Effects of Census Accuracy on Apportionment of Congress and Allocations of Federal Funds

Zachary H. Seeskin[1,2] and Bruce D. Spencer[2,3]

Abstract

The accuracy needed for the 2020 census depends on the cost of attaining accuracy and on the consequences of imperfect accuracy. While the cost target for the 2020 census of the United States has been specified, and the Census Bureau is developing projections of the accuracy attainable for that cost, it is also important to have information about the consequences of the accuracy that is attainable for a given cost. To assess the consequences of imperfect census accuracy, we consider alternative profiles of accuracy for states and assess their implications for apportionment of the U.S. House of Representatives and for allocation of federal funds. An error in allocation is defined as the difference between the allocation computed under imperfect data and the allocation computed with perfect data. Estimates of expected sums of absolute values of errors are presented for House apportionment and for federal funds allocations.

Key Words: Data Use, Data Quality, Data Cost, Cost-Benefit Analysis, Population, Government Statistics

1. Introduction

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.
>
> — The Constitution of the United States, Article I, Section 2, as amended by the 14th amendment

Although the Constitution requires a census, it does not say how accurate the census should be. Accuracy and cost are closely related. The Census Bureau can increase accuracy by spending more money, at least up to a point. As the great demographer Nathan Keyfitz (1979, 46) 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

[1]Doctoral Candidate, Department of Statistics, and Graduate Research Assistant, Institute for Policy Research, Northwestern University, Evanston, IL, 60208-4070, U.S.A. Email: <z-seeskin@u.northwestern.edu>.

[2] Research supported by NSF grant SES-1129475, "NSF-Census Research Network: Census Bureau Data Programs as Statistical Decision Problems."

[3] Professor, Department of Statistics, and Faculty Fellow, Institute for Policy Research, Northwestern University.

manufactured products often have defects, or why the police cannot catch all criminals." Accuracy can be increased through investment of more resources in the census, but the accuracy will never be perfect.

The government's current strategy for choosing census accuracy is to specify a cost target and optimize the accuracy that can be attained for that cost. The cost target, as we understand it, is consistent with recommendation of the National Research Council (2011, Recommendation 3) that the cost per housing unit for 2020 be kept at the same (inflation-adjusted) level as for 2010. The 2010 census was estimated to be quite accurate for the total U.S. population, so that the national net undercount was estimated to be nearly zero (Census Bureau 2012). However, most uses of the census depend on population sizes for geographic areas or demographic subgroups, and the census was estimated to have a 0.8 percent national overcount of non-Hispanic whites, a 2.1 percent national undercount of blacks, a 1.5 percent undercount of Hispanics (Census Bureau 2012). There were net overcounts for some states and net undercounts for others.

The question of what accuracy is attainable for the specified cost is complex and is being studied by the Census Bureau. In this paper, we address the related question of the consequences of a given profile of census accuracy. We refer to profiles of accuracy rather than levels, because census accuracy, like census statistics, is multi-dimensional. For example, census population numbers are produced for state and local governments and much smaller areas, and for demographic subgroups both nationally and by geography.

To understand the consequences of imperfect accuracy, one needs to know how census data get used. 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 than what they would be if the census numbers had no error. We will refer to the differences as, respectively, errors in allocation (misallocations) of funds, errors in apportionment (malapportionment) and errors in district sizes.

Not all uses of the census are or can be known, and it is important to acknowledge 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, but we conjecture that it is important. 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. Given the challenges involved in study these other uses of statistics, this paper focuses on studying the effects of census accuracy for Congressional apportionment and allocations of federal funds. Methods are described in Section 2, followed by results in Section 3 and conclusions in Section 4.

2. Methods

2.1 Overview
We measure the distortions in allocations of representation and funding among states that are projected to occur at alternative profiles of accuracy. The funding formulas and the apportionment algorithm are treated as fixed, and the allocations that would occur with error-free statistics are treated as *true values* for the allocations. The allocations that occur with the census population numbers are the *estimated* allocations a_j^{est}. The difference $a_j^{est} - a_j^{true}$ is the *error in allocation*, or misallocation, to state j. The number of misallocated seats (or funds) is defined as the sum of absolute errors (i.e., sum of absolute values of errors in apportionments), $\sum_j |a_j^{est} - a_j^{true}|$. The value of improving accuracy to reduce misallocations is a political question that relates to the question of how much it is worth spending on the census, but that we do not address here.

We represent accuracy in terms of the multivariate distribution of errors in census estimates of population for states and the District of Columbia (D.C.). The focus on the states and D.C. level is consistent with the uses of census data for apportionment of House seats among states and for allocation of federal funds to states and D.C. The error in an estimate or statistic is the difference between the statistic and its true value. The mean squared error (MSE) is the expected value of the squared error, and it is equal to the square of the bias plus the square of the standard deviation. The root mean squared error (RMSE) is the square root of the MSE, and the relative root mean squared error (relative RMSE) is the RMSE expressed as a percentage of the true value being estimated. If the estimate is unbiased, the RMSE is equal to the standard deviation or standard error of the estimate, and the relative RMSE is equal to the coefficient of variation (c.v.). We consider several alternative accuracy profiles for the census, as shown in Table 1.

Accuracy Profile	Description
Base Case	errors for all states (and D.C.) are independently normally distributed, with zero bias, and equal relative RMSE
Correlated Case	same as base case, except estimates for all areas have constant correlation of 0.5
Accurate Small States Case	errors for all states (and D.C.) are independently normally distributed, with zero bias; the smallest 25 states and D.C. have zero relative RMSE; the largest 25 states have constant relative RMSE, such that the average of RMSEs for all 50 states and D.C., weighted by the 2010 state census populations, equals the corresponding relative RMSE for the base case
Differential Bias Case	errors for all states (and D.C.) are independently normally distributed, with equal relative RMSE and with absolute value of the relative bias equal to the c.v.; the bias for the smallest 25 states and D.C. has one sign, and the bias for the largest 25 states has the opposite sign

Table 1. Alternative accuracy profiles for the U.S. Census.

2.2. Apportionment
Since 1941, House apportionment has been determined from census populations using the Method of Equal Proportions (also known as Hill's Method and Huntington's Method). If

fractions of seats could be allocated, then state j could simply receive its quota, q_j, defined as the number of seats h in the House of Representatives (currently $h = 435$) times the fraction of all 50 states' census population held by state j. Letting p_j denote the population of state j and p denote the population of all 50 states, we have $q_j = (p_j / p) \times h$. However, the allocations a_j of seats to state j must be whole numbers – no fractional allocations are allowed. The Method of Equal Proportions chooses positive integers a_j that minimize $\sum_j a_j (p_j / a_j - p / h)^2$ when the quotas q_j are given (Balinski and Young 1982, 1980, 1975; Spencer 1985). The apportionments are computed by the Census Bureau and provided to the President, who transmits them to Congress. Computation of apportionment is described by Balinski and Young (1977) and by Census Bureau (2015). The Method of Equal Proportions is computed in practice by first awarding the first fifty seats one to each state. Seats 51 to 435 are awarded iteratively, each one to the state with the largest value of $p_i / \sqrt{n_i(n_i - 1)}$, where p_i is the census population of state i and n_i is the number of seats already awarded.

The sensitivity of the apportionment to census accuracy depends in part on the values of the underlying true populations of the states. The requirement that the numbers of seats held by states must be integers implies that, for some configurations of states' populations, a change of just a single person can cause the numbers to shift (Keyfitz 1979). For such configurations, even the smallest errors in census numbers will shift the allocations of seats. To analyze how sensitive apportionments are to changes in census quality, we considered a joint distribution for the true state populations and the census numbers for states. To formulate the joint distribution, it is sufficient to consider the distribution for the true population sizes and the conditional distribution for census error given the true population. The error distribution was specified using an accuracy profile from Table 1, with alternative levels of average relative RMSE. The mean of the distribution of true population sizes was set equal to population projections prepared by the Weldon Cooper Center (2013a, 2013b). The true population sizes were taken to be independent, and the coefficient of variation for each state's population was chosen to be consistent with the past level of error in state population forecasts with similar time horizon.

We estimated the distribution of sums of absolute errors in states' allocations of seats by drawing population numbers (true and census numbers) from the joint distribution, as described above. Then the true apportionments a_j^{true} and estimated apportionments a_j^{est} were calculated for each state j, and the sum of absolute misallocations was calculated. This process was repeated 5,000 times independently, and the average sum of absolute misallocations was used to estimate the expected number of malapportioned seats.

2.3. Allocation of federal funds

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. It would be highly complex to jointly forecast the values of all such statistics ahead to 2020, and the results would likely be uncertain. Therefore, we took the simpler approach of obtaining the latest values we could of the statistics used to calculate allocations for the 18 programs we studied, and treating those as if they were error-free. To allow for error in the census, we used the accuracy profiles (Table 1) to first develop a distribution of census error, and we developed a distribution of error in the population statistics used to compute the allocations.

Blumerman and Vidal (2009) identified 140 federal grant and direct assistance programs that distributed approximately $446.4 billion in funds in FY 2007 at least partly on the basis of population and income data from the U.S. Census Bureau. The largest of these is the Medical Assistance Program, also known as Medicaid. Grants to states are equal to state medical expenditures times the Federal Medical Assistance Percentage (FMAP). The FMAP depends on per capita income, which is calculated as the ratio of census population to Bureau of Economic Analysis (BEA) personal income. The formula can be written as

$$\text{FMAP} = \min\left\{ \max\left[1 - 0.45\left(\frac{I_i / I_+}{P_i / P_+} \right)^2, 0.50 \right], 0.83 \right\}, \tag{1}$$

where I_i is BEA personal income, P_i is the census population of state i, $I_+ = \sum_j I_j$, and $P_+ = \sum_j P_j$ (NRC 2003).

For analysis of total misallocated funds across all 140 programs, we selected the 8 largest programs (in terms of FY 2007 obligated amount) with certainty, which accounted for about 80.1% of the total FY 2007 obligations. From the remaining 132 programs, we selected a disproportionate stratified sample of 10, so that larger programs had a higher chance of selection. The programs we selected are shown in Table 2. We used sample-weighting methods to get unbiased estimates of totals for all 140 programs. For each program, results are weighted by the ratio of N_h, the number of total programs in stratum h, to n_h, the number of programs sampled in the stratum. We formed the weighted sum of estimates for sampled programs to estimate the total for all 140 programs. Thus, in total, the 18 programs represent $445.6 billion in funds distributed in 2007 on the basis of population and income data from the U.S. Census Bureau. Note that this number is slightly different from the $446.4 in funds actually distributed in FY 2007, due to the random sampling of programs. For each program studied, we estimated the expected annual amount of absolute errors in state allocations that would arise from alternative specifications of census error distributions. Sampling errors and approximate 95% confidence intervals were also estimated using theory for stratified samples. Table 2 shows the sampled programs.

A variety of statistics were used to allocate funds across the 18 sampled programs.[4]

- Annual mid-year population estimates from the Population Estimate Program are used in 9 of the 18 programs.[5]

- Two programs use model-based estimates for small-area populations that include Census Bureau population data in the models. Title I Grants to Local Education

[4] Information on grant and assistance programs was collected from NRC 2003, Blumerman and Vidal 2009, Catalogue of Federal Domestic Assistance <https://cfda.symplicity.com> and legal codes at <http://law.justia.com/codes/us>.
[5] See <http://www.bea.gov/newsreleases/regional/spi/spi_newsrelease.htm>, accessed June 3, 2015, for information on the use of census mid-year estimates for per capita income.

Agencies uses Small Area Income and Population Estimates 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.

Strat h	N_h	n_h	Program	FY 2007 Obligation ($ Bill.)	Weighted FY 2007 Obligation ($ Bill.)
			Medical Assistance Prog.	$203.5	$203.5
			Unemployment Insurance	$35.9	$35.9
			Highway Planning and Construction	$34.2	$34.2
			Supp. Nutrition Assist. Prog. (SNAP)	$30.3	$30.3
			Temporary Assist. for Needy Families	$16.5	$16.5
			Federal Pell Grant Prog.	$13.7	$13.7
			Title I Grants to Local Educ. Agencies	$12.8	$12.8
1	8	8	Special Ed. – Grants to States	$10.8	$10.8
			Head Start	$6.9	$10.3
2	3	2	State Children's Insurance Prog.	$5.9	$8.9
			Special Supp. Nutrition Prog. for Women, Infants, and Children (WIC)	$5.5	$16.6
3	6	2	Child Care Mandatory & Matching Fund	$2.9	$8.7
			Child Care and Development Block Grant	$2.1	$12.3
4	12	2	Social Services Block Grant	$1.7	$10.2
			English Language Acquisition Grants	$0.6	$4.9
5	16	2	Special Ed. – Grants for Infants & Families	$0.4	$3.5
			Nonpoint Source Implementation Grants	$0.2	$9.5
6	95	2	Title V Delinquency Prevention Prog.	$0.1	$3.0
Total	140	18			$445.6

Table 2. Sampled programs allocating federal funds.

- 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.[6] 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 assistance. The CPS uses decennial census information for its sampling frame.

- Three programs, Supplemental Nutrition and Assistance Program, Pell Grants and Head Start, all make awards based on poverty thresholds. The poverty thresholds are estimated using the Consumer Price Index for all Urban Workers (CPI-U),

[6] ACS Public Use Microdata obtained from Ruggles et. al. (2010).

which 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.

- We found that 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.

Several analytic simplifications were necessary for analyzing the effect of census error on the allocations. Except as noted, the simplifications were chosen to have the effect of overstating the effect of census error on error in allocation.

(i) Unlike apportionment, which depends only on census population, the fund allocation programs involve other statistics in addition to census population. To fully model the diverse sources of error is too vast an undertaking for this project. Spencer (1980a, 67-150) demonstrates the kind of investigations that would be needed. For example, BEA personal income is used in multiple allocation formulas, but its accuracy is unknown. We use an approximation that conditions on the observed values of the non-census statistics. If we represent the allocation to a state by $f(x,y)$, where y denotes the census estimates and x denotes other statistics, then the expected absolute misallocation may be expressed as $E\,|\,f(x,y)-f(x^{true},y^{true})\,|$, where x^{true} and y^{true} denote the true values of x and y. We approximate this by $E\,|\,f(x,y)-f(x,y^{true})\,|$, conditioning on the observed values of x. Work in progress suggests that the approximation overstates the effect of census error in some general scenarios and that the potential understatement is smaller than the potential overstatement.

(ii) For cases where population enters the allocation formula as a mid-year population estimate, which adjusts 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 (Spencer 1980b). 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.

(iii) 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, we approximated the relative error in the model-based or ACS estimate of population of the subgroup by the relative error in the underlying base census number. Since the errors in model-based and ACS estimates of fractions are largely 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.

(iv) To model the effect of census error on CPS unemployment rates, we first estimated the relationship between census error and unemployment rate error using differential net undercount estimates by age, sex and race in 2010 and applying these to unemployment estimates for these three groups. We then 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 census errors on unemployment rate estimates. For CPI-U, we proceeded similarly using differential price indices for renters and owners together with information on renter and owner net census undercount.

(v) Title I Grants to LEAs provide grants to sub-state areas, namely school districts. We take the simplifying approach of studying errors in allocation at the state-level alone. Our 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.

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

We estimated the expected sum of absolute errors in allocations for the year for which the most recent data was available. In order to obtain estimates corresponding to FY 2007, we ratio-adjusted the estimates of sum of absolute errors by the ratio of the FY 2007 program obligations to the allocations for the year for which allocations were analyzed. Typically, this was a downward adjustment. We conducted 5,000 independent simulations of census numbers and found absolute errors for each federal program analyzed. We used sampling theory to estimate the total expected misallocated funds for FY 2007 for all 140 programs.

3. Results

3.1. Apportionment

Figure 1 presents results for the relative RMSE of state population estimates and the expected number of House seats going to the wrong states across the four different census accuracy profiles studied. The relationship is linear for each of the accuracy profiles. With high levels of census inaccuracy, there can be a large number of House seats malapportioned, with an expected malapportionment of 13.3 in the base case with a 4.0% average relative RMSE. The results are sensitive to the accuracy profile, with the base case and accurate small states cases associated with the greatest malapportionment and the correlated case associated with the smallest malapportionment. The estimates presented in Figure 1 are shown in Table 3. Standard errors for all estimates are less than 0.05 House seats.

For any particular census with an associated accuracy profile, the number of seats malapportioned could be much greater than the expected number. Figure 2 presents the estimated probabilities of *k* or more malapportioned House seats under the base case accuracy profile. With relative RMSE, or c.v., equal to 1%, the expected number of malapportioned seats is estimated to be 3.4, but there is a 1 in 6 chance that 6 or more seats are malapportioned. With relative RMSE equal to 4%, there is a 30% chance that 16 or more seats are malapportioned.

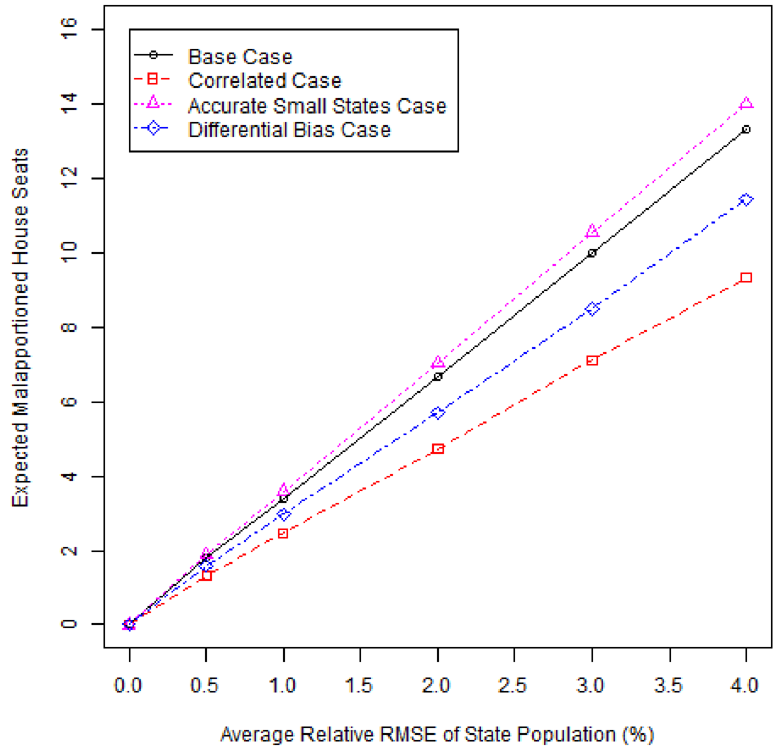

Figure 1. Expected sum of malapportioned seats in the U.S. House of Representatives under alternative profiles of census accuracy.

Estimated Expected Number of Malapportioned Seats						
	Average Rel. RMSE of State Population Numbers					
Accuracy Profile (Case)	0.0%	0.5%	1.0%	2.0%	3.0%	4.0%
Base	0.00	1.79	3.38	6.66	10.00	13.32
Correlated	0.00	1.32	2.46	4.74	7.11	9.33
Accurate Small States	0.00	1.88	3.59	7.03	10.56	14.01
Differential Bias	0.00	1.59	2.96	5.70	8.51	11.44

Table 3. Estimated expected number of malapportioned seats in the U.S. House, by different census accuracy profiles. (Estimated standard errors for all numbers do not exceed 0.05.)

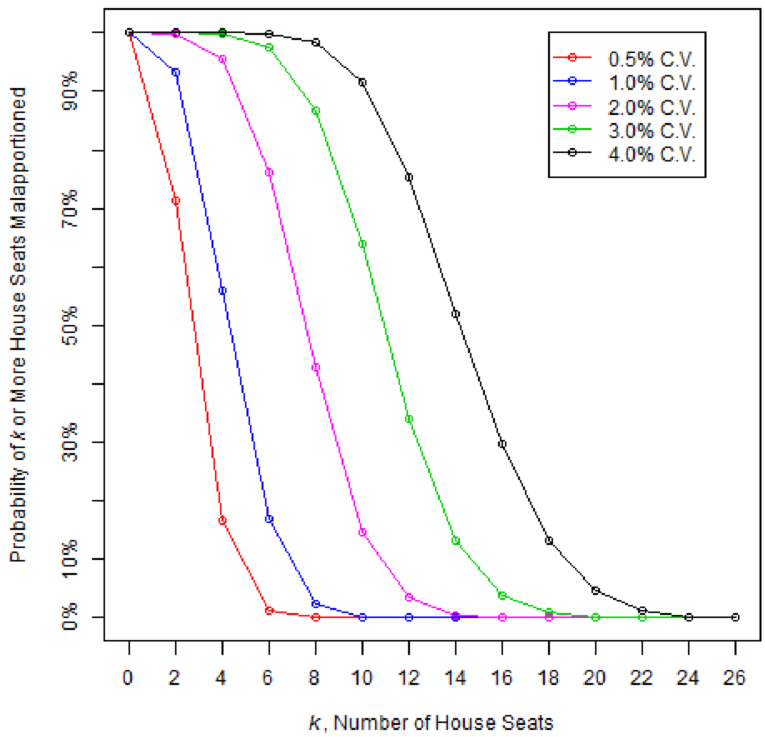

Figure 2. Estimated distribution of number of House seats malapportioned, base case census accuracy profile.

3.2. Formula-based allocations of federal funds

Estimates of the expected misallocations due to census error across all federal programs are presented in Figure 3 and in Table 4. Figure 3 presents results over 10 years by taking the results for fiscal year 2007 and multiplying by them by 10, while Table 4 presents estimated expected misallocations for 2007 alone. Error bars indicate 95% confidence intervals for the estimates. Most of the uncertainty in estimates is due to the random sampling of federal programs, but some is due to variance in results across the simulations. The right hand vertical axis shows the estimated expected sum of misallocations as a percentage of the total funds allocated across the 140 programs.

Like apportionment, the relationship between the average relative RMSE of state population numbers and expected misallocations is approximately linear for all four accuracy profiles. For a 4.0% average relative RMSE, we estimate that the expected misallocation over four years is around $80 billion, or about 1.8% of total allocations. The base case and accurate small states cases are associated with the greatest misallocations, while the correlated case is associated with the smallest misallocation.

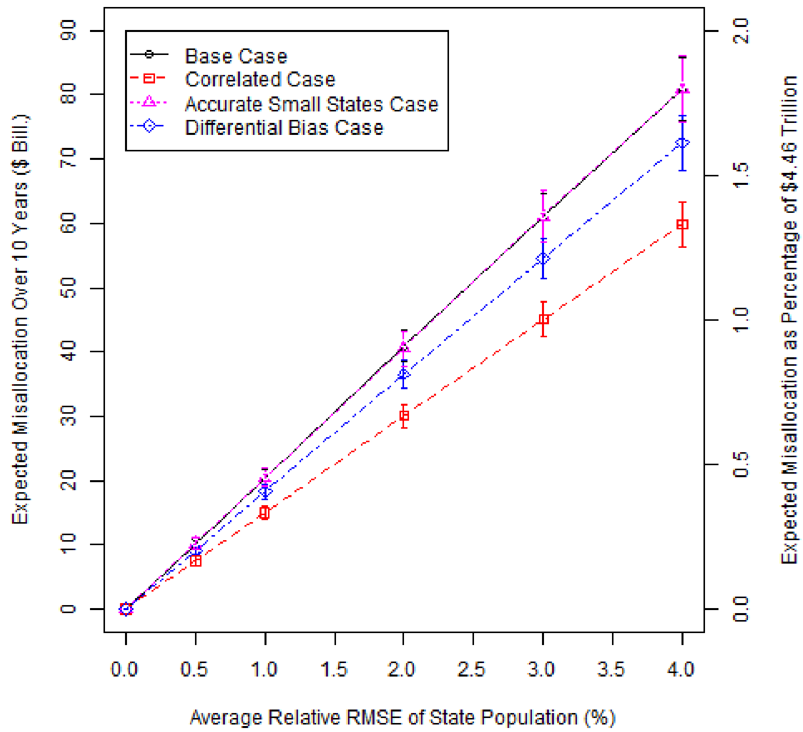

Figure 3. Expected sum of absolute errors in federal fund allocations to states over 10 years, as related to census accuracy. Error bars represent 95% confidence intervals.

	Estimated Expected Misallocated Funds in One Year ($ Billions)				
	Average Relative RMSE of State Population Numbers				
Accuracy Profile (Case)	0.5%	1.0%	2.0%	3.0%	4.0%
Base	1.02	2.05	4.10	6.09	8.08
standard error	*0.04*	*0.07*	*0.13*	*0.19*	*0.25*
Correlated	0.75	1.50	3.01	4.51	5.99
standard error	*0.03*	*0.05*	*0.09*	*0.14*	*0.18*
Accurate Small States	1.02	2.05	4.05	6.11	8.09
standard error	*0.04*	*0.07*	*0.14*	*0.20*	*0.27*
Differential Bias	0.92	1.83	3.66	5.45	7.25
standard error	*0.04*	*0.06*	*0.11*	*0.16*	*0.22*

Table 4. Estimated expected absolute misallocations of federal funds for FY 2007, by different census accuracy profiles.

Estimates of the FY 2007 expected sum of misallocations by federal program for the base case accuracy profile, with results weighted by the sampling strata, are presented in Table 5. As the results are weighted, the numbers reflect each program's contribution to the overall estimate of misallocated funds across all 140 federal programs. Medicaid represents more than half of the total estimated misallocated funds. Some programs from the probability sample, such as the Child Care and Development Block Grant, contribute large amounts to the estimate of total misallocated funds due to the weighting.

	Expected Weighted Sum of Absolute Errors in FY 2007 Allocations Due to Census Inaccuracy ($ Mill.)				
	Average Relative RMSE of State Population Numbers				
Program	0.5%	1.0%	2.0%	3.0%	4.0%
Medical Assistance Prog.	$635	$1,297	$2,606	$3,864	$5,113
Unemployment Ins.	$53	$106	$212	$317	$424
Supp. Nutrition (SNAP)	$14	$29	$58	$86	$115
Pell Grants	$8	$15	$31	$46	$61
Title I Grants to Local Educ.	$48	$97	$194	$290	$387
Spec. Ed. – State Grants	$28	$53	$105	$157	$210
State Children's Ins. Prog.	$15	$31	$63	$94	$125
Spec. Supp. Nutrition (WIC)	$42	$85	$170	$254	$339
Child Care Mand. & Match.	$20	$39	$78	$117	$156
Child Care & Development	$74	$129	$249	$370	$490
Social Services Block Grant	$40	$79	$159	$237	$317
English Lang. Acquis. Grant	$18	$37	$74	$110	$147
Spec. Ed. – Infants & Fam.	$12	$25	$50	$74	$99
Title V Juv. Delinquency	$12	$24	$48	$71	$95
FY 2007 Total	$1,019	$2,046	$4,097	$6,088	$8,080
Total Over 10 Years ($ Bill.)	$10.2	$20.5	$41.0	$60.9	$80.8

Table 5. Expected sum of absolute errors in FY 2007 allocations due to census inaccuracy, base case census accuracy profile.

4. Conclusions

Census inaccuracy can have a large effect on apportionment of the U.S. House of Representatives and on the distribution of more 4 trillion dollars in federal funds over the decade (Figures 1, 3). Our analysis suggests that, if the average absolute relative error in census numbers for states is 4%, then across the scenarios studied, the expected value of the number of House seats going to the wrong state (relative to having perfect data) is in the range of 9 to 14, and between $60 billion and $80 billion dollars in federal grants in aid will go to or from the wrong states over ten years. There is also a possibility of higher error in apportionment and allocation than the expected value. For example, if the average absolute relative error in state census numbers is 4%, then in the base case scenario, there is a 30% chance that 16 or more seats are malapportioned.

It is important to understand the limitations of the scope of the current analysis. In this paper we study only direct and specific "instrumental" uses census statistics for allocating funds and House seats. We have not studied the effect of census quality on "conceptual" uses of census data for scientific research or for policy-making, uses which are vastly more difficult to identify and describe (Beyer 1977). Part of the reason conceptual uses resist study is that they are hidden in chains of analysis. For example, policy X is adopted or theory Y is accepted on the basis of cited research that depended in part on supporting research that depended on past census data, but the role of the census is not apparent. For another example, former OMB Director Peter Orszag (2009, 40) noted that the educational policy goal of increasing the number of postsecondary education was developed to reduce social inequality, based on empirical research of Goldin and Katz (2008, 2007) that relied in key ways on decennial census data from 1940-1980 and on Iowa State Census data from 1915. Not only is it difficult to identify such uses of census data after they have occurred, but it is even more difficult to anticipate them ahead of time. As noted by J. G. March (1994, 246),

> Having knowledge when it is needed often requires an investment in knowledge that is not known to be needed at the time it is acquired. The returns from knowledge may occur in a part of the system quite different from the part where the costs are paid.

Another kind of use statistics is for window dressing, or "using research results to legitimate and sustain predetermined positions" (Beyer 1977, 17). Symbolic uses of data can be sensitive to data quality, as explained by Boruch (1984) among others. Suppose that a decision maker simply wants to use data as window dressing to defend a decision already made. If the data are high quality, then they will more accurately describe the true state. If the decision maker needs false information to justify the decision, which by itself raises questions about the validity of the decision, then that will be more difficult with high quality data.

In addition, the census is used to adjust or calibrate the results of virtually all national sample surveys of the U.S. population in the public and private sectors.

In conclusion, the effects of inaccuracy in the 2020 census on apportionment and allocation of federal funds can be appreciable, and depending on the degree of inaccuracy, large distortions in apportionment and allocation could be expected.

Acknowledgements

We are grateful for the assistance and comments of many people. While noting that responsibility for all views expressed, and any errors in the report, are solely those of the authors, we do gratefully acknowledge and thank the following individuals.

Wesley Basel (Social, Economic, and Housing Statistics Division, Census Bureau), William R. Bell (Associate Director for Research and Methodology), Patrick Cantwell (Decennial Statistical Studies Division), Robert W. Colosi (2020 Research and Planning Office, Census Bureau), Mark Cooley (Center for Medicaid and State Operations, DHHS), Richelle Davis (Office of Special Education Programs, DE), Dennis Finney (Grants Administration Division, EPA), Gay Gilbert (Office of Unemployment Insurance, DOL), Mark Glander (Office of Elementary and Secondary Education, ED), Robert M. Groves (Georgetown University), Howard R. Hogan (Director's Office, Census Bureau), Al Jones (Office of Special Education Programs, DE), Christa D. Jones (Director's Office, Census Bureau), Larry V. Hedges (Northwestern University), Edward L. Kobilarcik

(Decennial Management Division, Census Bureau), Thomas A. Louis (Director's Office, Census Bureau), Charles F. Manski (Northwestern University), Mary H. Mulry (Center for Statistical Research Methodology, Census Bureau), Laura Nixon (Governments Division, Census Bureau), Nancy A. Potok (Director's Office, Census Bureau), Burton H. Reist (2020 Research and Planning Office, Census Bureau), Todd Stephenson (Office of Elementary and Secondary Education, ED), Richard Strauss (Center for Medicaid and State Operations, DHHS), John H. Thompson (Director, Census Bureau), Robert Trombley (Office of Planning, Evaluation and Policy Development, ED), Philip Vidal (Governments Division, Census Bureau), Frank A. Vitrano (2020 Research and Planning Office, Census Bureau), Daniel H. Weinberg (Associate Director for Research and Methodology, Census Bureau), Patrick Wells (Administration for Children and Families, DHHS), Everett G. Whiteley (Budget Office, Census Bureau), Ron Wilus (Office of Unemployment Insurance, DOL).

References

Balinski, M. and H. Young (1975) The quota method of apportionment. *American Mathematical Monthly* 82, 701-730.

Balinski, M. L. and Young, H. P. (1977) On Huntington methods of apportionment. *Siam Journal on Applied Mathematics*, 33, 607-618.

Balinski, M. L. and Young, H. P. (1982) *Fair Representation: Meeting the Ideal of One Man, One Vote*. New Haven: Yale University Press.

Beyer, J. M (1977) Research utilization: bridging the gap between communities. *Journal of Management Inquiry* 6, 17-22.

Blumerman, L. M. and Vidal, P. M. (2009) 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. Washington, D.C.: U.S. Census Bureau.

Boruch, R. F. (1984) Research on the use of statistical data. *Proceedings of the Social Statistics Section, American Statistical Association*, 52-57.

Bureau of Economic Analysis (2015). State Personal Income press release for March 25, 2015. <http://www.bea.gov/newsreleases/regional/spi/spi_newsrelease.htm>

Bureau of Labor Statistics (2007). *BLS handbook of methods*: Chapter 17, The Consumer Price Index. Washington, D.C., U.S. Department of Labor.

Catalogue of Federal Domestic Assistance (2015). Washington, D.C: Office of Management and Budget. <https://cfda.symplicity.com>

Census Bureau (2012) Press release for May 22, 2012. <http://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html>

Census Bureau (2015). Computing apportionment. <http://www.census.gov/population/apportionment/about/computing.html>

Goldin, C. and Katz, L. F. (2007) The race between education and technology: the evolution of U.S. educational wage differentials, 1890 to 2005. NBER Working Paper No. 12984. Cambridge: National Bureau of Economic Research.

Goldin, C. and Katz, L. F. (2008) *The race between education and technology*. Cambridge: Harvard University Press.

Keyfitz, N. (1979) Information and allocation: two uses of the 1980 census. *The American Statistician*, 33, 45-50.

March, J. G. (1994) *A Primer on Decision Making: How Decisions Happen*. New York: The Free Press.

National Research Council (2003) *Statistical Issues in Allocating Funds by Formula*. Panel on Formula Allocations, Committee on National Statistics, Division on Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

National Research Council (2011) *Change and the 2020 Census: Not Whether But How*. Panel to Review the 2010 Census, T. M. Cook, J. L. Norwood, and D. L. Cork, eds., Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press.

Orszag, P. R. (2009) Federal statistics in the policy process. *The Annals of The American Academy of Political and Social Science* 631, 34-42.

Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M. B. and Sobek, M (2010). *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota.

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

Spencer, B. D. (1980b) Models for error in postcensal population estimates. Pp. 217-228 in National Research Council, *Estimating Population and Income of Small Areas*. Report of Panel on Small-Area Estimates of Population and Income. Committee on National Statistics, Assembly of Behavioral and Social Sciences. Washington, D. C.: The National Academies Press.

Spencer, B. D. (1985) Statistical aspects of equitable apportionment. *Journal of the American Statistical Association*, 80, 815-822.

United States Code (2012). Washington, D.C: Office of the Law Revision Council of the U.S. House of Representatives. Available at <http://law.justia.com/codes/us>.

Weldon Cooper Center for Public Service (2013a) *National and State Population Projections*. Demographic Research Group, University of Virginia, Charlottesville. <http://www.coopercenter.org/demographics/virginia-population-projections>.

Weldon Cooper Center for Public Service (2013b) *State and National Projections Methodology*. Demographic Research Group, University of Virginia, Charlottesville. <http://www.coopercenter.org/sites/default/files/node/13/National_Projections_Methodology.pdf>.