

WHAT IS THE ROLE OF DEMOGRAPHIC ANALYSIS IN THE 2000 UNITED STATES CENSUS?

J. Gregory Robinson¹

ABSTRACT

Demographic analysis (DA) is a well-developed coverage measurement and evaluation program in the United States. DA has served as the standard for measuring coverage trends in recent censuses and differences in coverage by age, sex, and race at the national level. In this paper, we explore the role that demographic analysis can play in the 2000 census:

* Should DA be only a coverage evaluation tool in support of the survey-based coverage estimates (CensusPlus or Dual System Estimation) used in the Integrated Coverage Measurement (ICM) operations, or

* Should DA be formally integrated with the survey estimates into the ICM coverage measurement process, drawing from the particular strengths of the demographic approach?

The role of DA should be based on balancing the strengths and limitations of the demographic method and of the survey-based coverage estimates. We believe that demographic analysis can play an important and expanded evaluation role in the 2000 census. DA also has the potential to enhance the ICM coverage measurement in the areas where DA is strong and the survey estimates have been weak--(1) the measurement of undercoverage of adult Black men and (2) the production of detailed age, sex, and race estimates that are both longitudinally and internally consistent. By integrating the DA results into the 2000 ICM, the age-sex-race differences between DA and survey estimates will be reconciled before producing the one-number census estimates, not after.

KEY WORDS: Demographic Analysis; Coverage Evaluation; Undercount.

1. INTRODUCTION

One of the goals of the 2000 census is to reduce the differential undercount and cost of the census with the use of sampling and estimation. We will use statistical techniques, and administrative records where possible, to estimate the number and types of people missed in the census. We will do this with the Integrated Coverage Measurement (ICM) program. The missed persons will be added to produce a "one-number" census total by the December 31, 2000 release deadline.

The ICM program was tested for the first time in the 1995 census test. We tested two coverage measurement techniques--CensusPlus (CP) and Dual System Estimation (DSE). These are sample survey-based estimates, involving case-by-case matching of persons in an independent survey with persons in the census (see Mulry and Singh, 1995, for a description of the CensusPlus and DSE methodology). In the 1990 coverage measurement program, the survey estimates were based on the Post Enumeration Survey (PES) (see Hogan, 1993). The 1990 PES used a dual-system estimation technique.

The Census Bureau has another coverage measurement and evaluation program--Demographic Analysis. Demographic analysis (DA) represents a macro-level approach to measuring coverage, where analytic estimates of net undercount are derived by comparing aggregate sets of data. The demographic approach differs fundamentally from the survey estimates, which represent a micro-level approach.

The method of demographic analysis relies heavily on aggregate administrative records, which are essentially independent of the census. The estimates for the population below age 65 are derived by the basic demographic accounting equation:

$$\text{Population} = \text{Births} - \text{Deaths} + \text{Immigrants} - \text{Emigrants}$$

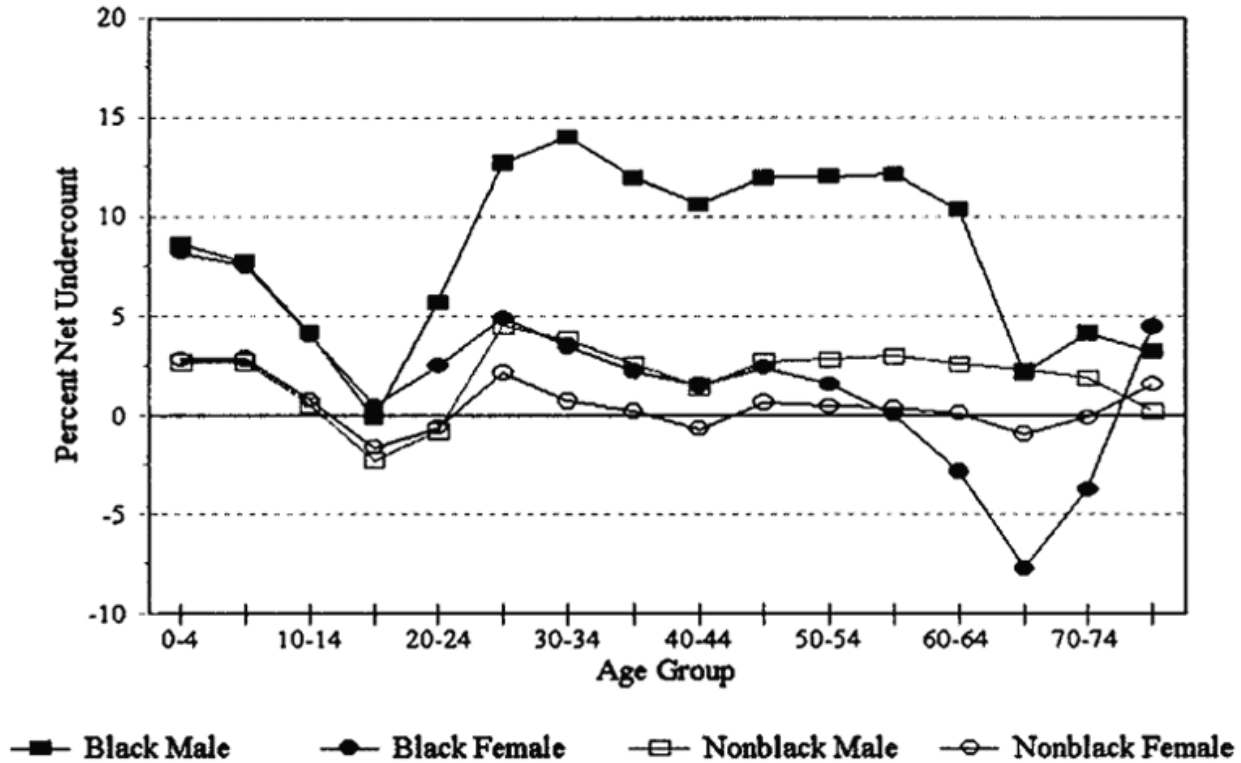
Aggregate Medicare data are used to estimate the population aged 65 and over.

The estimation process involves a number of assumptions about the completeness of the administrative data used to develop the demographic estimates. Also, since there are no records for some population groups such as undocumented immigrants, the size of some groups must be estimated (see Robinson et al, 1993, for a discussion of the 1990 demographic results; and Himes and Clogg, 1993, for an excellent overview of the use of demographic methods).

Demographic analysis as a tool for coverage evaluation has been well developed over time. The national demographic estimates have become the benchmark for assessing differences in coverage by age, sex, and race. Figure 1 displays demographic undercoverage rates for 1990--the figure shows the relative high undercount of Black children and adult Black men. The most notable pattern is the high levels of undercount of

Black men between ages of 25 and 64, where the estimated undercount ranges from 10 to 15 percent. A principal goal of Census 2000 is to reduce these differentials. In keeping with the theme of this conference, this is a nonsampling error we are trying to reduce in the census.

Figure 1. Percent Net Undercount by Race, Sex, and Age: 1990 DA



¹ J. Gregory Robinson, Chief, Population Analysis and Evaluation Staff, Population Division, United States Bureau of the Census, Washington, DC 20233-8800

2. THE POSSIBLE ROLES FOR DA IN 2000

2.1 Coverage Evaluation versus Coverage Measurement

How can demographic analysis be used in Census 2000, where we plan to release a "one-number" census? DA can serve one or two roles in 2000:

(1) As a coverage evaluation tool only, in support of the CensusPlus or DSE-driven ICM operations.

In 1990, DA served as a coverage evaluation program. It was used to evaluate the quality of the PES results, and provided important historical benchmarks (1940-1980) to assess completeness of coverage in 1990. Research conducted over the past four years demonstrates that DA can play an expanded evaluation role in the 2000 census, including the use of subnational DA benchmarks (see Robinson, 1994; and Robinson and Kobilarcik, 1995).

(2) As a coverage evaluation and an "active" coverage measurement program, where the demographic coverage estimates would be **integrated** with the CensusPlus or DSE survey estimates.

This "best set" would serve as the ICM standard for producing the one-number 2000 census products. In 1990, the PES estimates were used exclusively as the coverage measurement vehicle for any adjustment of the 1990 census counts. DA estimates were not used, because we believed the limitations of DA at that time (e.g., no geographic detail, uncertainty of the estimates) offset its strengths (e.g., independence, internal consistency).

2.2 Strengths and Limitations of DA

Should DA be only an evaluation tool in 2000--or should it also play an active integrated role in the ICM coverage measurement operations? These decisions will depend on how we can minimize its limitations and more clearly maximize its strengths. In the following review, we will identify where the strengths or limitations have changed since 1990 to build a stronger case for the integration of DA.

2.2.1 The Limitations of DA

1. **Lack of geographic detail**--Independent DA estimates in full age-sex-race detail are not available below the national level. For coverage measurement purposes in 2000, the survey-based estimates would remain the principal vehicle for the subnational ICM estimates.

Since 1990, extensive research has been conducted to develop "subnational" DA benchmarks of coverage, mainly for ages under 18 and 65+ for States and large county areas. For the younger ages, birth and death data are readily available and school enrollment data can provide an independent source for measuring the school-aged population and estimating migration. Administrative Medicare data are an excellent independent source for the population 65 and over. Further, sex ratio analysis provides clues about coverage differentials for ages 18-64 (see Robinson, 1994). Finally, we are developing a housing unit estimates program (for States and counties), which may ultimately be integrated with and strengthen the population estimates.

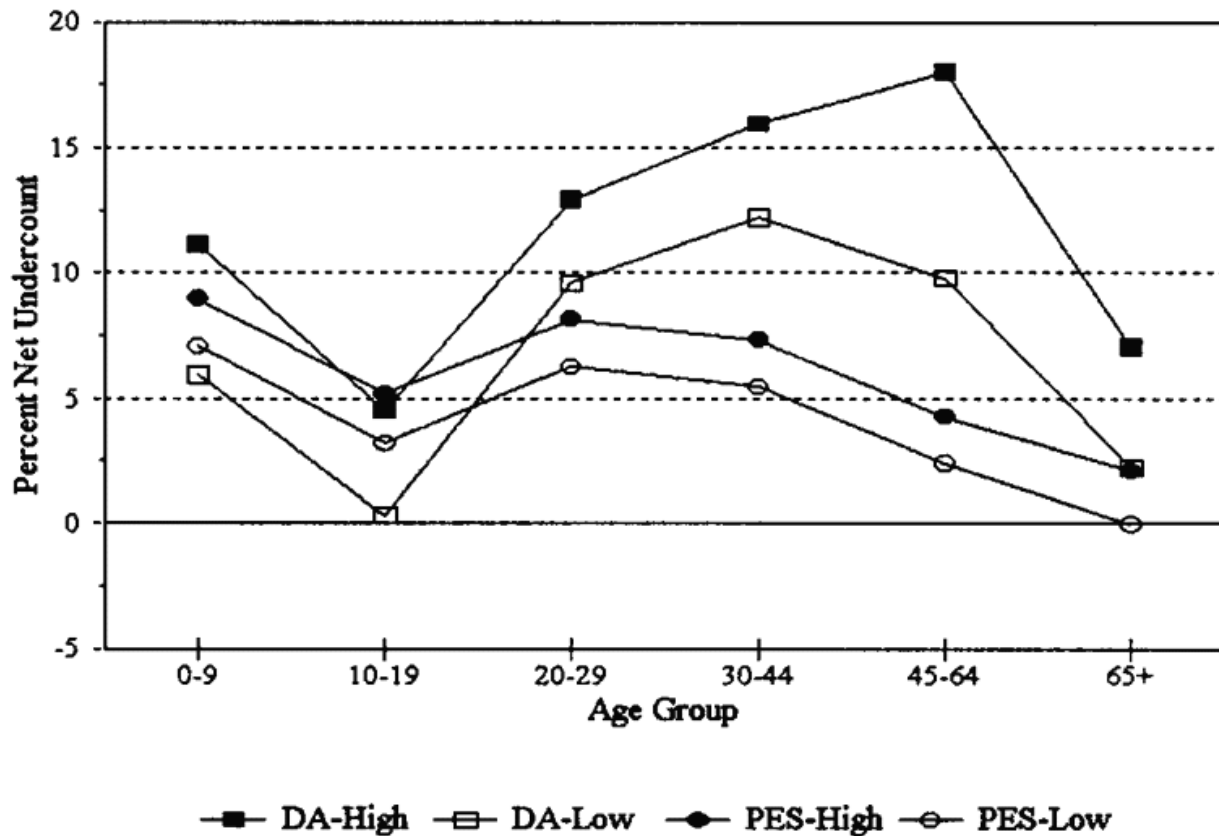
So there is a new geographic dimension to the demographic program, which can serve as an important evaluation tool in 2000 to compliment the survey-based ICM activities. We successfully used DA to evaluate the CensusPlus and DSE results in the 1995 test sites of Oakland, CA, Paterson, NJ, and six rural parishes in Louisiana (Robinson, 1996).

2. **Limited race/ethnic detail**--The principal DA race categories are Black and Nonblack. Although research is being conducted to produce DA estimates for Hispanics and Asians, these measures would not be as reliable as those for Blacks and Nonblacks. The CensusPlus/DSE would provide the coverage measurement standard for Hispanics, Asians, and American Indians (as well as important classifications by tenure).

3. **Inconsistencies in race classification**--The DA estimates of net undercount will be biased if persons who are classified as Black in DA are reported as another race in the census. We need to conduct more research to assess the degree of inconsistency and identify ways this "classification error" can be minimized. Also, the effect of a possible multiracial designation in the census race question for 2000 needs to be considered.

4. **Uncertainty in the DA estimates**--The principal concern regarding the DA estimates in 1990 was the uncertainty of the measured undercounts. For the first time, the 1990 DA estimates were accompanied by statistically-based measures of uncertainty (Das Gupta, 1991). The results demonstrated the DA estimates were subject to considerable uncertainty in measured undercount levels (see Figure 2 for 95 percent confidence intervals for the 1990 DA and PES undercount estimates of Black men). Nonetheless, it is clear that the demographic estimates of percent undercount for Black adult males remain relatively high under any reasonable "uncertainty" assumption. The "lowest" alternative estimate for Black males is above 8 percent for each broad age group between 20 and 64. And these lowest DA estimates were consistently higher than the comparable PES estimates that included uncertainty bounds (see Adlakha, et al, 1991).

Figure 2. Undercount Confidence Intervals, Black Males:1990 DA and PES



It is important to note that the DA estimates are subject to less uncertainty in terms of measuring differences in coverage according to age, sex, and race. This property--that demographic analysis provides better measures of coverage differences rather than absolute coverage levels--is attributable to the fact that many of the errors in the estimates are consistent and hence tend to "cancel" in comparisons across sex, race, and time. This particular strength could be exploited in 2000. For example, the DA sex ratios (ratio of males to females) are less error-prone than the DA undercount estimates themselves.

2.2.2 The Strengths of DA

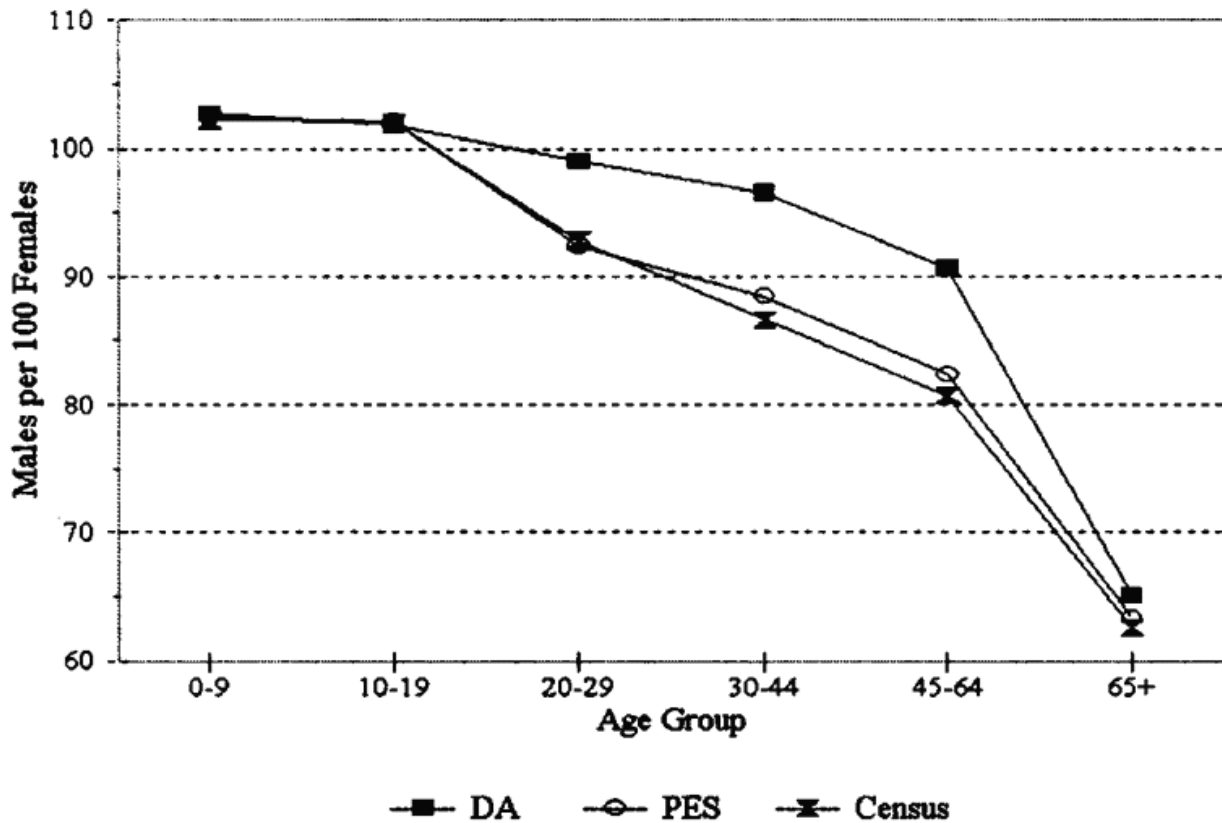
Demographic analysis possesses certain advantages over the survey-based approach that can be utilized in a comprehensive ICM system for 2000. Some of these strengths, while existing in the 1990 census setting, gain "standing" in the cost-conscious, one-number census environment of 2000:

1. **Low cost**--With the reduction of cost an important goal of the 2000 census, the relative low cost of the DA program becomes very attractive. DA is very inexpensive because it draws extensively from the Census Bureau's ongoing population estimates program. Even with a stepped-up research program, the DA method is much less expensive than the survey-based approach.
2. **Operational feasibility**--The DA method is battle-tested in previous censuses, with continued improvements in data and techniques and results available for review. The CensusPlus technique is still in the testing phase; in fact, it encountered unforeseen problems in the 1995 test. The DSE approach faces a very tight critical path to produce results by the December 31, 2000 deadline. The independent, administrative record-based DA estimates would provide a back up if the CensusPlus or DSE encounters problems.
3. **Timeliness**--Since field operations or census matching aren't involved, the DA estimates will be available in 2000 before the CensusPlus or DSE. First, independent housing unit benchmarks could evaluate completeness of the Master Address File even before the 2000 census begins. Second, DA population estimates can give important readings on the differential undercount in the "pre-ICM" counts (e.g., July-August of 2000). For example, the indication from low sex ratios of large relative undercounts of adult Black men (like in previous

censuses) would stress the importance of the ongoing ICM operations. Of course, the DA estimates will also be available to immediately evaluate the survey estimates when those are ready (October-November 2000).

4. **Independence**--Since DA is based largely on aggregate administrative records, it provides an independent basis to validate the ICM survey estimates. In 1990, the independent DA undercount estimates (1.85 percent) were used to validate the overall PES estimate (1.58 percent). The detailed DA estimates indicated, however, that the PES significantly understated the net undercount of adult Black men--the well-known "correlation bias" problem. For example, Figure 3 shows how the 1990 PES sex ratios for Blacks are much closer to the implausible census sex ratios than to the DA ratios. Even after taking into account the measured uncertainty of the DA and PES estimates, the DA sex ratios are significantly higher than the PES or census ratios.

**Figure 3. 1990 Expected Sex Ratios:
Comparison of DA and PES to Census**



For 2000, we are looking at ways to integrate the DA results (such as sex ratios) in the ICM to minimize this problem. Here, DA would clearly be serving a dual coverage evaluation and coverage measurement role (see Wolter, 1990, and Bell, 1993, for research on the use of DA sex ratios in coverage estimation).

5. **Internal consistency**--The foundation of the demographic method is the logical and longitudinal consistency of the underlying demographic data. DA follows the demographic process of population change as it occurs, starting with births, then incrementing or decrementing cohort size with subsequent information on mortality and net migration. The estimates created for 2000 from this process will be longitudinally and internally consistent. The time series linkage of the DA estimates (for multiple censuses) provide a consistent basis to assess the plausibility of the demographic estimates themselves. On the other hand, the survey estimates have no longitudinal dimension and cannot check for both longitudinal and cross-sectional consistency.

One distinct advantage of the DA method in this regard is that it provides single-year of age estimates. The administrative data for DA is virtually complete (no samples involved) and available annually (e.g., births, deaths, immigration data). The demographic process automatically produces single-year of age estimates. The survey estimates are necessarily based on sample data, which will compromise the quality of detailed age estimates. Among other uses, accurate single-year data are an important ingredient for the Census Bureau's annual

population estimates program. The quality of the ICM age data for 2000 could be enhanced if the DA estimates were integrated in the coverage measurement process.

6. **Historical benchmarks**--A major goal of the 2000 census is to reduce the differential undercount. The DA estimates provide the only consistent historical series of detailed age-sex-race undercount factors to document the possible reduction of undercount in 2000 compared to earlier censuses. The survey estimates do not have this historical dimension. Further, the detailed 1990 PES estimates for Blacks are flawed for the purposes of making valid 1990-2000 comparisons (e.g., the PES sex ratios for Blacks are implausible compared to the DA estimates).

3. DISCUSSION

In designing a comprehensive Integrated Coverage Measurement system for the 2000 census, we need to balance the strengths and weaknesses of DA and survey-based techniques. Clearly, demographic analysis should play an important role in the evaluation of the census and the ICM operations. The independent demographic estimates will be available on a timely basis to take multiple readings on coverage patterns, before and after the ICM. And it can do this at a relative low cost.

The question is: Should we take the next step forward and formally integrate demographic analysis into the ICM coverage measurement process? In particular, can we enhance the ICM estimates in the areas where DA is strong and the survey estimates have been weak?--(1) the measurement of undercoverage of adult Black men and (2) the production of detailed estimates by age, sex, and race (Black/Nonblack) that possess the demographic properties of longitudinal and internal consistency. By integrating the DA results into the 2000 ICM, the age-sex-race differences between the DA and survey estimates will be reconciled before producing the one-number census estimates, not after.

We are developing a research agenda that will spell out how DA can be integrated in the ICM. This agenda also documents the research tasks needed to improve the basis DA estimates themselves. Our goal is clear: To selectively and creatively draw from the unique strengths of demographic analysis to enhance the survey-based ICM estimates used in the final one number census counts for 2000.

4. REFERENCES

Adlakha, Arjun, Hogan, Howard, and Robinson, J. Gregory. "A Report on the Internal Consistency of the Post-Enumeration Survey Estimates," *1990 Coverage Studies and Evaluation Memorandum Series S-1*, U.S. Bureau of the Census, 1991.

Bell, William R. 1993. "Using Information from Demographic Analysis in Post-Enumeration Survey Estimation," *Journal of the American Statistical Association*, Vol. 88, No. 423, P. 1106-1118

Das Gupta, Prithwis. 1991. DA Evaluation Project D10: "Models for Assessing Errors in Undercount Rates Based on Demographic Analysis." *Preliminary Research and Evaluation Memorandum No. 84*, U.S. Bureau of the Census.

Fernandez, Edward W. 1995. "Using Analytic Techniques to Evaluate the 1990 Census Coverage of Young Hispanics," *Technical Working Paper Series No. 11*, Population Division, U.S. Bureau of the Census.

Himes, Christine L. and Clogg, Clifford C. 1993. "An Overview of Demographic Analysis as a method for Evaluating Census Coverage in the United States," *Population Index*, 58(4): 587-607, Winter 1992.

Hogan, Howard. 1993. "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of the American Statistical Association*, Vol. 88, No. 423, P. 1047-1060.

Mulry, Mary H. and Rajendra P. Singh. 1995. "Development and Evaluation of Census Methodology for 2000 Census," *Proceedings of the International Conference on Survey Measurement and Process Quality*, Bristol, United Kingdom, April 1-4.

Robinson, J. Gregory Robinson and Edward L. Kobilarcik. 1995. "Identifying Differential Undercounts at Local Geographic Levels: A Targeting Database Approach." Paper presented at the Annual Meeting of the Population Association of America. San Francisco.

Robinson, J. Gregory. 1994. "Use of Analytic Methods for Coverage Evaluation in the 2000 Census." Paper presented at the Population Association of America, Miami, May 5-7.

Robinson, J.G., Ahmed, B., Das Gupta, P., and Woodrow, K.A. 1993 . "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis," *Journal of the American Statistical Association*, Vol. 88, No. 423, P. 1061-1071.

Wolter, Kirk M. 1990 "Capture-Recapture Estimation in the Presence of a Known Sex Ratio, *Biometrics*, Vol. 46, P. 157-162.

Author: J. Gregory Robinson
Created: May 10, 2000

Source: U.S. Census Bureau, Population Division
[Questions?](#) / 1-866-758-1060