Technical Report No. 455, April 20, 1996
To appear in *Evaluation Review*

PLANNING FOR THE CENSUS IN THE YEAR 2000

David A. Freedman
Kenneth W. Wachter
University of California, Berkeley

ABSTRACT

Considering the difficulties, the Census Bureau does a remarkably good job at counting people. There are two current techniques for evaluating or adjusting the census: (i) demographic analysis uses administrative records to make independent population estimates, which can be compared to census counts; (ii) data from an independent sample survey can be used to estimate population coverage. 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. This paper reviews plans for Census 2000, including proposals for adjustment, in the light of past experience.

1. INTRODUCTION

The census has been taken every ten years since 1790. It is a sophisticated enterprise whose scale is remarkable. In round numbers, there are 10,000 permanent staff. Between October 1999 and September 2000, the staff will open 500 field offices, in which they will hire and train 300,000 temporary employees. In spring 2000, a media campaign will encourage people to cooperate with the census, with community outreach efforts targeted at hard-to-count groups. With respect to volume of advertising during this critical period, the census may be ahead of Coca Cola.

According to Bureau planning documents, the population of the United States in 2000 will be 275 million persons in 119 million housing units, distributed across 7 million "blocks," the smallest units of census geography.

(These blocks differ considerably in size.) Statistics for larger areas like tracts, cities, or states are obtained by adding up component blocks. From the perspective of a census-taker, there are three types of areas to consider. In "city delivery areas," the Bureau will develop a Master Address File (MAF); questionnaires are mailed to each address in the file; about 60% of these questionnaires will be filled out and returned by the respondents. Then, "Non-Response Follow-Up" procedures go into effect: for instance, census enumerators may go out and attempt to contact non-responding households; more about this, below. City delivery areas, which are high-density urban areas where good address lists are relatively easy to develop, comprise 100 million housing units out of the 119 million total.

"Update/leave" areas, comprising about 18 million households, are mainly suburban and have lower population densities; address lists are more difficult to construct. In such areas, the Bureau delivers the census questionnaire while it is updating the Master Address File; beyond that, procedures are similar to those in the city delivery areas. In "update/enumerate" areas, the Bureau tries to interview respondents as it updates the MAF: these areas are mainly rural, and post-office addresses are poorly-defined, so address lists are quite difficult to construct. Perhaps a million housing units fall into such areas.

The census operates on a ten-year cycle; planning the next census starts almost immediately after the current census is taken. The process is evolutionary, and new techniques are tested on a relatively small scale at intermediate times. For instance, there was a test census in 1995 in Oakland (among other sites); results will be discussed below. Census data give us a statistical portrait of the United States at ten-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—$60 billion in 1989, for example—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 counts 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 be made to improve the census. In 1980 and 1990, some statisticians argued that the undercount could have been estimated well enough for adjustment purposes. Others were skeptical: a bad adjustment may be worse than nothing. In 1980, the government decided not to adjust the census, and was upheld by the courts (Freedman and Navidi, 1992). In 1990, the government decided again not to adjust, and was upheld by the U. S. Supreme Court; the litigation will be discussed in section 5, below.

CENSUS 2000

For 2000, the Bureau has elected to "reengineer" the census. We discuss four of the main innovations: (i) sample-based non-response follow-up; (ii) Integrated Coverage Measurement (ICM) and the "One-Number Census"; (iii) use of administrative records; (iv) new questionnaires.

Sample-based non-response follow-up. In 1990, the plan was to follow up *all* non-responding households; in practice, virtually all the non-respondents were reached. (Some were imputed into the census, and some remained missing; missing data, errors in imputations, and incomplete address lists all contribute to the undercount.) In the year 2000, the Bureau thinks that it may be impractical to do 100% follow-up, because there is likely to be a severe shortage of part-time workers. The Bureau is now planning to follow up only a sample of non-respondents. More particularly, it will do complete follow-up in each county until it has 90% of the housing units; then it will sample non-responding housing units at the rate of 1 in 10. Sampling will be done within each block. The rationale for these design decisions is not completely apparent. Different cutoffs could be used in each county, and blocks could be sampled rather than housing units within block.

Of greater concern, a sample-based non-response follow-up is not likely to yield better results than 100% follow-up. In principle, using highly-trained personnel could help. In practice, sampling seems to be a more complex operation than taking a census, and coverage may suffer in consequence. For instance, it is estimated that the Current Population Survey only reaches 70% to 99% of its target population; these "coverage ratios" are fairly stable over time, but depend strongly on age, sex, and race. A second example: the Post Enumeration Survey in 1990 had at best the same coverage as the Census. (On the first point, see, for instance, U. S. Bureau of the Census, 1993a, page D-4; on the second point, we are reporting our own analysis of the data, and discussions with Bureau personnel.) In short, we are concerned that sample-based non-response follow-up will aggravate coverage problems in the census.

Integrated Coverage Measurement (ICM) and the One-Number Census. As part of its plans for the Year 2000, the Bureau proposes "Integrated Coverage Measurement (ICM)" for adjusting Census 2000. Among other things, the proposal is *not* to release census numbers and adjusted numbers, but to give out only the final set of results for each area—hence the term, "One-Number Census." ICM and sample-based non-response follow-up are often presented as indivisible; for instance, see (Belin and Rolph, 1994, p.502) or (Steffey and Bradburn, 1994, pp.21–22). To us, however, these programs look quite separable. ICM protects against failures in census coverage. From our perspective, ICM is expensive, inefficient, and itself quite error-prone, because it is so complex. Thus, other kinds of protection are necessary, and we propose an alternative—"raking" to independently-estimated national totals by age, sex, race, and ethnicity; such estimates can be derived by demographic analysis. (There will be more discussion of ICM, demographic analysis, and raking, below.)

Adminstrative records. The Bureau plans to use administrative records, like Internal Revenue Service tax forms or Social Security data. These records will be used in several distinct ways: (i) to correct the Master Address File; (ii) to correct the census roster for each housing unit in the MAF; (iii) to impute data for the first percentage point or so of non-responding housing units. At this time, details on (ii) or (iii) are not available; apparently, however, the activity will be quite model-dependent. Moreover, use of administrative records may create much uneasiness: respondents will need to be told that the Census Bureau gets data from the IRS or the Social Security Administration, but does not turn over data to those agencies—which may seem implausible. The likely gains in accuracy, or savings in cost, pale by comparison. We recommend that the Bureau sharply curtail its plans for using administrative records in 2000.

New questionnaires. One key to getting public cooperation is making the forms intelligible. Questionnaire design is critical. It is also important to make forms available in a language that respondents can understand. On these two fronts, we think real progress has been made since 1990.

2. TECHNICAL BACKGROUND

This section of the paper presents some technical background on the following topics: (i) demographic analysis; (ii) Integrated Coverage Measurement—which has two versions, CensusPlus and Dual System Estimation; (iii) smoothing, which may be used in an effort to reduce sampling error; (iv) evaluations of the census and adjustments, and (v) Loss Function Analysis, which attempts to provide overall measures of error in the census and the adjustment. In section 3, we comment on the National Academy reports; section 4 reports on other relevant literature. In section 5, we discuss the recent Supreme Court decision on the 1990 census, and section 6 summarizes the discussion. Readers can by-pass the technical details by going directly to section 5.

DEMOGRAPHIC ANALYSIS AND RAKING

Demographic analysis estimates the U. S. population using birth certificates, death certificates, and other administrative record systems. The estimates are made for national demographic groups—defined by age, sex, and race (black, white, other); it seems likely that the estimates could be modified to account for ethnicity as well. According to demographic analysis, the undercount in 1970 was about 3% nationally; in 1980, it was 1% to 2%; the result for 1990 was similar. Demographic analysis reports the undercount for blacks at about 5 percentage points above whites, in all three censuses.

Census figures could be scaled up to match the demographic analysis subtotals for national subgroups of the population; that would be a simple alternative to ICM. Scaling-up is also called "raking." The people in a demographic group who are thought to be missing from the census are added back, in proportion to the ones who are counted—state by state, block by block.

Demographic analysis starts from an accounting identity:

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

There are some problems with the "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. Allowances can be made for illegals, but these are (necessarily) somewhat speculative.

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 (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: proposed changes to the algorithm would reduce estimated undercount rates for young black children by 2 to 5 percentage points (Passel, 1990).

Gaps in the vital statistics are 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 40 and 65 in 2000, birth certificate data must be adjusted for under-registration; and the adjustment is based on previous census data. In short, before birth certificate data can be used to adjust the census, census data must be used to adjust the birth certificates.

Prior to 1935, many states did not have birth certificate data at all: and the further back in time, the less complete is the system. This makes it harder to estimate the population aged 65 and over. Vital statistics were estimated for the period prior to 1935; these estimates were used in 1960, 1970, and 1980. Persons born before 1935 are over the age of 65 by 2000; demographic analysis estimates the number of such persons starting from Medicare records (adjusted for under-enrollment).

Demographic analysis has many flaws, and raking to demographic analysis totals should not be oversold. The technique makes national totals match independent estimates, without markedly degrading the quality of state or small-area totals; improvement is quite possible. The ICM is complicated enough that it could go badly wrong, in ways that will not be noticed until too late. By comparison, raking is much simpler, so there is little likelihood of any major disasters.

INTEGRATED COVERAGE MEASUREMENT

Two versions of ICM are being considered for 2000. One is called "CensusPlus," the other is the "Dual System Estimator," or DSE. With either version, the population is divided into "post strata." These are demographic subgroups, for instance, black male renters age 30–49 in rural areas. A sample of blocks (minimal units of census geography) is drawn, and a special survey is done to estimate the "true" counts for each post stratum, using data collected on the sample blocks. CensusPlus and the DSE use slightly different survey procedures, and then somewhat different estimation procedures. In 1990, the proposed adjustment was based on the DSE, and the special survey for adjustment purposes was called the "Post Enumeration Survey"; the PES is discussed in some detail, below. The post strata for Census 2000 remain to be defined; choice of post strata had a substantial impact on the 1990 results (Woltman, 1991).

ICM fieldwork is done in the sample of blocks just discussed; with either version of ICM, an independent listing is made of the housing units in those blocks. We now give details on the CensusPlus version of ICM (also see Steffey and Bradburn, 1993, pp.109ff); details on the DSE will follow.

CensusPlus. For the CensusPlus version of ICM, the independent listing of housing units in the sample blocks is reconciled with the Census "Master Address File," to produce an enriched listing. Interviews are then conducted shortly after census follow-up ends. First, an independent household roster is constructed; second, an attempt is made to reconcile this roster with the previously-obtained census roster; the census roster is enriched from administrative records, as noted earlier.

Interviewing is computer-assisted. The interviewers carry laptops into respondent households; the various rosters—and questions about them—pop up onto the screen. While the interviewers ask the questions and collect the information, the final reconciliation is done on the computer at data processing centers.

The end result is a set of corrected counts, by post strata, in the sample blocks. These counts are projected from the sample to the universe of all blocks. That gives an estimated "true count" in each post stratum. The ratio of the estimated true count to the census count is an adjustment factor, which can be used to adjust blocks not in the sample, post stratum by post stratum. This is where the "homogeneity assumption" or "synthetic assumption" comes into play—undercount rates are taken as constant within post strata across geography.

The DSE. We turn now to the Dual System Estimator—the second version of ICM. Many details for 2000 remain to be defined, but a discussion of the 1990 PES illustrates the principles. Estimates of coverage are made using capture-recapture techniques. Capture is in the census; recapture is in a sample survey conducted after the census—the Post Enumeration Survey or PES. The 1990 PES was based on a stratified sample of about 12,000 blocks, with 150,000 households and 400,000 people.

The PES estimates the number of people missed by the census, or "gross omissions," and the number who are counted in error—the "erroneous enumerations". Erroneous enumerations include babies who were counted in the census but were born just after census day, fabrications, duplicates, people counted at the wrong address, and so forth. To a first approximation, the net undercount is estimated as the difference

$$\text{gross omissions} - \text{erroneous enumerations};$$

the exact formula is given below.

To make its estimates, the PES uses a "P-sample" and "E-sample." The P-sample consists of all the people (in the 12,000 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.

Persons in the P-sample but not the E-sample (or surrounding blocks) are considered to have been missed by the census. There is one major exception: if a P-sample person has moved into a sample block after the census, matching to the census is attempted at the previous address, which typically falls outside the sample blocks. Persons who turn up in the E-sample and not the P-sample represent potential erroneous enumerations. However, they may have been correctly enumerated by the census and missed by the PES; field work is done to resolve the status of these persons, and of certain potential gross omissions. Finally, there are persons in neither sample: their existence cannot be demonstrated using census and PES data, but their number is estimated by the capture-recapture model. For 2000, the search area for matches remains to be decided; the evidence in 1990 suggests that the search area matters (Breiman, 1994, p.473).

The population in 1990 was divided up into 1,392 post strata, demographic subgroups that were considered by the Census Bureau to be relatively homogeneous for undercounts. For example, one post stratum comprised Hispanic males age 0–9 living in owner-occupied housing in central cities in the Pacific Division (California, Oregon, Washington, Alaska, Hawaii). PES data were used to compute the "raw" dual system estimator, or DSE, for the population count in each post stratum. Then, a "raw adjustment factor" was computed for each post stratum:

$$\text{raw adjustment factor} = \frac{\text{raw DSE}}{\text{Cen}}.$$

The ratio (raw DSE)/Cen is called a "raw adjustment factor" because it adjusts the census count Cen in a post stratum to the raw dual system estimate for that post stratum. In 1990, these raw factors were "smoothed" using an empirical Bayes sort of algorithm (see below).

The formula for the raw dual system estimator, which is used to estimate the population in each post stratum, can be stated as follows:

$$\text{raw DSE} \;=\; \frac{\text{Cen} - II}{M/N_p} \times \left[1 - \frac{EE}{N_e}\right].$$

In this formula, the raw DSE is the dual system estimate of population; Cen is the census count; II is the number of persons imputed into the census; M is the estimated total number of matches obtained by weighting up sample matches; N_p is the estimated population obtained by weighting up P-sample block counts; EE is the estimated number of erroneous enumerations; and N_e is the estimated population obtained by weighting up E-sample block counts.

The "match rate" M/N_p appears in the denominator of the DSE. Intuitively, the complement of the match rate estimates the gross omissions rate in the census. Likewise, EE/N_e estimates the rate of erroneous enumerations in the census. The object of the PES is to estimate these fractions. Cen and II come from census records. Some persons are counted in the census without enough detail for matching; such persons are classified as erroneous enumerations, as are persons who have been counted more than once. Further details are given by Fay (1992), Hogan (1993) or Freedman et al. (1994, pp.248ff, p.265).

Comparisons with 1990. Comparison of ICM with 1990 may be useful. ICM will be done shortly after census follow-up, reducing problems with movers. However, there may be a compensating problem, if the ICM changes response rates to the census (compare Steffey and Bradburn, 1993, p.115): basically, it will be harder to extrapolate from the sample to the population, if the sample has become different as a result of the ICM treatment. ICM will use a much larger sample, which will reduce sampling error, and should help with heterogeneity. There may be offsetting errors in the fieldwork, if bigger sample surveys are harder to control administratively and process statistically—especially given very tight constraints on time and money.

The Oakland test census. As noted above, the Bureau elected to sample households within blocks, rather than whole blocks, for non-response follow-up. This would create some tension with ICM—how would they take an independent sample, or do the matching and reconciliation? To avoid such difficulties, non-response follow-up in ICM blocks will be done on a 100% basis. The Bureau believes that treating ICM blocks differently from non-ICM blocks with respect to follow-up will not create any differences in coverage. Their evidence is a study done as part of the 1995 Oakland test census. The data are not yet available. However, it is somewhat questionable whether the test has enough power to detect important differences in coverage, within post strata between treatments. (An order-of-magnitude calculation suggests there is enough power to detect a 5 percentage point difference, but not a 2 percentage point difference.)

Another finding in the Oakland test census is of some concern. It appears that many CensusPlus interviewers did not follow instructions. It also appears that in a substantial percentage of cases, the interviewers were not given the census roster in time, so fieldwork on reconciliation could not be done. The ability of the interviewers to reconcile the rosters was checked by a computerized matching algorithm; it is not obvious why the algorithm is thought to be giving ground truth. Apparently, the search for duplicate enumerations is done within-household only; if so, the impact on data quality will not be positive.

Comparison of ICM and raking. Will ICM produce better estimates than raking to demographic analysis totals? For an affirmative answer, (i) the post stratification must control heterogeneity to low levels, and (ii) error rates in the field work and computerized matching must be controlled to low levels. Convincing evidence for the superiority of ICM will depend on the ability to measure heterogeneity and other errors. Experience from 1980 and 1990 suggests that such evidence will be quite hard to develop (see below). As good as the 1990 PES was, it was not good enough to accomplish its intended purpose—to produce an adjustment that was demonstrably more accurate than the census. Why is the ICM expected to do better?

Of course, demographic analysis has problems of its own. Indeed, detailed results from demographic analysis are somewhat inconsistent. Furthermore, results for any given decade have to be revised from time to time, in an effort to force consistency; the revisions give some indication of the uncertainty in the results. Finally, there seems to be substantial heterogeneity even after post stratification (Freedman and Wachter, 1994); on this score, raking must be worse. In 1980 and 1990, it was doubtful whether adjusting the census by raking to demographic analysis totals would have improved the accuracy of state population shares. However, in 2000, we think the problems with raking will be distinctly minor—by comparison with the problems in ICM. That is because raking is simpler, and more robust. The argument may be informal, but we find it quite persuasive.

THE SMOOTHING MODEL

The 1990 adjustment factors were smoothed separately in four geographical regions. In the Northeast, for instance, there were 300 post strata; index these by i. Let γ_i be the "true" adjustment factor for post-stratum i, and Y_i the raw adjustment factor derived from the DSE. In essence, the Bureau was using a hierarchical linear model. The first equation in the model can be stated as follows:

$$Y_i = \gamma_i + \delta_i,$$

where δ_i is a random error term. Let δ be the vector whose ith component is δ_i. The model assumes that $\mathrm{E}(\delta) = 0$ and $\mathrm{cov}(\delta) = K$, where K is a 300×300 matrix, which is unknown but to be estimated from the data.

The second equation states that

$$\gamma = X\alpha + \epsilon,$$

where X is a matrix of covariates computed from census data; the ith row describes post stratum i; again ϵ is a vector of random errors, assumed independent of δ, with $\mathrm{E}(\epsilon) = 0$ and $\mathrm{cov}(\epsilon) = \sigma^2 I$. Here, σ^2 is an unknown parameter, and α is a vector of unknown parameters. The random errors are assumed to be multivariate normal.

To begin with, K is estimated from the sample data by a version of the jackknife (Fay, 1990), but then variances are "pre-smoothed," that is, replaced by fitted values from some auxiliary regression model. Call the resulting estimate $\hat{K}$. (The Bureau defined the off-diagonal elements of $\hat{K}$ so the correlation matrix was preserved.) Next, choose some initial version X_0 for X. Then σ^2 and α can be estimated by maximum likelihood, and a new version of X can be chosen by a "best subsets" algorithm; iterate to convergence. (In the Northeast, for instance, the algorithm selected 18 variables out of 32.)

Write X for the "final" version of the design matrix; denote the MLE for σ^2 by $\hat{\sigma}^2$. Let P_X be the OLS projection matrix, $P_X = X(X'X)^{-1}X'$. Define the 300×300 matrix $\hat{\Gamma}$ as follows: $\hat{\Gamma}^{-1} = \hat{K}^{-1} + \hat{\sigma}^{-2}(I - P_X)$. Ordinarily, the vector of "smoothed" adjustment factors would be $\hat{\gamma} = \hat{\Gamma}\hat{K}^{-1}Y$, with covariances estimated as $\widehat{\mathrm{cov}}(\hat{\gamma} - \gamma) = \hat{\Gamma}$. However, the Bureau "benchmarked" $\hat{\gamma}$ so the total population in each of the four regions was unaffected by smoothing: this necessitated a more complex formula for the covariances. Allowance was made for uncertainty in $\hat{\sigma}^2$, but no allowance was made for uncertainty in $\hat{K}$, or in the choice of X; no allowance was made for specification error.

In the model, smoothing reduces variance in the adjustment factors without introducing bias, so variance can be interpreted as mean square error. But these results depend quite strongly on the assumptions; if the assumptions are violated, the real variances may be quite a bit larger than the model outputs, and bias may be appreciable. Smoothing and pre-smoothing appeared to reduce variances in estimated adjustments by a factor of about 2. However, simulation studies and sensitivity analysis suggest that, if anything, smoothing actually increased mean square error. If so, error rates estimated from the model are too small by a factor of 2 or 3.

Moreover, the detailed structure of the model seems quite arbitrary, and details have major impacts. For instance, the proposed adjustment in 1990 would have shifted population share from the Northeast and Midwest to the South and West; pre-smoothing and benchmarking account for 1/3 of the effect. There was some conflict, never resolved, between smoothing and benchmarking: if the model is right, benchmarking is counter-productive. See Fay (1992) or Freedman et al. (1993) for a detailed critique; Rolph (1993) takes a more positive view. Fortunately, the Bureau is not committed to smoothing in 2000. That may change, if the budget for sampling is cut. Waiting in the wings are models hauntingly like the smoothing model of 1990: see Zanutto and Zaslavsky (1995ab).

EVALUATING THE ADJUSTMENT

In both versions of the ICM, success depends on the ability to "link" or "match" records, or reconcile inconsistencies across surveys. Evaluations of the ICM turn on the ability to measure error rates in such reconciliations. In 1990, among other things, the Bureau used yet another survey—called the "Evaluation Follow-up," or EFU—to measure the error rates in the Post Enumeration Survey. It is helpful to consider four kinds of errors in the PES:

(i) "Sampling error," that is, random errors introduced by the luck of the draw in choosing the sample from which adjustments are estimated.

(ii) "Processing error," also called "non-sampling error." An example is clerical error in carrying out survey operations. These errors were measured, in part, by the EFU; we refer to them as "measured errors."

(iii) "Correlation bias" is created, for instance, when certain kinds of people are more likely than others to be missed both by the census and by the surveys used to adjust the census—even within a post stratum. This bias is not measured, even in principle, below the national level. The effect of correlation bias will be discussed later.

(iv) Failures in the "homogeneity assumption" are termed "heterogeneity."

Estimated error rates were used by the Bureau in its "Total Error Model" and "Loss Function Analysis"; the latter compares overall accuracy of the census and the adjustment, and is the topic of the next section. The Bureau's error rates were on the whole too low, and some sources of error were omitted from the model. As a result, while the net undercount in 1990 was estimated at 5.3 million persons—using the raw DSE on the PES data—PES errors of one kind or another account for 2.8–4.2 million persons out of the 5.3 million total. For instance, a coding error in the computer programs contributed by itself 1 million people to the proposed adjustment. In other words, a large part of the proposed adjustment to the 1990 Census resulted from measured errors in the PES. (The higher estimate of error rates comes from Breiman, 1994; the lower, from the Bureau itself: see U. S. Bureau of the Census, 1993b, p.75; also see Mulry, 1991, Table 15; the coding error in 1990 was this: in parts of the country, a PES in-mover who matched to the census was incorrectly declared to be a correct enumeration.)

Correlation bias, as defined above, is the error that occurs when people missed by the Census enumeration are also missed by a Post-Enumeration or ICM survey in greater numbers than the estimators provide for. In other words, there are "unreached people." At the national level, correlation bias offsets the measured errors to some degree. Differential correlation bias occurs when unreached people are concentrated in certain areas, making the estimates in those areas differentially too low. In 1990, the Bureau's models for the geographical distribution of correlation bias seemed quite unrealistic. Unreached people may well have been concentrated in the inner cities of the Northeast and Midwest. The 1990 adjustment, surprisingly enough, would have reduced the population shares of states like Pennsylvania, Ohio, Massachusetts, Michigan, New Jersey, and New York. Correlation bias is partly or largely to blame.

In the Total Error Model and Loss Function Analysis, the Bureau assumed that heterogeneity was trivial. However, there is strong evidence to show that heterogeneity was substantial, and the error analyses may have given quite a distorted picture for that reason alone. It is not clear how the Bureau will evaluate ICM in Census 2000; apparently, evaluation studies will not be published until well after the event.

LOSS FUNCTION ANALYSIS

In brief, "Loss Function Analysis" attempts to make unbiased estimates of "risk." Various loss functions can be considered, and different levels of geography. We focus here on squared error in population shares for 51 areas—the 50 states and D. C. With the Bureau's assumptions, the risk difference is estimated as 687 ± 281 favoring adjustment; our best estimate is 250 ± 821 favoring the census. (For details, see Freedman et al., 1994, or Table 1 below; units are parts per 100 million: on this scale, an error of 0.001 in estimating a share change corresponds to a squared error of $1/1,000^2 = 1/1,000,000$, that is, 100 parts per 100 million.)

Index the areas by $k = 1, \ldots, 51$. Let μ_k be the error in the census population share for area k. Let Y_k be the production dual-system estimate for μ_k, derived from the PES and the smoothing model. The bias in Y_k is denoted β_k; this is estimated from EFU as $\hat{\beta}_k$. The Bureau's model can be stated as follows: $Y \sim N(\mu + \beta, G)$, $\hat{\beta} \sim N(\beta, H)$, and $\hat{\beta}$ is independent of Y. Here, Y, μ, β, and $\hat{\beta}$ are 51-vectors of share changes; G and H are 51×51 matrices of rank 50: shares add to unity, so share changes add to 0, and one degree of freedom is lost. The smoothing model provides an estimator $\hat{G}$ for G; the Bureau has an estimator $\hat{H}$ for H; these estimators are assumed to be nearly correct; likewise, the model assumes $\hat{\beta}$ to be an unbiased estimator for β. We find these assumptions to be extremely questionable; for example, $\hat{\beta}$ is severely biased—as discussed above. (For details, see Breiman, 1994; also see Freedman et al., 1994.)

For area k, the risk from the census is μ_k^2, which can be estimated as $(Y_k - \hat{\beta}_k)^2 - \hat{G}_{kk} - \hat{H}_{kk}$; the risk from adjustment is $\beta_k^2 + G_{kk}$, which can be estimated as $\hat{\beta}_k^2 + \hat{G}_{kk} - \hat{H}_{kk}$. The estimated risk difference is

$$\hat{R}_k = (Y_k - \hat{\beta}_k)^2 - \hat{\beta}_k^2 - 2\hat{G}_{kk}.$$

Now we have unbiased estimates of variance, risks, and so forth—given the model, and given that

$$\mathrm{E}(\hat{G}) = G, \quad \mathrm{E}(\hat{H}) = H.$$

If variability in $\hat{G}$ can be ignored for a moment, the covariance in estimated risk differences is

$$\mathrm{cov}(\hat{R}_i, \hat{R}_j) = 4\mu_i\mu_j G_{ij} + 2G_{ij}^2 + 4E(Y_iY_j)H_{ij}.$$

The displayed expression can be estimated from sample data, as

$$\hat{C}_{ij} = 4(Y_i - \hat{\beta}_i)(Y_j - \hat{\beta}_j)\hat{G}_{ij} - 2\hat{G}_{ij}^2 - 4\hat{G}_{ij}\hat{H}_{ij} + 4Y_iY_j\hat{H}_{ij}.$$

Let ERD be the estimated risk difference summed over all 51 areas, that is, $\sum_j \hat{R}_j$; var ERD can now be estimated as $\sum_{ij} \hat{C}_{ij}$; this estimate will be used in Table 1 below.

Recall that G is the covariance matrix for proposed adjustments to the population shares of our 51 areas. As discussed above, the Bureau's estimate $\hat{G}$ was too small, by a factor of 2 to 3; furthermore, measured biases in the PES—response errors, matching error, errors in the imputations for missing data, the coding error, and so forth—amount to well over half the estimated undercount of 2.1%. Thus, $\hat{\beta}$ is severely biased. Now, another difficulty: the Bureau's estimate for $H = \mathrm{cov}(\hat{\beta})$ was off by a very large factor.

Bias was measured for 13 evaluation post strata, and allocated to 1,392 post strata. The allocation assumes bias to be proportional to some covariate for post strata within evaluation post strata, and constant across states within post strata. Bias is allocated down, then reaggregated to the 51 areas. (Table 1 reports on "PRODSE," which is one of the two schemes used by the Bureau for allocating bias from evaluation post strata to individual post strata; otherwise, we allocate bias to states in proportion to estimated undercounts.)

The EFU was about 7% of the size of the PES; there was no smoothing (factor of 2 reduction in apparent variance). Thus, trace$(\hat{H})$ should be $14 \times 2 \times \mathrm{trace}(\hat{G})$; instead, it is $.33 \times \mathrm{trace}(\hat{G})$. Consequently, $\hat{H}$ is off by $14 \times 2 \times 3 = 84$. In brief, the Bureau's allocation scheme converts variance in $\hat{\beta}$ to bias, and the model assumes that bias in $\hat{\beta}$ vanishes. Of course,

β is probably smaller than μ, which offsets our factor of 84 on $\hat{H}$ to some degree.

Table 1. Impact of allocation schemes for state-level biases, correction of final variances, and correction factors to variance estimates. ERD is the estimate of

$$\text{census risk} - \text{adjustment risk},$$

summed over the 51 areas. The "SE" column gives the estimated standard error of ERD under the various allocation schemes. Units are parts per 100 million.

Allocation	$\hat{G}$	$\hat{H}$	ERD	SE
PRODSE	1	1	667	281
PRODSE	1	50	667	890
PRODSE	2	1	542	371
PRODSE	2	50	542	885
.25 × undercount	1	1	193	199
.50 × undercount	1	1	−125	156
.50 × undercount	1	50	−125	859
.50 × undercount	2	1	−250	169
.50 × undercount	2	50	−250	821

Table 1 shows what happens when we revise the Loss Function Analysis to correct its various problems. The first line is our replication of the Bureau's Loss Function Analysis. Based on the Bureau's models, the estimated risk difference—summed over all 51 areas—is 667 with an SE of 281. (Units are parts per 100 million.) In the next line, we multiply $\hat{H}$ by a factor of 50. That does not change the estimated risk difference, but the SE shoots up to 890. The other lines in the table can be read the same way; in the last line, for instance, bias is taken to be half the undercount,

$\hat{G}$ is doubled, and $\hat{H}$ is multiplied by 50. Now it is adjustment that has the higher estimated risk, by 250 parts per 100 million; of course, the SE is rather large. Evaluating the 1990 adjustment is quite problematic, because we still do not know where the undercounted people were living: the PES and the EFU do not provide enough evidence to decide the merits of adjustment.

To summarize: when we corrected the Loss Function Analysis to take PES errors into better account, we found that the adjusted figures may be more accurate—or the census. The difference in accuracy may be statistically significant, or just at the chance level: conclusions are driven by assumptions, not by data. A more rudimentary loss function analysis: the proposed adjustment in 1990 would have moved two seats in Congress. After correction for the coding error—but not the other measured errors—only one seat moves. Much of the adjustment is due to error.

For details on Loss Function Analysis and the Total Error Model, including caveats on the t-statistics implied by Table 1, see Breiman (1994), Freedman and Wachter (1994), or Freedman et al. (1994). For a view from the Bureau, see Mulry and Spencer (1993); that paper does not respond the issues raised here. For a response, see Belin and Rolph (1994) or Ericksen, Fienberg and Kadane (1994); but also see Breiman (1994, pp.521–27), Freedman and Wachter (1994, pp.527–37).

3. THE NATIONAL ACADEMY OF SCIENCES

The NAS studies may be viewed by some as evidence for the soundness of ICM. The two most recent reports are Edmonston and Schultze (1995) and Steffey and Bradburn (1994). We focus on the latter. Its arguments are generally theoretical rather than empirical, and normative rather than descriptive. Even so, the report concedes that many issues will need to be resolved before taking ICM into the field for Census 2000 (pp.115–35).

Few details are given on problems in 1990. The report does acknowledge a "controversy" about smoothing models (p.124), it notes concerns about correlation bias (p.109), and says there is a "lively debate" about heterogeneity (p.125). On p.121, the report says, "'Loss measures' such as weighted mean squared error are credible ways of combining estimated bias and variance into a single measure at any selected level of geography." So they would be—if the underlying estimates of bias and variance were credible. However, difficulties with the total error model and loss function analysis are not discussed. Therefore, Steffey and Bradburn cannot explain how these difficulties will be surmounted in 2000.

There is some discussion (pp.129–131) of demographic analysis, which cites many of the difficulties we have mentioned. Steffey and Bradburn do not discuss the idea of raking to DA totals as an alternative to ICM. But they do consider (and reject) the much less appealing idea of incorporating raking as part of ICM: the hybrid may inherit the worst features of both parents.

With reference to CensusPlus, the report says (p.110), "The assumption of complete coverage replaces the independence assumption. . . ." Assumptions by themselves only serve to focus discussion. The critical issue is whether CensusPlus can be implemented so as to translate its basic assumptions into reality. When discussing this issue, the report promises that (p.110), "subsequent sections [will] consider possible problems and discuss evaluations relevant to this assumption." By p.116, the report concedes that complete coverage is illusory, but

> "we caution against using this concern by itself to dismiss any ICM method. Even if no method can totally eliminate the differential undercount, which is probably the case, ICM may well lead to substantially better estimates."

Of course, ICM *may* lead to better estimates; however, the experience of 1990 suggests that adjustment may also lead to worse estimates. The NAS reports have not addressed the hard questions.

4. CITATIONS TO THE LITERATURE

The synthetic method for adjusting small areas is explained in Freedman and Wachter (1994, pp.476–77); for the impact of heterogeneity on loss function analysis, see Freedman and Wachter (1994, pp.483ff, pp.532ff). On demographic analysis, see Passel (1990); Freedman (1991) discusses some of the uncertainties; also see Robinson et al. (1993), but the "uncertainty intervals" in that paper seem quite misleading, see Clogg and Himes (1993). Freedman and Navidi (1992, p.62) discuss the revision history of demographic analysis estimates; that paper also reviews the evidence on proposed adjustments for 1980. For a defense of synthetic estimation in 1980, see Schirm and Preston (1987) or Wolter and Causey (1991); but also see Freedman and Navidi (1992, pp.20ff and 64ff). Fay et al. (1988) discusses evaluations of the 1980 census; Warren and Passel (1987) discusses illegal aliens counted in that census. A view from the Bureau on 1990 is outlined in Wolter (1991); also see Fienberg (1993). A more recent evaluation by senior Bureau personnel—with which we generally agree—is Fay and Thompson (1993). The decision on the 1990 adjustment is U. S. Department of Commerce (1991), which includes Passel (1990) on demographic analysis, and Murray (1991); the latter paper describes the impact of adjustment on the allocation of tax moneys. Also see "Dividing the Dollars: Issues in Adjusting Decennial Counts and Intercensal Estimates for Funds Distribution," S. Prt. 102-83; this Senate Report, by the Committee on Governmental Affairs (1992), estimates that about $35 billion a year in tax funds would have been affected by adjustment, with a shift of about 1/2 of 1% of the resources.

Correlation bias is discussed in Freedman and Wachter (1994, pp.533–34); also see (Wachter, 1993, pp.110–13). Our view of the Bureau's models for correlation bias is explained in Freedman and Wachter (1994, pp.533–34); a Bureau view is given by Bell (1993), also see Mulry and Spencer (1993). These papers do not respond to our concerns; for a response, see Belin and Rolph (1994) or Ericksen, Fienberg and Kadane (1994); but also see Breiman (1994, pp.521–27), Freedman and Wachter (1994, pp.527–37). The legal discussion (below) is paraphrased in part from Freedman (1993), Freedman et al. (1994), Kaye and Freedman (1996).

Plans for Census 2000 are described in U. S. Bureau of the Census (1996). Our information about Census 2000 and the 1995 test census in Oakland comes from this document, discussions with Bureau personnel, and memoranda provided by the Bureau; as always, we are indebted to the Bureau for its cooperation. Clark et al. (1994) has the sample sizes used in Oakland; Killion (1996) and Miskura (1995) discuss interviewer behavior; Leslie (1995) discusses missing data; Neugebauer (1996) discusses administrative records; Thompson et al. (1995) provides an overview of plans for 2000, including use of administrative records and possible use of models; Treat (1995) discusses linking; Whitford (1995) gives some details on the ICM. Our reading of these memoranda is of course "selective," in the sense that the Bureau seems quite sanguine about prospects for ICM in the year 2000. Any technical difficulties in the 1995 Oakland pretest will, they must think, be resolved in time for Census 2000. However, the Bureau was also quite sanguine about the PES in 1990. This time, they may have it right. We do not yet see the argument.

5. THE SUPREME COURT

For 1990, the Department of Commerce decided not to adjust the census (U.S. Department of Commerce, 1991). Three major court cases were brought to challenge this decision (citations are given, below). In Detroit v. Franklin, the courts saw no real evidence of differential undercounts, and held that neither statutes nor the constitution required adjustment. In Tucker v. Commerce, the courts held that there were legitimate technical arguments on both sides of the adjustment issue, and it was up to the executive branch to call the shot: the issue was not reviewable by the courts.

In New York v. Commerce, the trial court ("District Court") held that the issue was reviewable, and the main test was this: did the Secretary of Commerce make a reasonable decision? Evidently, the court leaned toward adjustment, but it ruled that a decision not to adjust was reasonable:

> "Midst all the sturm und drang, after all is said and done, the question before the Court distills to this: did the Secretary act reasonably? This, of course, depends mainly upon the evidence he had before him. In his testimony, Dr. Robert E. Fay, one of the principal statisticians at the Census Bureau (who, incidentally, voted to adjust) pierced right to the heart of the case: 'I told the Secretary that ... reasonable statisticians could differ on this conclusion ....' The Court agrees, and therefore, concludes that the Secretary's decision not to adjust the 1990 census count was neither arbitrary nor capricious." (822 F.Supp. 928, that is, page 928 in volume 822 of the *Federal Supplement*, which reports court opinions.)

New York appealed to the "Second Circuit"—the appeals court for that jurisdiction. The appeals court vacated the judgment of the district court and remanded for further proceedings. Generally, an appeals court will rely on the district court's findings of fact. However, the appeals court held that adjustment was more accurate than the census, paraphrasing the district court to say that "for most purposes and for most of the population ...

adjustment would result in a more accurate count than the original census." (34 F.3d 1129, that is, *Federal Reporter*, 3rd series, volume 34, page 1129.) However, the Second Circuit seems to have picked and chosen among the findings of the district court. For example, according to the district court, plaintiffs had failed to "illustrate affirmatively the superior accuracy of the adjusted counts [at the state and local level] for any reasonable definition of accuracy." (822 F.Supp. 924) The Second Circuit simply ignored the findings that did not suit.

The Second Circuit emphasized numeric accuracy, and criticized the Secretary of Commerce for giving priority to "distributive accuracy," that is, accuracy of population shares for geographic areas (34 F.3d 1131). Curiously enough, the Second Circuit based its own legal argument on cases (including Baker v. Carr, Wesberry v. Sanders, Reynolds v. Sims, and Karcher v. Daggett, 34 F.3d 1125–29) that deal with distributive accuracy. The Second Circuit ruled that the Secretary's decision was subject to "heightened scrutiny:" to prevail, the Federal Government needed to show some compelling reason of state not to adjust. The Second Circuit ordered the district court to decide the case on that basis.

The State of Wisconsin had intervened on the government side in New York v. Commerce, and was therefore entitled to appeal to the Supreme Court. (It was joined by the Federal Government.) In Wisconsin v. City of New York, the Supreme Court unanimously over-ruled the Second Circuit, and upheld the Secretary's decision not to adjust the census. The Supreme Court determined that the exacting requirements of the equal protection clause, as explicated in congressional redistricting and state reapportionment cases, do not "translate into a requirement that the Federal Government conduct a census that is as accurate as possible" and do not provide any basis for "preferring numerical accuracy over distributive accuracy." Concluding that the government had shown a "reasonable relationship" between the decision not to adjust and "the accomplishment of an actual enumeration of the population, keeping in mind the constitutional purpose of the census . . . to determine the apportionment of the Representatives among the States," the Court held that the Secretary's decision satisfied the constitution. Indeed, having rejected the argument that the constitution compelled statistical adjustment, the Court noted that the constitution might prohibit such adjustment (footnote 9). Other quotes from the opinion may also be of interest:

"In 1990, the Census Bureau made an extraordinary effort to conduct an accurate enumeration, and was successful in counting 98.4% of the population.... The Secretary then had to consider whether to adjust the census using statistical data derived from the PES. He based his decision not to adjust the census upon three determinations. First, he held that in light of the constitutional purpose of the census, its distributive accuracy was more important than its numerical accuracy. Second, he determined that the unadjusted census data would be considered the most distributively accurate absent a showing to the contrary. And finally, after reviewing the results of the PES in the light of extensive research and the recommendations of his advisers, the Secretary found that the PES-based adjustment would not improve distributive accuracy. Each of these three determinations is well within the bounds of the Secretary's constitutional discretion.

"Moreover, even those who recommended in favor of adjustment recognized that their conclusion was not compelled by the evidence.... Therefore, and because we find the Secretary's ... prior determinations ... to be entirely reasonable, we conclude that his decision not to adjust the 1990 census was 'consonant with ... the text and history of the Constitution....'" (Supreme Court Opinion dated March 20, 1996, in cases 94-1614, 94-1631, and 94-1985, pp.18 and 22.)

6. SUMMARY AND CONCLUSIONS

There will be four options in the year 2000: the census itself, raking, CensusPlus and the DSE. Which is to be preferred? Given the experience of 1990, we doubt any convincing quantitative evaluation will be achieved. If the census is not considered viable, we prefer raking. Our option is simpler than ICM, and there is less chance of statistical disaster.

REFERENCES

Belin, T. R. and J. E. Rolph. 1994. Can we reach consensus on census adjustment? *Statistical Science* 9: 486–504, 520–21 (with discussion).

Bell, W. R. 1993. Using information from demographic analysis in Post-Enumeration Survey estimation. *Journal of the American Statistical Association* 423: 1106–18.

Breiman, L. 1994. The 1991 census adjustment: undercount or bad data? *Statistical Science* 9: 458–75, 521–27 (with discussion).

Clogg, C. C. and C. L. Himes. 1993. Comment: uncertainty in demographic analysis. *Journal of the American Statistical Association* 423: 1072–74.

Edmonston, B. and C. L. Schultze, eds. 1995. *Modernizing the U. S. Census.* Washington, D. C.: National Academy Press.

Ericksen, E. P., S. E. Fienberg, and J. B. Kadane. 1994. Comment. *Statistical Science* 9: 511–15.

Fay, R. E. 1990. VPLX: Variance estimates for complex samples. Prepared for the Annual Meetings of the American Statistical Association, August 1990, Anaheim, CA. Technical report, Bureau of the Census, Washington, D. C.

Fay, R. E. 1992. Inferences for small domain estimates from the 1990 Post Enumeration Survey. Technical Report, Bureau of the Census, Washington, D. C.

Fay, R. E., J. S. Passel, J. G. Robinson, and C. D. Cowan. 1988. *The Coverage of the Population in the 1980 Census*. Washington, D.C.: U. S. Department of Commerce, Government Printing Office.

Fay, R. E. and J. H. Thompson. 1993. The 1990 Post Enumeration Survey: Statistical lessons in hindsight. In *Proceedings of the 1993 Annual Research Conference*, pp. 71–91. Washington, D. C.: U. S. Bureau of the Census.

S. E. Fienberg (1993). The New York City census adjustment trial: witness for the plaintiffs. *Jurimetrics* 34: 65–83.

Freedman, D. A. 1993. Adjusting the census of 1990. *Jurimetrics* 34: 99–106.

Freedman, D. A. 1991. Adjusting the 1990 census. *Science* 252: 1233–36. Some paragraphs in the present paper are paraphrased from this source, copyright by AAAS.

Freedman, D. A. and W. C. Navidi. 1992. Should we have adjusted the U. S. census of 1980? *Survey Methodology* 18: 3–74 (with discussion).

Freedman, D. A. and K. W. Wachter. 1994. Heterogeneity and census adjustment for the inter-censal base. *Statistical Science* 9: 476–85, 527–37 (with discussion). Some paragraphs in the present paper are paraphrased from this source, copyright by IMS.

Freedman, D., K. Wachter, D. Coster, R. Cutler, and S. Klein. 1993. Adjusting the census of 1990: the smoothing model. *Evaluation Review* 17: 371–443. Some paragraphs in the present paper are paraphrased from this source, copyright by Sage Publications.

Freedman, D., K. Wachter, R. Cutler, and S. Klein. 1994. Adjusting the census of 1990: loss functions. *Evaluation Review* 18: 243–280. Some paragraphs in the present paper are paraphrased from this source, copyright by Sage Publications.

Hogan, H. 1993. The 1990 Post Enumeration Survey: operations and results. *Journal of the American Statistical Association* 423: 1047–60.

Kaye, D. H. and D. A. Freedman. 1996. Statistical Proof. To appear in *West's Companion to Scientific Evidence*. D. Faigman et al., eds. Some paragraphs in the present paper are paraphrased from this source, copyright by D. H. Kaye and D. A. Freedman.

Mulry M. H., and B. D. Spencer. 1993. Accuracy of 1990 census and undercount adjustments. *Journal of the American Statistical Association* 423: 1080–91.

Murray, M. P. 1991. Census adjustment and the distribution of federal funding. Technical report, Bates College. Reprinted in U. S. Department of Commerce (1991).

Passel, J. S. 1990. *Demographic analysis: a report on its utility for adjusting the 1990 Census.* Technical Report, Urban Institute, Washington, D. C. Reprinted in U. S. Department of Commerce (1991).

Robinson, J. G. et al. 1993. Estimation of population coverage in the 1990 United States Census based on demographic analysis. *Journal of the American Statistical Association* 423: 1061–79 (with discussion).

Rolph, J. E. 1993. The census adjustment trial: reflections of a witness for the plaintiffs. *Jurimetrics* 34: 85–98.

Schirm, A. L. and J. Preston. 1987. Census undercount adjustment and the quality of geographic population distributions. *Journal of the American Statistical Association* 82: 965–90 (with discussion).

Steffey, D. L. and N. M. Bradburn, eds. 1994. *Counting People in the Information Age*. Washington, D. C.: National Academy Press.

U. S. Bureau of the Census. 1996. The Plans for Census 2000. Washington, D. C.

U. S. Bureau of the Census. 1993a. *Money Income of Households, Families, and Persons in the United States: 1992*, Current Population Reports P60-184. Bureau of the Census, U. S. Department of Commerce, Washington, D. C.

U. S. Bureau of the Census. 1993b. Decision of the Director of the Bureau of the Census on whether to use information from the 1990 Post-Enumeration Survey (PES) to adjust the base for the intercensal population estimates produced by the Bureau of the Census. *Federal Register* 58: 69–77 (January 4th).

U. S. Department of Commerce. 1991. Office of the Secretary. Decision on Whether or Not a Statistical Adjustment of the 1990 Decennial Census of Population Should Be Made for Coverage Deficiencies Resulting in an Overcount or Undercount of the Population; Explanation. Report dated July 15, Washington, D. C. Portions are reprinted in *Federal Register* 56: 33582–33642 (July 22nd).

Wachter, K. W. 1993. The census adjustment trial: an exchange. *Jurimetrics* 34: 107–115.

Warren, R. and J. S. Passel. 1987. A count of the uncountable: estimates of undocumented aliens counted in the 1980 United States Census. *Demography* 24, no. 3.

Wolter, K. 1991. Accounting for America's uncounted and miscounted. *Science* 253: 12–15.

Wolter, K. and B. Causey. 1991. Evaluation of procedures for improving population estimates for small areas. *Journal of the American Statistical Association* 86: 278–84.

Woltman, H. 1991. Estimated state level adjusted counts based on revised state groups. Bureau of the Census, memorandum dated 7/15/91.

Zanutto, E. and A. M. Zaslavsky. 1995a. A model for imputing nonsample households with sampled nonresponse follow-up. Technical report, Department of Statistics, Harvard.

Zanutto, E. and A. M. Zaslavsky. 1995b. Models for imputing nonsample households with sampled nonresponse followup. Technical report, Department of Statistics, Harvard.

BUREAU MEMORANDA

Clark, J. et al. 1994. Integrated coverage measurement status report.

Killion, R. 1996. Results from the 1995 census test: responses to the coverage probes in the Integrated Coverage Measurement persons inteview—Project 4. No. 21 in the series.

Leslie, T. 1995. November 7, 1995 debriefing results: automated matching evaluation. No. 22 in the series.

Miskura, S. 1995. Results from the 1995 Census test: the Integrated Coverage Measurement evaluation of tape recorded interviews with behavior coding—Project 2. No. 18 in the series.

Neugebauer, S. 1996. Administrative records file acquisition history for the 1995 Census test. No. 24 in the series.

Thompson, J. et al. 1995. Census 2000: statistical issues in reengineering the decennial census.

Treat, J. 1995. Integrated Coverage Measurement (ICM) evaluation project 5: ICM interviewer linking error. No. 20 in the series.

Whitford, D. 1995. Documentation of the design of the 1995 Integrated Coverage Measurement.

LEGAL CITATIONS

City of Detroit v. Franklin, 4 F.3d 1367 (6th Cir. 1993) cert. denied, 114 S. Ct 1212 (1994);

Tucker v. Dep't of Commerce, 958 F.2d 1411 (7th Cir.) cert. denied, 113 S. Ct. 407 (1992).

New York City v. United States Dep't of Commerce, 822 F. Supp. 906 (E.D.N.Y. 1993); reversed and remanded, 34 F.3d 1114 (2d Cir. 1994); overruled sub nom. Wisconsin v. City of New York, 1996 WL 119521 (U.S. Mar.20, 1996)

AUTHOR'S FOOTNOTE

The authors testified as expert witnesses for the government in the 1980 and the 1990 lawsuits over Census adjustment. Wachter was a member of the Special Advisory Panel on Census Adjustment of the Secretary of Commerce, 1989–1991. Paul Humphreys, Richard Olshen, John Rice, Terry Speed, and an anonymous referee all made helpful comments.

ABOUT THE AUTHORS

David A. Freedman is Professor of Statistics and Mathematics, University of California, Berkeley. His research interests are in the foundations of statistics, modeling, and policy analysis.

Kenneth W. Wachter is Professor of Demography and Statistics, University of California, Berkeley. His research interests include mathematical demography, demographic heterogeneity, computer microsimulation, kinship forecasting, and statistical astronomy.