Adjusting the 1990 Census

DAVID A. FREEDMAN

CONSIDERING THE DIFFICULTIES, THE CENSUS BUREAU does a remarkably good job at counting people. In this article, I outline the process and review the two current techniques for evaluating or adjusting the census. In demographic analysis administrative records are used to make independent population estimates, which can be compared to census counts. With capture-recapture methods, data from an independent sample survey are used to estimate population coverage in the census.

If there is a large undercount, these techniques may be accurate enough for adjustment. With a small undercount, it is unlikely that current adjustment methodologies can improve on the census; instead, adjustment could easily degrade the accuracy of the data.

An overview of the census. The census has been taken every 10 years since 1790. It is a sophisticated enterprise whose scale is remarkable. There are about 9,000 permanent staff. Between October 1989 and September 1990, the staff opened 500 field offices, in which they hired and trained 500,000 temporary employees. In spring 1990, a media campaign encouraged people to cooperate with the census and asked, "Were you counted?" in English, Spanish, and several Asian languages. The census was ahead of Coca Cola, with respect to volume of advertising.

The population of the United States in 1990 was about 250 million persons in 100 million households, distributed across 6.5 million "blocks," the smallest units of census geography (*1*). Statistics for larger areas like tracts, cities, or states are obtained by adding up component blocks (*2*).

I will outline census procedures for two types of areas: tape address register (TAR) and pre-list. Most people live in TAR areas, which are mainly urban. The census began there by acquiring and merging commercial address lists. Each address was assigned to a census block (geocoded) on the basis of previously developed computer files. Address lists were checked for completeness by the Post Office and in a pre-canvass by census personnel who walked the streets and looked for omitted housing units. In pre-list areas, typically suburban, the address list was developed by census personnel who canvassed the area, making spot maps to show the location of each residence, which could then be geocoded.

In TAR and pre-list areas, census forms were mailed to each household, filled out by a respondent, and mailed back. Nationally, the mail-back rate was 78% in 1970 (when the mail-out–mail-back procedure was used in TAR areas), 75% in 1980, and 63% in 1990. In 1990, for the first time, computers were used to log forms out and in, and to organize "non-response follow-up." Households that did not return forms were followed up by enumerators: three visits (and three telephone calls) were made, and "last resort" information was obtained from neighbors, building superintendents, and so on. Forms were microfilmed, scanned by optical readers, and checked for incomplete or inconsistent entries. Errors triggered follow-up by telephone or personal visit.

Census data give us a statistical portrait of the United States at 10-year intervals. Geographical detail makes these data unique. However, the counts have more than academic interest: they influence the distribution of power and money. The census is used to apportion Congress as well as local legislatures and to allocate tax money—$40 billion per year in the late 1980s—to 39,000 state and local governments. For these purposes, the geographical distribution of the population matters, rather than counts for the nation as a whole. Indeed, the census is used as a basis for sharing out fixed resources: if one jurisdiction gets more, another must receive less. Adjusting the census is advisable only if the process brings us closer to a true picture of the distribution of the population.

The undercount. A small undercount is thought to remain in the census, and this undercount is unlikely to be uniform. People who move at census time are hard to count; in rural areas, maps and address lists are incomplete. Central cities have heavy concentrations of poor and minority persons who may not cooperate with government agencies.

If the undercount can be estimated with good accuracy, especially at the local level, adjustments can—and should—be made to improve the census. Some statisticians argue that the undercount can be estimated well enough, others are skeptical: a bad adjustment may be worse than nothing (*3, 4*).

Because of its resource implications, the undercount has attracted considerable attention—in the media, the Congress, and the courts. After the 1980 census, New York City (and other jurisdictions) sued the Department of Commerce, seeking to compel an adjustment based on demographic analysis and capture-recapture techniques. The Commerce Department resisted this pressure and was upheld by the court, which found "as a matter of fact that a statistical adjustment of the 1980 census is not feasible" (*5*).

The Department of Commerce also decided not to adjust the 1990 census, and was again sued by New York City and other jurisdictions that were plaintiffs in 1980. The issues in the two cases seem quite similar. One part of the 1990 suit was settled before trial: the Secretary of Commerce agreed to reconsider and make a new decision on adjustment by 15 July 1991.

How can the undercount be estimated? One direct method is to take a sample of small areas, and count them more accurately. Census counts could then be calibrated, by comparison. However, current methods for estimating the undercount do not work that way and are quite indirect.

Demographic analysis makes independent estimates of the national population from administrative records (*6*). The starting point is an accounting identity:

$$\text{Population} = \text{Births} - \text{Deaths} + \text{Immigration} - \text{Emigration}$$

The estimates are made by age, sex, and race (white, black, other) and are compared to census counts. According to demographic analysis, the undercount in 1970 was about 3% nationally; in 1980, it was 1 to 2%; the result for 1990 is likely to be around 2%. Demographic analysis reports the undercount for blacks at about 5 percentage points above whites in 1970 and in 1980.

There are some problems with the accounting identity, however. Data on emigration are incomplete. And there is substantial illegal immigration, which cannot be measured directly. In 1980, for instance, it is estimated that roughly 3 million illegal immigrants were living in the United States; about 2 million are thought to have been counted in the census (*7*).

Evidence on differential undercounts depends on racial classifications, which may be problematic; and procedures vary widely from one data collection system to another. For the census, race of all household members is reported by the person who fills out the form. On death certificates, race of decedent is often determined by the undertaker. Birth certificates show the race of the mother and

The author is professor of statistics, University of California, Berkeley, CA 94720.

(usually) the race of father; procedures for ascertaining race differ from hospital to hospital. A computer algorithm is used to determine race of infant from race of parents: changing the algorithm would reduce estimated undercount rates for young black children by 2 to 5 percentage points (*8*).

Coverage of vital statistics is another problem, forcing demographic analysis to use different techniques for different age groups, with further variations by race and sex. In the period 1935 to 1960, the coverage of the birth certificate system was far from complete, especially for blacks. To estimate undercount rates for persons between the ages of 30 and 55 in 1990, birth certificate data must be adjusted for underregistration; and the adjustment is based on census data. In short, before birth certificate data can be used to adjust the census, the census must be used to adjust the birth certificates (*9*).

Prior to 1935, many states did not have birth certificate data at all; the further back in time, the less complete is the system. This makes it harder to estimate the population aged 55 and older. For the period 1925 to 1935, a set of data for whites was created by Whelpton (*10*). For black females, another set of imputed data is used, and the number of black males is estimated from expected sex ratios; "imputed data" are themselves estimated from other sources. There is yet another set of imputations for persons of other races (Asians, for example). Finally, persons born before 1925 are over the age of 65 by 1990; demographic analysis estimates the number of such persons starting from Medicare records (adjusted for underenrollment).

Detailed results from demographic analysis—especially estimated undercount rates by age for black males—suggest that there are problems in the procedure. In 1980, for instance, demographic analysis produced an estimated undercount of nearly 0 for the late teens, rising to 18% for ages 40 to 44, then dropping again. (Fig. 1) A more plausible pattern is just the reverse: high undercount rates for teenagers and low rates for the middle-aged.

Again, according to demographic analysis (Fig. 1), the age group with the highest undercount rate was 20 to 24 in 1960, 30 to 34 in 1970, and 40 to 44 in 1980. Now there may be a cohort of hard-to-count people who were 20 to 24 in 1960, and who remained hard to count as they aged through the census of 1970 and 1980. A more plausible explanation is statistical artifact—including overadjustment of births in the 1930s (*8*, *11*). The validity of demographic analysis depends on a series of complicated adjustments to a variety of administrative statistics. The errors may be small, but so is the undercount.

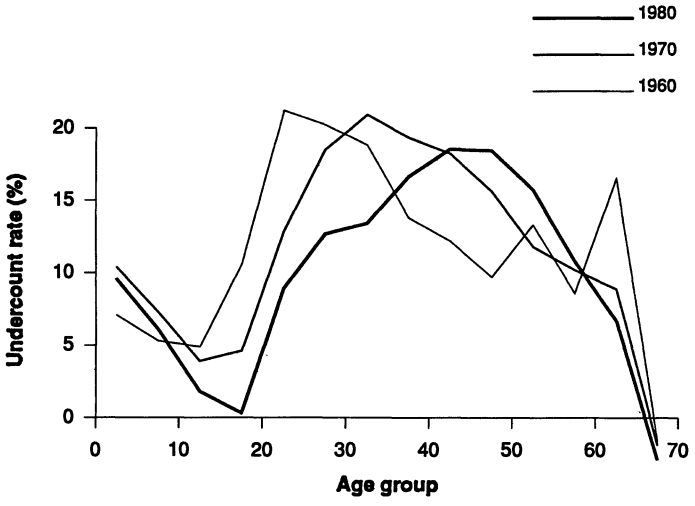

Fig. 1. Results from demographic analysis: estimated undercount rates for black males, by age group, in the census of 1960, 1970, and 1980.

One limitation of demographic analysis is widely recognized. The estimate are national rather than local because data are lacking on internal migration. Of course, national undercount rates could be applied to small areas, a process called "synthetic estimation." For instance, if the undercount rate for black males 40 to 44 years of age is estimated as 18% for the country as a whole, the number of such persons in every block could be increased by the "adjustment factor" $1/(1 - 0.18) = 1.22$. This method is of doubtful utility for making small-area estimates because undercount rates must vary by substantial amounts from place to place.

The dual system estimator (DSE). To estimate the undercount in geographic detail, proponents of adjustment suggest using the DSE on capture-recapture data. This is easiest to explain in a hypothetical example: estimating the number of fish in a lake. You catch 1000 fish, tag them, and throw them back (the "capture"). Then catch another 100 fish ("recapture"). Say that 90 of the recaptured fish are tagged, indicating that $90/100 = 90\%$ of the fish in the lake are tagged. The total population, tagged and untagged, would be estimated as $1000/(0.90) \approx 1111$. The idea is appealing, but the practical difficulties are serious. Recapture has to be done at random (equal probability of recapture for all fish), tags have to stay on the fish, and the population of the lake has to stay the same between capture and recapture.

For the DSE, "capture" means being counted in the census. Recapture is by a special sample survey, the PES (post-enumeration survey). The PES is based on a stratified sample of about 5,000 blocks, with 150,000 households, and 400,000 people (*12*). In principle, it may be easy to see whether a fish is tagged. With people, the problem is more complicated. Records from the PES have to be matched against records from the census, to determine whether a person in the PES was captured in the census.

About 70% of the matching is done by computer; the rest, by clerks. The census does not collect unique identifiers like social security numbers (*13*); the matching algorithm uses name, address, age, sex, race, and ethnicity. However, some of the data are inaccurate, on the PES side as well as the census side. There are variations in spelling, and some persons give fictitious names. Demographic characteristics (even sex) sometimes appear to change from one interview to another.

For many of the records, match status cannot be determined on the basis of the information obtained in the census and the initial PES interview. Such cases are re-interviewed ("sent to follow-up"). Despite this effort, some cases remain "unresolved," and statistical models are used to impute match status. The validity of these models is questionable. In 1980, roughly 8% of the PES cases were imputed. In 1990, the problem is expected to be less severe; however, a small percentage of unresolved cases spells trouble when the undercount rate is small (*14*).

Even with complete data, record matching is a complex and error-prone process. The typical mistake is a false non-match—a failure to link two records that really refer to the same person. That reduces the denominator of the DSE and inflates the estimated undercount (see Eq. 1). Each percentage point of false non-matches, no matter what the source, inflates the estimated undercount by nearly 1 percentage point.

The census is taken in the spring and counts people at their usual place of residence on census day (1 April). The PES is done in the summer, with follow-up interviews in the fall and winter. About 20% of the population moves every year, late spring and early summer being peak times. In some areas, roughly one third of the population has moved between census day and PES interviewing; 5 to 10% may be more typical. PES interviewers must get the right census-day address for respondents. This is a critical step, and one that may be quite hard to do. Failure in tracking people back to their

census-day address is likely to create false non-matches, with a serious impact on estimated undercounts.

The DSE must deal with erroneous enumerations in the census, as well as double-counting and fabrication (*15*). To handle these problems, and limit the search area for matches to manageable scale, the DSE uses the "P-sample" and "E-sample". The P-sample consists of all the people (in the sample blocks) found by the PES interviewers. The E-sample consists of census records for these same blocks.

An attempt is made to match persons in the two samples: a match validates both the census and the PES records (*16*). Persons in the P-sample but not the E-sample are considered to have been missed by the census. (Of course, they may also have been counted in error by the PES.) Persons who turn up in the E-sample, but who are not found in the P-sample, represent potential erroneous enumerations. (In the alternative, they were correctly enumerated by the census, but missed by the PES; fieldwork is needed to resolve the status of these persons.) Finally, there are persons in neither sample; their existence cannot be demonstrated directly, but their number is estimated by the capture-recapture model. The resulting classification is shown in Fig. 2.

The formula for the dual system estimate (*DSE*) is

$$DSE = \frac{Cen - EE - II}{M/N} \quad (1)$$

where *Cen* is the census count, *EE* is the estimated number of erroneous enumerations, and *II* is the number of imputations and unmatchable persons in the census. (Some persons are imputed into the census, or counted without enough detail for matching; such persons have to be subtracted out of the census count.) *M* is the estimated total number of matches obtained by weighting up sample matches and *N* is the estimated population obtained by weighting up P-sample block counts (*17*). The object of the PES matching operation is to provide estimates of *M* and *EE* (Fig. 2); *Cen* and *II* come from census records.

"Weighting up" may be an unfamiliar term, but the idea is easy: if you sample one block in 1000, say, then each sample block counts for 1000 blocks in the country—and therefore gets a "sample weight" of 1000. Population subtotals (for matches, say) are estimated by adding up sample weights for corresponding people. In practice, there are different sample weights in different strata, and adjustments are made for nonresponse in the PES.

Intuitively, the "match rate" *M/N* in the denominator of the *DSE* estimates the fraction of the population counted by the census. The ratio *DSE/Cen* is an "adjustment factor": it adjusts the census count to the dual system estimate.

Post-strata and smoothing. Actually, many adjustment factors are computed. Different kinds of people are likely to have unequal probabilities of responding to the PES, violating the randomness assumption for recapture. As a partial solution, the PES sample is "post-stratified" by six age groups, by sex, race, ethnicity, and housing "tenure" (owner or renter). There are about 1400 post-strata (*18*). The DSE and corresponding "raw" adjustment factor are computed separately for each one. On average, we expect 400,000/1,400 ≈ 300 sample people in each post-stratum, and only a few gross omissions and erroneous enumerations. Resulting estimates would be quite unstable, due to sampling error.

		E-sample (census)	
		In	Out
P-sample (PES)	In	Matches	Gross omissions
	Out	Erroneous enumerations	Not found

Fig. 2. Capture-recapture: classification of records in the P- and E- samples.

To reduce sampling error, statistical smoothing techniques are used to combine results from similar post-strata. More technically, regression models are used to predict adjustment factors from some of the variables that define strata and post-strata, and predicted factors are averaged with raw factors. However, the models do not fix bias due to matching error, for example. Furthermore, statistics that measure reliability of smoothed results can be quite misleading unless the models are validated (*4*).

Block-level counts. Once adjustment factors are computed, the proposal is to correct block-level counts by the synthetic method. For example, take one post-stratum: males who are black or Hispanic, 45 to 64 years of age, living in central cities in New England. Suppose the DSE for this post-stratum is 10% over the census count, so that the adjustment factor is 1.1. Now suppose some central city block in New England has a census count of ten black or Hispanic males aged 45 to 64. According to the DSE, there are $1.1 \times 10 = 11$ such persons in the block (*19*). One of the ten real census records would be chosen at random and copied. The resulting fictitious person would be added to a special "adjustment category" in the block and come into all census of population tables for areas that include the block. This scenario would be repeated for every block, increasing the post-stratum count by 10%. Block-to-block variability would be ignored.

Some post-strata will have adjustment factors below 1.00, corresponding to apparent overcounts (*20*). Suppose a central city block in New England has 20 white males who are 45 to 64 years old, by census count; and the adjustment factor is 0.95. According to the DSE, there are only $0.95 \times 20 = 19$ such people in the block. One of the census records would be selected at random and a corresponding "negative person" put into that block's special adjustment category. This process would be applied uniformly to all blocks, reducing the post-stratum count by 5%. Real people wind up being subtracted from the census tables.

Variability is a major obstacle to adjustment. Indeed, undercount rates differ from one geographical area to another, and from one demographic group to another. That is why synthetic estimates for small areas, based on demographic analysis, have not been widely accepted. However, adjustment by the DSE is unsatisfactory for the same reason. For example, one post-stratum consists of Hispanics—cross-classified by age, sex, and housing tenure—in central cities in the Pacific Division (California, Washington, Oregon, Alaska, and Hawaii). In round numbers, the 1990 population of the Pacific Division is about 40 million with 8 million Hispanics, 5 million of the latter being in southern California.

Consider an adjustment for Stockton, a city of about 200,000 people in California's Central Valley, a 4-hour drive north of Los Angeles. The Hispanic population is about 50,000; there can be at most a few dozen Hispanics from Stockton in the PES, and a handful of gross omissions or erroneous enumerations. No stable estimates could be developed from a sample that small. Instead, estimates for Stockton would be based on the adjustment factor for the whole post-stratum, the numbers being driven by PES data from southern California. The basic assumption: undercount rates for Hispanics are the same in Stockton as in Los Angeles. There is no empirical evidence to support this assumption. And there is a similar problem for non-Hispanics. Indeed, adjustment factors for non-Hispanics in Stockton are driven by PES data on non-Hispanics in the whole Pacific Division. Apparently, Stockton's non-Hispanics are supposed to be like their counterparts in the north, while its Hispanics are taken to be southern. Stockton is the rule not the

exception. There are 39,000 state and local government areas to adjust; and only 5,000 sample blocks with PES data. Most jurisdictions would be adjusted on the basis of data from elsewhere—and the synthetic assumptions.

Quantitative evaluation of the 1990 PES waits on publication of the data. However, the methodology was tried in 1986 in the Los Angeles Test of Adjustment Related Operations (TARO) and in the 1988 Dress Rehearsal (*21*, *22*). In TARO, special follow-up research studies were done, and some information is available on the magnitude of the errors. The data can be interpreted in many ways. Still, more than half the estimated undercount seems to be due to PES errors. If so adjustment by the DSE would have moved the TARO census from an undercount of perhaps 4% to an overcount of 5% or more.

For example, about 5% of the data from the E-sample were imputed, with significant understatement of erroneous enumerations. Alternative and more plausible imputations reduce the estimated undercount by 2 percentage points. Another 3 percentage points of undercount came from respondents who gave bad address information at the PES interview. Many of these respondents were in fact movers but said otherwise, creating false non-matches and inflating the estimated undercount. Indeed, about half the group seems to have moved into the test site after the test census: such respondents are "out of scope."

Proponents of adjustment reply by estimating how many movers would have been correctly classified had they given correct addresses and been in scope, or how many fabricated interviews would have matched to the test census had they been real (*22*). Such estimates seem fanciful.

Conclusion. The census does a remarkably good job at counting people—given the difficulties in large-scale statistical work. Still, an undercount may be expected. Of the two current adjustment methodologies, demographic analysis must cope with small errors and inconsistencies in a variety of administrative data systems; its estimates are made only at the national level. The dual system estimator faces problems created by incorrect or missing data—especially for movers—which increase the error rate in record matching and inflate estimated undercounts. Variation in undercount rates from place to place is a reality faced by both methodologies.

There is little hard evidence to show that current adjustment methodologies would improve the accuracy of the census, and much can go wrong. In short, the present state of the art probably cannot support adjustment of the 1990 census.

REFERENCES AND NOTES

1. The census attempts to count "permanent residents" as of 1 April. This definition excludes foreigners temporarily resident in the United States; it includes aliens—both legal and illegal—living in the United States; it excludes U.S. citizens abroad, unless they are in the military or working for the U.S. government. At the margins, the definition is somewhat arbitrary and unclear.
2. Many details are omitted in this brief overview. For example, a small percentage of blocks cross administrative boundaries (for example, city lines) and are "split" by survey after the enumeration. List-enumerate methods for counting rural areas are not discussed at all.
3. C. F. Citro and M. L. Cohen, Eds., *The Bicentennial Census: New Directions for Methodology in 1990* (National Academy Press, Washington, DC, 1985); evaluation is discussed in chapter 4 and adjustment in chapter 7; E. P. Ericksen and J. B. Kadane, *J. Am. Stat. Assoc.* **80**, 98 (1985); E. P. Ericksen, J. B. Kadane, J. W. Tukey, *ibid.* **84**, 927 (1989); R. J. Beran *et al.*, statement on census adjustment presented to the U.S. House of Representatives subcommittee on census and population, 3 March 1988; E. P. Ericksen, *Stats 4* (Fall 1990), p. 4.
4. D. A. Freedman and W. C. Navidi, *Stat. Sci.* **1**, 1 (1986).
5. The 1980 case was decided in 1987: *Cuomo et al.* v. *Baldrige et al.* 674 *F. Supp.* 1089 (Southern District of New York).
6. R. E. Fay, J. S. Passel, J. G. Robinson, C. D. Cowan, *The Coverage of the Population in the 1980 Census* (Government Printing Office, Washington, DC, 1988). In 1980, demographic analysis series DA-2 estimated the national undercount rate for whites and other races as 0.7 of 1%; the rate for blacks as 5.9%; blacks were about 12% of the population (Table A.80.3).
7. R. Warren and J. S. Passel, *Demography* **24**, 375 (1987).
8. J. S. Passel, "Demographic analysis: A report on its utility for adjusting the 1990 census" (Technical Report, Urban Institute, Washington, DC, 1990).
9. The adjustment is done by capture-recapture. Recapture was in the census of 1940 or 1950; capture, in a sample of birth certificates covering several months before census day. A more recent test (1964–1968) was based on the Current Population Survey [See *Vital Statistics of the United States*, *1960, vol. 1, Natality* (U.S. Department of Health, Education, and Welfare, Washington, DC)].
10. P. E. Whelpton, *Vital Stat., Spec. Rep., Sel. Stud.* **33** (no. 8) (1950).
11. J. G. Robinson, P. Das Gupta, B. Ahmed, "A case study in the investigation of errors in estimates of coverage based on demographic analysis: black adults aged 35 to 54 in 1980." (Technical Report, Bureau of the Census, Washington DC, 1990); "Evaluating the quality of estimates of coverage based on demographic analysis" (Technical Report, Bureau of the Census, Washington, DC, 1990).
12 Strata are defined on the basis of geography and demographic characteristics from the 1980 census.
13. B. Bryant, *Science*, in press.
14. In 1980, six different sets of imputations were considered; and there were two "P-samples" (April and August Current Population Survey) for recapture. National undercount estimates ranged from −1.0% to 2.1% (*6*, table 7.1).
15. An "erroneous enumeration" is a person counted in error, for instance, a baby born on 2 April. Some people are double-counted, at their regular homes and at vacation homes, for example. A "fabrication" is made up, or "curbstoned" by an enumerator. Failure to take these problems into account could lead to an overestimate of the population. Similar problems in PES data are not dealt with; such errors, as well as matching errors and bad address information all tend to bias the estimated population upward. "Correlation bias" (dependence between capture and recapture or heterogeneity in recapture probabilities) may be a partial offset. So far, attempts to quantify this bias have been quite speculative.
16. Census records are searched for a match in the block where the PES person is found and surrounding blocks. (For movers, the search starts in the block where the person is reported to have lived on census day.) To avoid bias, levels of effort on the E-side (to identify duplications or erroneous enumerations, for instance) should correspond to those on the P-side, and search areas should be comparable. However, virtually all the potential erroneous enumerations are sent to follow-up; many gross omissions do not go to follow-up.
17. Movers in the PES are weighted up according to the block in which they are found, but contribute to the estimated population and number of matches according to reported place of residence on 1 April 1990.
18. Demographic categories are not fully crossed with geographical stratification.
19. This process usually leads to fractional numbers of people, which are rounded.
20. For data on undercounts in 1980, see Fay *et al.* (*6*), tables 7.5, D1-2. PEP 2-9 was the particular version of the DES favored by plaintiffs' experts in 1980. In 1980, there was an overcount in 7 out of 17 age groups for white-and-other-race males, according to PEP 2-9. The pattern for women was similar. Furthermore, overcounts were estimated in 11 out of 50 states.
21. G. Diffendahl, *Surv. Methodol.* **14**, 71 (1988); N. Schenker, *ibid.*, p. 87; C. D. Jones, "Evaluation of TARO" (Technical Report, Bureau of the Census, Washington, DC, 1988); M. M. Mulry and K. K. West, "Evaluation follow-up for the 1988 post-enumeration survey" (Technical Report, Bureau of the Census, Washington, DC, 1990); M. Mulry and B. Spencer, "Total error in PES estimates of population: The Dress Rehearsal Census of 1988" (Technical Report, Bureau of the Census Washington, DC, 1990).
22. H. Hogan and K. Wolter, *Surv. Methodol.* **14**, 99 (1988).

23. I testified for the Department of Commerce in the 1980 census adjustment case and am working for the Department of Justice on the 1990 case. Support from the Miller Foundation is gratefully acknowledged. L. Bazel, L. Breiman, P. Diaconis, E. Hoag, S. Klein, W. Kruskal, P. Meier, R. Purves, and A. Tversky all made useful comments. So did P. Bounpane, G. Robinson, C. Jones, and J. Thompson. J. Passel deserves thanks.