model for predicting expected sums of absolute misallocations was fit analogously to that for apportionment errors, with an achieved R^2 of 0.984. The nominal *p*-values (assuming normality) for all regression coefficients were below 0.001. Further details are in (*19*). The empirical findings about sensitivity of malapportionment to the error parameters are qualitatively similar for fund allocations; see (*19*) for details

***Table S8*. Sampled programs allocating federal funds.** Source: columns 2 – 4 from (*28*) and columns 5 – 6 from (*21*).

Stratum h	$\frac{n_h}{N_h}$	CFDA No.	Program	FY 2015 Obligation ($Billions)	Weighted FY 2015 Obligation ($Billions)
1	$\frac{8}{8}$	93.778	Medical Assistance Program (Medicaid)	$311.8	$311.8
		17.225	Unemployment Insurance	$3.0	$3.0
		20.205	Highway Planning and Construction	$38.5	$38.5
		10.551	Suppl. Nutrition Assistance Program (SNAP)	$71.0	$71.0
		93.558	Temporary Assist. for Needy Families (TANF)	$17.2	$17.2
		84.063	Federal Pell Grant Program	$29.9	$29.9
		84.010	Title I Grants to Local Educ. Agencies (LEAs)	$14.3	$14.3
		84.027	Special Education Grants to States	$11.4	$11.4
2	$\frac{2}{3}$	93.600	Head Start	$8.5	$12.8
		93.767	State Children’s Insurance Program (CHIP)	$4.2	$6.3
3	$\frac{2}{6}$	10.557	Special Supplemental Nutrition Program for Women, Infants, and Children (WIC)	$6.1	$18.2
		93.596	Child Care Mandatory and Matching Funds	$5.3	$15.9
4	$\frac{2}{12}$	93.575	Child Care and Development Block Grant	$0.0	$0.1
		93.667	Social Services Block Grant	$1.6	$9.5
5	$\frac{2}{16}$	84.365	English Language Acquisition Grants	$0.7	$5.8
		84.181	Special Ed. – Grants for Infants and Families	$0.4	$3.4
6	$\frac{2}{95}$	66.460	Nonpoint Source Implementation Grants	$0.1	$5.7
		16.458	Title V Delinquency Prevention Program	$0.0	$0.0
Total	$\frac{18}{140}$				$574.9

S6. Analyzing Effects of Census Error on Sampled Allocation Programs

S6.1. Roles of census population numbers in 18 sampled allocation programs

To analyze the effects of census error on allocations by the 18 sampled grant and assistance programs requires understanding how census numbers are used in each of the programs. Here we provide an overview. Details about each program are provided in the Appendix.

Table S9 shows the kinds of statistics used to allocate funds across the 18 sampled programs.

- Postcensal estimates from the Census Bureau's Population Estimate Program are used in 9 of the 18 programs.
- Two programs use model-based estimates for small-area populations, and census population data are directly or indirectly used to fit the models. Title I Grants to Local Education Agencies uses Small Area Income and Population Estimates (SAIPE) for school district school-age children in poverty. The Supplemental Nutrition and Assistance Program for Women, Infants and Children uses a model-based estimate of the number of children age 1 to 4 below 185% of the poverty line.
- Two programs use American Community Survey (ACS) estimates. Special Education Grants to State uses information on state Free Appropriate Public Education age children in poverty from ACS Public Use Microdata. English Language Acquisition Grants uses ACS data on Limited English Proficiency children and foreign-born children.
- Current Population Survey (CPS) unemployment rates help determine whether states are eligible for additional Unemployment Insurance (UI) assistance. The CPS uses postcensal estimates as ratio controls for totals.
- Three programs, Supplemental Nutrition and Assistance Program, Pell Grants and Head Start, all make awards based on poverty thresholds. The poverty thresholds are developed using the Consumer Price Index for all Urban Workers (CPI-U) as a measure of inflation. CPI-U is estimated in part with a sampling frame that uses the decennial census (BLS 2007).
- Five programs also use non-census statistics in formula-based allocation. For example, Medicaid awards use both census population numbers and BEA personal income.

- For 3 of the 18 selected programs, the allocations would not be affected by error in the most recent census: Highway Planning and Construction, Temporary Assistance for Needy Families and Nonpoint Source Implementation Grants. These three programs have used census data for past allocations, but future allocations are fixed to previous state shares.

Table S9. **Statistics used in formulas for allocating federal funds.** Source: (*39*)–(*41*).

	Statistic Depending on Census Population						
Allocation Program	Midyear Postcensal Pop. Est.	Model-based Pop. Est.	ACS Pop. Est.	CPS Unempl. Rate	CPI-Urban	Non-Census Stats Used	Latest Census Not Used
Medicaid	•					•	
CHIP	•					•	
Child Care Mandatory+Matching	•						
Child Care and Development	•					•	
Social Services Block Grant	•						
Special Ed. – Infants & Families	•						
Title V Delinquency Prevention	•						
Title I Grants to LEAs	•	•				•	
Special Ed. – States	•		•				
WIC		•					
English Language Acquisition			•				
Unemployment Insurance				•			
SNAP					•		
Pell Grants					•	•	
Head Start					•		
Highways							•
TANF							•
Nonpoint Source Implementation							•

S6.2. Approximations and analytic simplifications

Several analytic simplifications were used for analyzing the effect of census error on the allocations. Except as noted, the simplifications tend to overstate the effect of census error on allocations.

Unlike apportionment, which depends only on census population, the fund allocation programs involve other statistics in addition to census population. To avoid analyzing the accuracy of all of the statistics, we conditioned on the observed values of the non-census statistics. If the allocation to a state is represented by $f(x, y)$, where y denotes the census number and x denotes other statistics, then the expected absolute misallocation may be expressed as $E\,|\,f(x, y) - f(x^*, y^*)\,|$, where x^* and y^* denote the true values of x and y. We approximated this by $E\,|\,f(x, y) - f(x, y^*)\,|$, conditioning on the observed values of x. Analysis suggests that the approximation overstates the effect of census error in some general scenarios and that the potential understatement tends to be smaller than the potential overstatement (*19*).

Mid-year postcensal population estimates adjust the census estimate for births, deaths and net migration since the census. We approximated the relative error in the postcensal estimate by the relative error in the underlying base census number. This approximation overstates the effect of census error on the postcensal estimate, since the errors in estimates of change due to births, deaths, and net migration are only somewhat dependent on the census base (*34*). Specifically, the relative effect of census error on the census base overstates the relative effect of census error on the sum of the census base and other components only somewhat affected by census error.

Model-based and ACS population estimates are used to calculate the proportion of the population in a group or area. The proportion is multiplied by a census or postcensal estimate of total population to estimate the number in the group or area. Here too, the relative error in the model-based or ACS estimate of population of the subgroup is approximated by the relative error in the underlying base census number. Since the errors in model-based and ACS estimates of fractions are mostly independent of the census base, the effect of census error on the census base approximates the effect of census error on the product of the census base and the model-based or ACS estimate of the population proportion.

Census error affects CPS estimates of unemployment rates. To analyze the effect, we first estimated the relationship between census error and unemployment rate error by applying using differential 2010 net undercount estimates to unemployment estimates by age, sex and race. Next, we made the simplifying assumption that the effect of undercount by age, race and sex on unemployment rate estimates is proportional to the effect of state-level census errors on unemployment rate estimates (*19*, 182-183).

Census error affects the consumer price index CPI-U. To analyze the effect, we used a similar approach to that for the CPS unemployment rate estimates, using information about differential undercount for renters and owners (*19*, 183-184).

Title I Grants to LEAs (local education agencies) provide grants to sub-state areas, namely the LEAs, which are often school districts. We simplify the analysis by studying errors in allocation at the state-level alone. The models apply the state relative errors to each LEA population estimate within the state. We conjecture that this approach slightly understates the effect of census error on the LEA-level Title I allocations.

For programs that depend upon multiple census-based statistics, we assume that same relative errors apply to all statistics, which may overstate the effect of census error.

Appendix:

Use of Census Statistics in 18 Sampled Programs for Allocating Federal Funds

This appendix describes how census numbers are used for the 18 sampled grant and assistance programs, with each program's Catalog of Federal Domestic Assistance (CFDA) number denoted in parentheses. Sources of information about programs include (28), (39) – (41), program websites and correspondence with employees at government agencies. Further details are available in (*19*). Throughout this appendix, for statistic $X_i \geq 0$ in state i, denote state i's national share for the statistic by $\breve{X}_i = X_i / \sum_j X_j$.

A1. Medical Assistance Program (93.778)

The Medical Assistance Program pays a fraction of state medical expenditure as determined by the state's Federal Medical Assistance Percentage (FMAP), which is based on a state's relative per capita income. The FMAP has a minimum of 0.50 and a maximum of 0.83. Specifically, the grant amount G_i for state i is $G_i = E_i \times \max\left\{0.50, \min\left\{0.83, 1-0.45(\breve{I}_i / \breve{P}_i)^2\right\}\right\}$. Here, E_i is medical expenditures, I_i is BEA personal income and P_i is population. This formula does not apply to D.C., which has a fixed FMAP of 0.70.

A2. Unemployment Insurance (17.225)

When state unemployment rates are above certain thresholds, unemployment insurance recipients are eligible for extra weeks of compensation through Extended Benefits and, during the last recession, Emergency Unemployment Compensation. The U.S. government provides part of the funding for these two programs. State unemployment rates are estimated by the Current Population Survey.

A3. Highway Planning and Construction (20.205)

The Moving Ahead for Progress in the 21st Century Act passed in 2012 changed the funding formulas for programs administered through Highway Planning and Construction. Allocations for the various programs are fixed at proportions states received in previous years and do not depend on new population statistics.

A4. Supplemental Nutrition Assistance Program (10.551)

A recipient's eligibility for SNAP benefits and amount received are based on poverty threshold which are revised annually based on CPI-U. CPI-U is estimated in part using postcensal population estimates as ratio controls.

A5. Temporary Assistance for Needy Families (93.558)

The grant amounts to states are fixed to the proportions of the grants in 2002. New census statistics are not used in the determination of grant amounts.

A6. Federal Pell Grant Program (84.063)

A student's Pell Grant amount is determined by the cost of attendance and the Expected Family Contribution. The formula to determine the Expected Family Contribution is revised each year based on measures of inflation. The inflation measures are estimated in part using postcensal population estimates as ratio controls.

A7. Title I Grants to Local Educational Agencies (84.010)

Title I funding consists of four sets of grants to Local Educational Agencies (LEAs): Basic Grants, Concentration Grants, Targeted Grants and Education Finance Incentive Grants (EFIG).

All four grant programs depend on Small Area Income and Poverty Estimates (SAIPE) data for the age 5 to 17 population in poverty for each LEA.

The *Basic Grant amount* G_i^{Basic} for LEA i is specified by

$$G_i^{\text{Basic}} = N_i\left[\min\left(0.48E_{US}, \max\left(0.32E_{US}, 0.4E_i\right)\right)\right]\chi_{\{N_i \geq 10\}}\chi_{\{N_i/T_i \geq 0.02\}}. \tag{1}$$

Here, N_i is a measure of the number of children at-need in the LEA, specifically the SAIPE age 5—17 population in poverty; T_i is the total number of school-aged children; E_i is the per-pupil expenditures in the state that includes the LEA; E_{US} is the national per-pupil expenditure; $\chi_{\{A\}}$ is the indicator function taking the value 1 if A is true and 0 otherwise.

The *Concentration Grant amount* $G_i^{\text{Concentration}}$ for LEA i is specified by

$$N_i\left[\min\left(0.48E_{US}, \max\left(0.32E_{US}, 0.4E_i\right)\right)\right]\chi_{\{N_i \geq 6500 \text{ OR } N_i/T_i \geq 0.15\}}, \tag{2}$$

with all variables having the same definitions as they do for Basic Grants.

The *Targeted Grant amount* G_i^{Targeted} for LEA i is specified by

$$W_i N_i\left[\min\left(0.48E_{US}, \max\left(0.32E_{US}, 0.4E_i\right)\right)\right]\chi_{\{N_i \geq 10\}}\chi_{\{N_i/T_i \geq 0.05\}}, \tag{3}$$

with all variables having the same definitions as they do for Basic Grants, and including W_i, a weight between 1.0 and 4.0 that increases with N_i and N_i / T_i and depends on county or school district administration of the LEA.

The *EFIG amount* G_i^{EFIG} for LEA i is specified by

$$N_i\left[\min\left(0.46E_{US}, \max\left(0.34E_{US}, 0.4E_i\right)\right)\right] \times \text{Effort}_i \times \left(1.3 - \text{Equity}_i\right)\chi_{\{N_i \geq 10\}}\chi_{\{N_i/T_i \geq 0.05\}}, \tag{4}$$

where all variables have the same definitions as they do for basic grants, and including Effort_i, the effort factor, and Equity_i, the equity factor for the LEA's state. The effort factor depends upon per capita personal income (BEA) and thus indirectly on census population statistics. Specifically, $\text{Effort}_i = \left(E_i / E_{US}\right)\left(\breve{P}_i / \breve{I}_i\right)$, where E_i and E_{US} are as defined above and $\breve{P}_{State} / \breve{I}_{State}$ is algebraically equal to the ratio of US per capita income to state per capita income. The equity factor for the state, Equity_i, depends on a weighted coefficient of variation of LEA per-pupil expenditure within the state to which LEA i belongs, with the weighting depending on N_i and T_i.

All four grants are ratably reduced to sum to the total amount allocated for each grant program. Adjustments are made so each state receives a minimum amount for each of the four grants.

A8. Special Education Grants to States (84.027)

Grants to states use measures of states' Free Appropriate Public Education (FAPE) age population (usually age 3-21 population) and states' FAPE age population in poverty. The FAPE age population is taken from postcensal population estimates by single year of age. The Office of Planning, Evaluation and Policy Development at the Department of Education stated via personal correspondence the FAPE age population in poverty is determined by combining the postcensal single year of age data with American Community Survey (ACS) Public Use Microdata Sample estimates of the fraction of each age group in poverty. State i receives grant amount G_i, where $G_i = G_i^{(99)} + R\left[0.85\breve{P}_i + 0.15\breve{N}_i\right]$, $G_i^{(99)}$ is the grant amount in 1999, R is the total amount available for the program in excess of 1999 total amount, P_i is the FAPE age population, and N_i is the FAPE age population in poverty.

A9. Head Start (93.600)

Head Start agencies must have a certain percentage of children they serve be from families who are below poverty thresholds. The poverty thresholds are revised based upon CPI-U, which is estimated in part using the decennial census for a sampling frame. Because some Head Start agencies are not fully enrolled, and agencies can respond to changes in poverty threshold eligibility by either increasing or decreasing their effort to recruit students, we take the view that 2020 census error will have negligible effect on Head Start funding.

A10. State Children's Insurance Program (93.767)

The State Children's Insurance Program (CHIP) pays a fraction of state CHIP expenditure as determined by the state's Enhanced Federal Medical Assistance Percentage (eFMAP), which is based on a state's relative per capita income. The eFMAP has a minimum of 0.65 and a maximum of 0.85. The formula for the grant amount G_i for state i is

$G_i = E_i \times \max\left\{0.65, \min\left\{0.85, 1 - 0.315\left(\breve{I}_i / \breve{P}_i\right)^2\right\}\right\}$, where E_i is medical expenditures, I_i is BEA personal income and P_i is population. This formula does not apply to D.C., which has a fixed eFMAP of 0.79.

A11. Special Supplemental Nutrition Program for Women, Infants and Children (10.557)

WIC uses two grant programs, one for food costs and another for nutrition services and administrative (NSA) costs. Food grant amounts are proportional to a state's model-based estimate of the number of children age 1 to 4 below 185% of the poverty line. NSA grants use

BEA measures of food cost inflation. We treat the NSA grants as unaffected by census inaccuracy.

A12. Child Care Mandatory and Matching Funds of the Child Care and Development Fund (93.596)

Mandatory funds are allocated based on a state's share of expenditures for the now-repealed AFDC child care programs during the years 1992-1995, and thus not based on government statistics. Matching funds are allocated to be proportional to a state's population under age 13, which is determined from postcensal estimates by single year of age.

A13. Child Care and Development Block Grant (93.575)

Grants to states depend on a variety of statistics. Census statistics include state population and state population under age 5 (from postcensal single year of age estimates). Non-census statistics include BEA personal income and the number of children receiving free or reduced school lunch from the Department of Agriculture. Specifically, state i receives grant

$$G_i = 0.5A\left(\breve{Y}_i H_i / \sum_j \breve{Y}_j H_j + \breve{L}_i H_i / \sum_j \breve{L}_j H_j\right).$$

Here, A is the total amount allocated for the program, Y_i is population under age 5, L_i is free or reduced lunch population, and $H_i = \min\left\{1.2, \max\left\{0.8, \breve{P}_i / \breve{I}_i\right\}\right\}$, where P_i is population and I_i is state personal income. So H_i is the ratio of U.S. per capita income to state per capita income, constrained to be no less than 0.8 and no more 1.2.

A14. Social Services Block Grant (93.667)

Grants to states are proportional to state population, based on postcensal estimates.

A15. English Language Acquisition Grants (84.365)

Grants to states use measures for the number of Limited English Proficiency (LEP) children and the number of immigrant children and youth, both estimated from the American Community Survey. Specifically, state grants are proportional to the sum of 80% of the state's national share of LEP children plus 20% of the state's national share of immigrant children and youth.

A16. Special Education-Grants for Infants and Families (84.181)

Grants to states are proportional to state population age 0 to 2, which is obtained from postcensal estimates by single year of age. Each state receives a minimum of 0.5% of all funding allocated for the grant program.

A17. Nonpoint Source Implementation Grants (66.460)

Grants to states are determined using a formula based on a variety of government statistics, including 1990 census state population and 1987 postcensal state population estimates. New census statistics are not used in the determination of grant amounts.

A18. Title V Delinquency Prevention Program (16.548)

Grants to states are proportional to a state's youth population under the maximum age of original juvenile court delinquency jurisdiction, which varies by state and is obtained from postcensal population estimates by single year of age.