

ESTIMATING COVERAGE OF THE 1980 UNITED STATES CENSUS: DEMOGRAPHIC ANALYSIS

Jeffrey S. Passel and J. Gregory Robinson, U.S. Bureau of the Census

The Census Bureau's plan for evaluating the completeness of coverage of the 1980 census relies on three basic approaches. Demographic analysis will be used to produce official coverage estimates at the national level. Estimates for States and some major SMSA's and cities are planned to be developed from several combined samples of the Current Population Survey. Persons interviewed in the surveys would be matched to the census on a case-by-case basis to produce undercount estimates. Further matching to various administrative files is also being planned to improve these estimates. The method of demographic analysis will also be used to prepare State undercount estimates but the quality of the resulting estimates is uncertain. On the basis of these projects, the Census Bureau is planning to issue official estimates of census coverage for the nation and States as well as the largest SMSA's and cities. Finally, the possibility of using statistical techniques to develop experimental coverage estimates for counties and local areas and preliminary coverage estimates for States is under investigation.

Demographic analysis as a tool for census evaluation involves developing expected values for the population in various categories (such as age, sex, race categories) at the census date by the combination and manipulation of various types of demographic data and then comparing these values with the corresponding census counts. The foundation of the demographic method is the logical consistency and interrelationships of the underlying demographic data. These data are drawn from sources essentially independent of the census, such as birth, death, and immigration statistics, historical series of census data, and data from sample surveys. The data are corrected for various types of errors and, as such, are assumed to be more accurate than the census being evaluated. The accuracy of the method obviously depends on the quality of the demographic data and the corrections.

The estimates resulting from the application of demographic analysis provide measures of net underenumeration of the total population, but combine coverage and classification error for age, sex, and race groups. Following the 1970 census, demographic analysis was used to produce the official estimates of coverage for the United States as a whole. This method is considered by the census staff to be more effective than a post-enumeration sample survey for developing satisfactory estimates of net undercounts at the national level; consequently, the demographic estimates are envisioned as the official national estimates of census coverage for 1980.

National Estimates

The particular procedure used to estimate the coverage nationally for the various demographic subgroups, notably age groups, depends on the nature of the available data and on the timing

requirements of the overall evaluation program. For the population under age 45 in 1980, i.e., persons born after 1935, estimates of the corrected population will be developed from birth, death, and immigration statistics. For the population over age 65, aggregate Medicare data will provide the basis for coverage estimates. For the remaining ages, 45 to 64 years, the coverage estimates will be extensions of the estimates for ages 35 to 54 in 1970; these were derived from analysis of previous censuses. Actual death statistics will be used to allow for mortality up to age 74. Official immigration statistics supplemented by estimates of other immigration and emigration will be used to allow for net immigration.

Different methods may be used for the same age groups in preliminary and revised estimates because of the availability of different data. In some instances, it is not possible to specify the choice among alternatives for the revised estimates at this time. It should be noted that demographic analysis has not proven successful in developing coverage estimates for the Hispanic population. Estimates of coverage for this group in 1980 are expected to be obtained from match studies.

Birth and Death Statistics

Registered births over several decades provide a direct basis for estimating the corrected numbers of persons in most age groups in 1980. Statistics on registered births are available (by race and sex) for all States since 1933. In addition, tests of birth registration completeness were conducted for 1940, 1950, and 1964-68 which provide correction factors for these years; factors for other years can be obtained by interpolation and extrapolation. These data will be used in estimating census coverage for persons under age 45 in 1980, age groups which will comprise almost 70 percent of the population.

Research conducted at the Census Bureau suggests that the results of the birth registration tests for Whites can be improved. Regression estimates of birth registration completeness are derived on the basis of the test results using States as the units of analysis. The independent variables include the percent of registered births occurring in hospitals and dummy variables for location as well as various characteristics of each State registration system. Interpolation and extrapolation for non-test years is done on a State-by-State basis and uses interpolated regression coefficients. The results for States are aggregated to give national estimates of corrected births for years since 1935. The possibility of extending this work to Blacks-and-other-races is currently under investigation.

Another project, if successful, would extend the estimates of corrected births to years prior to 1935. Estimates of the completeness of birth

registration for the White population of States in 1920 and 1930 are being developed from demographic models and regression techniques. Results for 1930 especially appear quite promising. This project would permit the estimation of coverage nationally and for States from birth statistics for ages 45 to 64 in 1980 in addition to the younger ages. This technique is believed to be superior to that used to produce coverage estimates for ages 35 to 54 in 1970, especially for States.

Current plans call for using registered deaths in the demographic estimates with no correction for underregistration or misreporting of the characteristics of decedents. One improvement in this area is a method of allocating deaths to cohorts which relies on smoothing the death rates rather than the distribution of deaths. Alternative calculations will permit investigation of the effect on coverage estimates of allowances for possible underregistration of deaths, particularly for infant deaths in earlier years. Similarly, effects of age misreporting on death certificates may be investigated.

Immigration Statistics

Data on legal immigrants admitted to the United States from 1935 to 1980 are used in estimating the corrected population under age 65 in 1980. The Immigration and Naturalization Service (INS) publishes data on immigrants classified by age, sex, and country of origin. The immigration component of the estimates of corrected population also includes other data items, some supplied by INS. These include net arrivals in the United States from Puerto Rico, net arrivals of other civilian citizens, and parolees. Investigation of the quality of these data, particularly net arrivals from Puerto Rico, as well as research into various estimation techniques is being undertaken.

One component of the expected population in 1980 for which data are lacking is illegal immigration. Obviously, because of the nature of this population, an accurate estimate of its size will be quite difficult, if not impossible, to make. The range of existing estimates for the illegal population in recent years is quite broad. Although a variety of estimation techniques have been used to try to establish the number of illegal residents, no satisfactory estimates of either the net flow or the size of the illegal population in the United States are available and the true number remains unknown.

The Census Bureau has undertaken an evaluation of the existing studies of illegal immigration to the United States and is investigating various approaches to the estimation problem in addition to those previously employed. A previous study which investigated changes in age-specific death rates through 1975 is being extended to 1980. Intercensal cohort analysis of the Mexican censuses of 1960, 1970, and 1980 would be expected to yield information on net immigration (legal and illegal combined) from Mexico. In a similar vein, a project attempting to reconstruct the population of Mexico from vital statistics might

yield similar information once the 1980 census of Mexico is available. On the basis of its own research and of others, the Bureau hopes to develop a range of working estimates of the illegal alien population to be included in the estimate of the expected population in 1980.

Movement out of the United States (by both citizens and aliens) is another component for which little satisfactory data exist. The methods used in the past to develop emigration estimates present problems in terms of timeliness, coverage, consistency and accuracy of the estimates over time, and the scope of the assumptions required. Consequently, the Census Bureau has been considering a test of network (multiplicity) sampling in conjunction with the Current Population Survey to investigate the feasibility of obtaining information on emigration from the United States.

With the multiplicity technique, respondents are asked whether certain specified relatives have emigrated from the United States. Persons with such relatives are asked further questions to obtain information about the emigrants and for weighting the sample responses. No results from the proposed test of the multiplicity survey will be available for the preliminary estimates of undercount so that indirect estimation techniques will have to be employed. These techniques include analysis of INS alien registration data, data on Social Security beneficiaries abroad, IRS data on tax filers abroad, and analysis of foreign censuses. Should a test of the multiplicity approach prove successful, it may be possible to develop emigration estimates on the basis of a full-scale survey conducted some time in 1981.

Revisions of estimates of emigration for 1950-60 and 1960-70 are being considered. For the latter period, estimates based on intercensal cohort analysis of the foreign-born population have already been published. The same technique is being applied for the earlier period.

Medicare Statistics

The corrected population aged 65 years and over for age, race, and sex categories will be developed from aggregated Medicare statistics. Those statistics will be adjusted for estimated underenrollment and then compared with census counts to obtain undercount estimates. The factors for adjusting the Medicare data are to be obtained as a by-product of proposed match studies involving the Current Population Survey, to be conducted in 1981.

In the meantime, several alternatives are available for providing preliminary coverage estimates for the population 65 and over in age-sex-race categories. One alternative would involve carrying forward the corrected population aged 55 and over in 1970 with estimates of deaths and net immigration from abroad. Another would utilize preliminary aggregated Medicare data for 1980. These data would be corrected for underenrollment on the basis of correction factors developed from a test of record matching techniques involving CPS, IRS, and Medicare data for February 1978. A third approach would involve

correcting the Medicare data for groups which are known to be excluded. For any approach, estimates of the corrected population in some groups can be derived using expected sex ratios. This was in fact done in 1970.

Preliminary and Revised Estimates

The Census Bureau is planning to release at least two national coverage estimates based on demographic analysis: (1) preliminary estimates to meet the demand for estimates at the earliest possible date; and (2) revised estimates to incorporate data and research findings which become available later. Preliminary estimates of coverage for the total population will be released about January 1, 1981; preliminary coverage estimates for age, sex, and race categories (White, Black, and Other or Black and Other combined) will be released shortly thereafter, i.e., as soon as review of appropriate census tabulations can be completed. Because of the need to analyze the census data fully (particularly the racial categories) and to complete ongoing research, the revised "demographic" estimates of coverage are not likely to be available until mid-1982.

As noted earlier, the preliminary and revised estimates will use different data and methods to estimate various components of the corrected population in 1980. For ages 45 to 64, the preliminary estimates will most likely be extensions of the estimates for ages 35 to 54 in 1970, which are based on analysis of previous census data. The revised estimates for these ages may be based on survivors of births as corrected at the State level on the basis of research now being conducted. Other revisions involve replacing provisional data on births, deaths, and immigration for 1979 and 1980 with final data. Because of the uncertainty involved in estimating illegal immigration and the availability of new data, it is very likely that any estimate of the illegal alien population included in the coverage estimates will be modified in the revised coverage estimates if illegal aliens can be included at all. Furthermore, in both instances, a range of estimates may be employed.

Problems in obtaining the required data prevent completion before 1982 of revised estimates of the population over 65 years of age. All Medicare enrollees as of April 1, 1980, will not be included in the data files until the end of 1980. More importantly, however, since a match involving the census or the coverage evaluation survey and administrative records is required to develop the revised underenrollment factors for adjusting the Medicare data, the estimates cannot be completed until the match results are known, analyzed, and incorporated into the estimation procedure.

Alternative Racial Groupings

Two major changes in coding and classifying racial groups in the 1980 census will create inconsistencies between the 1980 census data and historical data series (including censuses, vital statistics, and immigration records). In the 1980 census, "Asian Indian" has been added as a

separate response category to the race question. Furthermore, this category will be included in the "Asian and Pacific Islander" grouping in collapsed tabulations. In past censuses and in vital statistics, persons of Asian Indian descent have been classified as "White." (This group numbered about 75,000 in 1970.) The other change, of much greater numerical impact, is in the treatment of persons who write in a national origin designation of a Latin American country (e.g., Mexican, Venezuelan, Cuban, etc.) in response to the race question. In 1980, these persons will remain in the "Other" category; in previous censuses and in vital statistics, such responses have been coded into the "White" category. Because so many Hispanics are likely to mark "Other" race, this classification difference could affect 4-5 million persons in 1980.^{1/}

Because of the necessity for agreement between the racial groupings in the expected population as calculated from demographic analysis and the census population, either the expected population or the census population must be adjusted to conform to the other.^{2/} Adjusting the expected population would necessitate at least changing birth, death, and immigration statistics for a period of 45 years as well as Medicare records. Unfortunately, the data do not exist within these data systems or any others for making the required adjustments. Consequently, it will be necessary to adjust the 1980 census data to conform to the historical categories. The resulting preliminary estimates of undercount for the 1980 census as derived by demographic analysis will apply to the racial groupings of the adjusted data rather than the data as collected and published in almost all census reports.

The adjustment procedures are relatively straightforward. The Asian-Indian group will simply be included with the White category. To produce census data consistent with the corrected population for the "Hispanic-Others" a cross-classification of persons giving age, sex, race, and Hispanic origin will be required. Within each age-sex category, persons of Hispanic origin and "Other" race will be distributed to the designated racial categories. Two distribution schemes are under consideration: one would place all persons of Hispanic origin and "Other" race into the White category; the other would distribute such persons on a pro rata basis within age-sex categories according to the racial distribution of all or part of the remaining Hispanic origin population. The effect of alternative distribution schemes will be investigated. In addition, the geographic level at which the redistribution will be carried out will vary. In the preliminary coverage estimates, preliminary State-level census data will provide the basis for distributing the "Hispanic-Other" group. For the revised estimates, smaller areas, possibly Revenue Sharing areas, will be considered.

Demographic Estimates for States

A "demographic" approach to estimating the coverage of State populations was attempted for 1970. The basic approach involved first estimating the coverage of the population born in each

State (for ages under 35 in 1970) by comparing survivors of births with the census data on the population born in each State. Then, several different procedures and assumptions were used to convert the coverage estimates for the population born in each State into estimates for the population living in each State. The procedures and assumptions involved a number of parameters for which no direct estimates were available. Accordingly, a number of alternative sets of State coverage estimates for the population under 35 were derived rather than a single preferred set. Other uncertainties in the estimation procedure, particularly the lack of any reliable methodology for ages 35 to 64, led to other alternative estimates and to the characterization of the estimates as "developmental."

For 1980, we plan again to produce "demographic" estimates of census coverage for States. Several developments should remove some of the uncertainties in the estimation methods to the point where it may be possible to derive a usable set of estimates. The CPS-Census match studies, if successful, would provide estimates of the relative coverage of lifetime interstate migrants and nonmigrants. This parameter is crucial to the estimation procedure; values had to be assumed for 1970. The information from the match study should permit a simplification of the method, including elimination of the separate calculation of coverage estimates for State-of-birth populations and the resulting necessity of converting those estimates to represent States of residence.

The data on births which extend back to 1935 will cover a larger proportion of the population in 1980, i.e., under age 45. In addition, as previously mentioned, research is under way to extend the data on corrected births back to 1925

or 1915. If successful, these projects would virtually eliminate the need for ratio estimates in the middle age range. We also hope to remove some of the uncertainty in the coverage estimates for ages 65 and over with the results of studies being conducted at the Bureau of the Census on the accuracy of residence reporting in the Medicare files.

The "demographic" estimates of coverage for States are based on the place-of-birth data collected on the sample form. Since these data will not be available before 1982 and considerable analytic work is necessary, the coverage estimates for States will not be completed before late 1983. The quality of the data on State of birth has apparently deteriorated between the 1950 and 1970 censuses, e.g., higher nonresponse rates and evidence of greater misreporting. Should this trend continue, there will be serious problems with any methodology employing census data on State of birth.

Footnotes:

1/ In 1970, more than 300,000 persons of Hispanic descent who responded in the "Other" race category were not recoded initially into the "White" category as specified by the coding rules in effect for that census, thus necessitating revisions in the data. Pretests for the 1980 census indicated that many more Hispanics would respond in the "Other" category. The anticipated coding difficulties along with other concerns dictated the change in coding procedures.

2/ Adjustment, in this context, refers to revisions and reclassifications in the reported data because of classification concepts, not because of missing entries or missed persons.