

Opinion of O’CONNOR, J.

SUPREME COURT OF THE UNITED STATES

No. 01–714

UTAH, ET AL., APPELLANTS *v.* DONALD L. EVANS,
SECRETARY OF COMMERCE, ET AL.

ON APPEAL FROM THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF UTAH

[June 20, 2002]

JUSTICE O’CONNOR, concurring in part and dissenting in part.

In the year 2000 census, the Census Bureau used the statistical technique known as “hot-deck imputation” to calculate the state population totals that were used to apportion congressional Representatives. While I agree with the Court’s general description of the imputation process, its conclusion that the appellants have standing to challenge its use, and its conclusion that we otherwise have jurisdiction to consider that challenge, I would find that the Bureau’s use of imputation constituted a form of sampling and thus was prohibited by §195 of the Census Act, 13 U. S. C. §1 *et seq.* Therefore, while I concur in Parts I and II of the majority’s opinion, I respectfully dissent from Part III and have no occasion to decide whether the Constitution prohibits imputation, which the majority addresses in Part IV.

I

To conduct the year 2000 census, the Census Bureau (Bureau) first created a master address file that attempted to list every residential housing unit in the United States. See U. S. Dept. of Commerce, Economics and Statistics Admin., Census 2000 Operational Plan VI (Dec. 2000) (hereinafter Census 2000 Operational Plan). The Bureau

Opinion of O'CONNOR, J.

then conducted a survey of every address on that list, primarily through the use of mail-back questionnaires. See *id.*, at IX.A to IX.E; *ante*, at 1–2. As relevant here, these questionnaires requested the name of each person living at a given address. See Census 2000 Operational Plan V.B.

Because not every address returned a questionnaire, the Bureau had its enumerators attempt to contact nonresponding addresses up to six times by phone or in person in an effort to obtain population information for each address. See Declaration of Howard Hogan ¶73, App. 285 (hereinafter Hogan); Census 2000 Operational Plan IX.G. This was known as “nonresponse followup.” *Ibid.* Also during this follow-up procedure, addresses that appeared vacant were marked as such while addresses determined to be nonexistent were noted for later deletion. See Hogan ¶¶69, 73, App. 283, 285. When all follow-up procedures were completed, the Bureau still lacked population information for approximately 0.4% of the addresses on the master address list because the Bureau had been unable to classify them as either “occupied, vacant, or nonexistent.” *Id.*, at 188. Additionally, the Bureau lacked household size information for approximately 0.2% of addresses that were classified as occupied. See *id.*, at 191.

At this point, the Bureau employed the statistical technique known as “hot-deck imputation.” For each unsuccessfully enumerated address, the Bureau imputed population data by copying corresponding data from a “donor” address. *Ante*, at 2. The donor address was the “‘geographically closest neighbor of the same type (*i.e.*, apartment or single-family dwelling) that did not return a census questionnaire’ by mail.” *Ibid.* (quoting Brief for Appellants 7–8). What this means is that donor addresses were selected only from addresses that had been personally surveyed by the Bureau’s enumerators, primarily through the nonresponse follow-up procedure described

Opinion of O'CONNOR, J.

above. See App. 156. After imputation was completed, every address on the master address list was associated with a household size number that had been determined either by imputation or by enumeration (although that number was zero for addresses ultimately classified as vacant or nonexistent).

The Bureau used the imputation-adjusted data to calculate state population totals. *Ante*, at 3. Because these totals were used to determine the apportionment of congressional Representatives, *ibid.*, we must determine whether the Bureau's use of imputation constituted a form of sampling. If it did, it was prohibited by §195 of the Census Act, 13 U. S. C. §1 *et seq.* See *Department of Commerce v. United States House of Representatives*, 525 U. S. 316, 338 (1999).

II

As initially enacted, §195 provided that “[e]xcept for the determination of population for apportionment purposes, the Secretary [of Commerce] may, where he deems it appropriate, authorize the use of the statistical method known as ‘sampling’ in carrying out the provisions of this title.” 13 U. S. C. §195 (1970 ed.). As relevant here, Congress replaced “may, where he deems it appropriate” with “shall, if he considers it feasible” when it amended §195 in 1976. Pub. L. 94–521, 90 Stat. 2464. In *House of Representatives*, we found that this amended language “might reasonably be read as either permissive or prohibitive with regard to the use of sampling for apportionment purposes.” 525 U. S., at 339. Even so, we held that §195 maintained the prohibition on sampling with respect to apportionment given the “broader context” of “over 200 years during which federal statutes [had] prohibited the use of statistical sampling where apportionment [was] concerned.” *Id.*, at 339–341. With respect to §195, then, the only question is whether “hot-deck imputation” is a form of sampling.

Opinion of O'CONNOR, J.

To answer this question, I begin with the definition of sampling the Bureau provided to Congress in connection with the year 2000 census:

“In our common experience, ‘sampling’ occurs whenever the information on a portion of a population is used to infer information on the population as a whole[,] . . . [although] [a]mong professional statisticians, the term ‘sample’ is reserved for instances when the selection of the smaller population is based on the methodology of their science.” Report to Congress—The Plan for Census 2000, p. 23 (revised and reissued Aug. 1997).

Under this definition, the Bureau’s use of imputation was a form of sampling. The Bureau used a predefined, deterministic method to select a portion of the population and then used that portion of the population to estimate unknown information about the overall population. The Bureau’s imputation process first selected a group of “donor” addresses, one for each address that had not been successfully enumerated. This donor group was a subset of the overall population. Indeed, the donor group was actually a subset *of a subset* of the population because it was selected from only those addresses that had not returned an initial questionnaire but were successfully enumerated through other means. This highlights the Bureau’s reliance on a selected portion of collected data.

Next, the Bureau used the population of the donor group as a direct estimate of the number of people who had not been successfully enumerated. This estimate related to the “population as a whole” because it was an estimate of the overall number of people in the population who had not responded (or had not provided a consistent response, see *ante*, at 1–2) to the Bureau’s survey efforts. See, e.g., F. Yates, *Sampling Methods for Censuses and Surveys* 64, 130 (2d rev. ed. 1953) (describing the use of sampling to

Opinion of O'CONNOR, J.

estimate survey nonresponse); *ante*, at 16 (describing the sampling at issue in *House of Representatives* as one for estimating nonresponse). Because the imputation process selected a portion of the population to estimate the number of people who had not been successfully enumerated, the process constituted a form of sampling.

To counter this conclusion, the majority contends that the Bureau's use of imputation differs from sampling in several different ways. First, the majority argues that the Bureau's use of imputation differs quantitatively from other forms of sampling, suggesting that estimating nonresponse is not sampling when the amount of nonresponse is very small. See *ante*, at 16 (contrasting the use of sampling to estimate a 10% level of nonresponse with the use of imputation to estimate a 0.4% level of nonresponse). But the majority provides no statistical basis to suggest that sampling is confined to "large" estimates. Moreover, we have already decided that the *extent* of the Bureau's reliance on sampling is irrelevant when we held that §195 prohibits sampling for apportionment purposes regardless of whether it is used as a "substitute" for or "supplement" to a traditional enumeration. *House of Representatives, supra*, at 342.

Indeed, the majority more generally acknowledges that the Bureau's reliance on imputation may be distinguishable only in degree from other forms of sampling. See *ante*, at 16 (stating that the sampling at issue in *House of Representatives* differs "in degree if not in kind" from the imputation at issue here). But the majority provides no statistical basis for claiming a difference of degree matters to the question of what constitutes sampling, nor does it explain how a meaningful line between sampling and nonsampling could be drawn on such a basis.

Second, the majority contends that imputation is not sampling because the sample selection method used by the Bureau does not look like "typical," *ante*, at 11, selection

Opinion of O'CONNOR, J.

methods in terms of *when* or *how* the relevant sample is selected. With respect to when a sample is selected, the majority contends that imputation is not sampling because it occurs after all data have been collected. See *ibid.* This presumes that one cannot sample from already-collected data. But sampling from collected data is a recognized form of sampling, even when the collected data result from an attempt to survey the entire population. See Yates, *supra*, at 128.

With respect to *how* a sample is selected, the majority argues that imputation does not look like methods employed “to find a subset that will resemble a whole through the use of artificial, random selection processes.” *Ante*, at 11. But the Bureau’s “nearest neighbor” imputation process is just as artificial as any other form of nonrandom selection, and it is beyond dispute that nonrandom selection methods—including those that produce nonrepresentative samples—may be used for sampling. See, *e.g.*, W. Hendricks, *Mathematical Theory of Sampling* 239–241 (1956); P. Sukhatme, *Sampling Theory of Surveys with Applications* 10 (1954); F. Stephan, *History of the Uses of Modern Sampling Procedures*, 43 *J. Am. Statistical Assn.* 12, 21 (1948) (all indicating that nonrandom selection methods may be used for sampling); see also Yates, *supra*, at 17; R. Jessen, *Statistical Survey Techniques* 16 (1978); W. Deming, *Sample Design in Business Research* 32 (1960) (together indicating that the selection of nonrepresentative or “biased” samples may be permissible, preferred, or even deliberate). Finally, even if random and unbiased selection methods were assumed to be more accurate than other methods of sampling, it would make little sense to construe §195 as prohibiting only the most accurate forms of sampling.

Third, the majority contends that imputation is not sampling because the Bureau never meant to engage in sampling. Along these lines, the majority stresses that the

Opinion of O'CONNOR, J.

Bureau's "overall approach to the counting problem," *ante*, at 11, did not reflect a "deliberate decision," *ante*, at 16, to engage in sampling. Instead, according to the majority, the Bureau's "immediate objective was the filling in of missing data," in an effort to ascertain population information on "individual" units, not "extrapolating the characteristics of the 'donor' units to an entire population." *Ante*, at 11–12.

The majority provides no statistical basis for defining sampling in terms of intent or immediate objectives, however, and to do so would allow the Bureau to engage in any form of sampling so long as it was characterized as something else or appeared to serve some nonsampling objective. But that would render hollow the statutory prohibition on sampling for apportionment purposes. The majority allows this to happen, however, by focusing on the Bureau's "immediate objective" of filling in missing data, which overlooks the fact that the Bureau estimated nonresponse using a selected subset of the population and imputation was simply a means to that end.

Fourth, the majority contends that some definitions of sampling, if viewed broadly, contain no limiting principle and thus might encompass even "the mental process of inference." *Ante*, at 15. But recognizing the Bureau's use of imputation as a form of sampling does not require that sampling be read so broadly. Instead, sampling under §195 can be confined to situations where a selected subset of the population has been directly surveyed on a particular attribute and then that subset is used to estimate population characteristics of that *same* attribute. Such a limitation is neither ill defined nor all encompassing.

Apart from the above arguments, which primarily relate to the statistical characterization of imputation, the majority makes several additional arguments. It contends that Congress' use of the term "sampling" should be read narrowly, limited to what "the Secretary called 'sampling,'

Opinion of O'CONNOR, J.

at the time.” *Ante*, at 14. But the statutory prohibition was not written in terms of what the Secretary viewed as sampling, nor is there any reason to think Congress intended the term “sampling” to be read narrowly as a tight restriction on the Bureau’s ability to gather data for nonapportionment purposes. Rather, the “purpose . . . [was] to permit the utilization of *something less than a complete enumeration*, as implied by the word ‘census,’ . . . except with respect to apportionment.” H. R. Rep. No. 1043, 85th Cong., 1st Sess., 10 (1957) (emphasis added). This suggests “sampling” was meant in a broad rather than narrow sense.

Moreover, because the Bureau’s authorization to use sampling for nonapportionment purposes was simultaneously a prohibition on the use of sampling for apportionment purposes, it makes even less sense to construe “sampling” narrowly *when viewed as a prohibition* given the broader historical context in which §195 marked “the first departure from the requirement that the enumerators collect all census information through personal visits to every household in the Nation.” *House of Representatives*, 525 U. S., at 336. Finally, even if one were willing to assume that the statutory prohibition should not be read to cover statistical techniques the Bureau had used for apportionment purposes prior to 1957, that still would not justify the use of imputation since the Bureau had never before added people to the apportionment count using that process. See Hogan ¶¶39, 41, App. 266–268.

The majority also notes the possibility of *Chevron* deference with respect to the scope of the term “sampling.” *Ante*, at 17 (citing *Chevron U. S. A. Inc. v. Natural Resources Defense Council, Inc.*, 467 U. S. 837, 842–845 (1984)). But the majority ultimately does not rely on this form of deference, *ante*, at 17, nor does it indicate where the Bureau has provided an interpretation of §195 that would have the “force of law” on this issue. See *Christen-*

Opinion of O'CONNOR, J.

sen v. Harris County, 529 U. S. 576, 587 (2000) (explaining that agency “[i]nterpretations . . . which lack the force of law . . . do not warrant *Chevron*-style deference”). Additionally, based on the reasons provided by JUSTICE THOMAS’s partial dissent, I would find that the Bureau’s use of imputation to calculate state population totals for apportionment purposes at least raises a difficult constitutional question. This provides a basis to construe §195 as precluding imputation, regardless of whether the Bureau is entitled to any form of deference. See *Edward J. DeBartolo Corp. v. Florida Gulf Coast Building & Constr. Trades Council*, 485 U. S. 568, 574–575 (1988).

The majority downplays the idea that imputation could be used to manipulate census results, arguing that “manipulation would seem difficult to arrange” in light of the “uncertainties as to what States imputation might favor.” *Ante*, at 16–17. But in every census where imputation would alter the resulting apportionment, the mere decision to impute or not to impute is a source of possible manipulation. While that might be averted if the Bureau were required to use imputation, I do not read the majority’s opinion to demand that. Moreover, in the past, we have given deference to the Secretary’s decision *not* to statistically adjust the census, even when a final decision on that matter was not made until after the census was completed. See *Wisconsin v. City of New York*, 517 U. S. 1, 10–11, 20–24 (1996).

Finally, the majority suggests that imputation is somehow “better” than making no statistical adjustment at all. *Ante*, at 14. But no party has cited a study suggesting that imputation improves distributive accuracy, and the Bureau admits that numeric rather than distributive accuracy “drove the process.” Hogan ¶34, App. 264; see also *id.*, at ¶¶34–35, App. 265 (acknowledging that it may be “impossible to know *a priori* the effects of a particular census operation on distributive accuracy” and that “[i]n

Opinion of O'CONNOR, J.

designing Census 2000, the Census Bureau did not reject operations that would improve numeric accuracy . . . even if these operations might affect distributive accuracy *negatively*” (emphasis added)). I therefore would not assume that imputation necessarily resulted in a “better” census given the recognized importance of distributive accuracy in assessing overall accuracy. See *Wisconsin, supra*, at 20 (stating that “a preference for distributive accuracy (even at the expense of some numerical accuracy) would seem to follow from the constitutional purpose of the census, viz., to determine the apportionment of the Representatives among the States”).

III

Because the Bureau used “hot-deck imputation” to make the same statistical inferences it could not make through more transparent reliance on sampling, I would find that the Bureau’s use of imputation was a form of sampling and thus was prohibited by §195. I therefore respectfully dissent from Part III of the majority’s opinion and have no occasion to decide whether the Constitution prohibits imputation, which the majority addresses in Part IV. For these reasons, I would reverse the judgment of the District Court.