Advances in the Use of Capture-Recapture Methodology in the Estimation of U.S. Census Coverage Error

Mary H. Mulry and Vincent T. Mule Jr.

Abstract A post-enumeration survey (PES) is an important tool for assessing the quality of a census and gaining information about how to improve census-taking methodology. The U.S. Census Bureau has implemented a PES to evaluate the coverage error in each U.S. Decennial Census since 1980. A PES uses a second enumeration implemented on a sample basis after a census and subsequently matched to the census using a combination of computer and clerical matching. Then, dual system estimation may be used to estimate the population size. The difference between the PES estimate of the population size and the census total yields an estimate of the net undercount. This chapter focuses on the methodology and estimation of net coverage error in the 2010 Census produced by the 2010 PES. The evaluations of U.S. censuses continue to use the PES methodology to evaluate the coverage of the decennial census. These implementations of the PES have built on the quality control methodology that Dr. Stokes developed for the 1990 PES.

1 Introduction

A post-enumeration survey (PES) is an important tool for assessing the quality of a census and for gaining information about how to improve census-taking methodology. The U.S. Census Bureau has implemented a PES to evaluate the coverage error in each U.S. Decennial Census since 1980. There are two types of coverage error. One type is *overcount*, which occurs when an enumeration is inappropriate, such as entries that are duplicates of other enumerations, for people

Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau

M. H. Mulry (✉) · V. T. Mule Jr.
U.S. Census Bureau, Suitland, MD, USA
e-mail: mary.h.mulry@census.gov

H. K. T. Ng, D. F. Heitjan (eds.), *Recent Advances on Sampling Methods and Educational Statistics*, Emerging Topics in Statistics and Biostatistics,
https://doi.org/10.1007/978-3-031-14525-4_5

born after Census Day or for people who died before Census Day. The other type is *undercount*, which occurs when a person who should be counted in the census is not enumerated. The *net coverage error*, which equals the *overcount* minus the *undercount*, provides a measure of the quality of a census.

A PES uses a second enumeration implemented on a sample basis after a census and subsequently matched to the census using a combination of computer and clerical matching. Then, dual system estimation, which is another name for capture-recapture estimation, may be used to produce an estimate of the population size. The difference between the PES estimate of the population size and the census total yields an estimate of the net coverage error. A PES that uses dual system estimation essentially applies a variation of the "capture-recapture" methodology designed for estimating the size of wildlife populations to human populations.

This chapter focuses on the methodology and estimation of net undercount in the 2010 Census produced by the 2010 PES. The data collection methods included new quality control procedures and an estimation approach that differed from the estimation used in the prior PES programs conducted from 1980 through 2000. The implementation of the 2020 PES used essentially the same methodology for data collection and estimation as that employed for the 2010 PES. However, the COVID-19 pandemic resulted in some unexpected delays in the 2020 PES data collection and processing. As a result, the estimates from the 2020 PES will not be available in time to meet the publication deadline for this volume.

Dr. S. Lynne Stokes contributed to the methodology for data collection and estimation for the Post-Enumeration Survey at different points in her career. The discussion of the PES methodology and the evolution of its implementation to evaluate census coverage at the U.S. Census Bureau will include descriptions of her contributions.

The discussion in this document focuses on the evolution of design of the PES as implemented to evaluate the coverage of the decennial censuses conducted from 1980 to 2010. These topics include the following:

- Section 1 is the introduction to the document.
- Section 2 has a brief overview of the recognition that there was a need to evaluate the coverage of the U.S. Censuses.
- Section 3 contains a description of the dual system estimator (DSE) that is used in estimating the coverage error in censuses, including the first attempt to implement a PES aimed at evaluating the coverage of the 1980 Census.
- Section 4 describes the 1990 PES and the role Dr. Stokes played in the evaluation program that informed a decision on whether to use PES estimates to adjust the 1990 Census for coverage error.
- Section 5 explains how the 2000 PES was designed to evaluate the coverage of the 2000 Census and describes evaluations used in the decision on whether the 2000 Census should be adjusted for coverage error.
- Section 6 discusses methodological challenges in the 2010 PES.

- Section 7 describes current research on developing methods for replacing data collected in PES fieldwork with administrative records and third-party data for the US population in DSEs to produce census coverage estimates.
- Section 8 is a summary.

2 Background

The U.S. Constitution requires that a census of the U.S. population be conducted every ten years for the purpose of the apportionment of seats in the House of Representatives among the states. Article 2, Section 2 of the Constitution, states that the "actual enumeration" be used to allocate the seats among the states. The current apportionment method, which was chosen by the House of Representatives, is the Method of Equal Proportions, but other methods have been used over the years (Spencer, 1985).

The first U.S. Census was conducted in 1790. As Secretary of State, Thomas Jefferson's duties included certifying the 1790 Census data. Even though Secretary Jefferson certified the census count, both he and President Washington thought the 1790 Census had undercounted the U.S. population by several hundred thousand (U.S. Census Bureau, 2021a). For years, the prevailing attitude was that the census provided the best information about the size and distribution of the U.S. population. And, even if the census was not perfect, certainly it had better coverage of the population than any source of administrative records available at the time.

New information about the coverage of the census appeared in the early 1940s when the Census Bureau conducted a study that compared the number of males of military age in the 1940 Census to the number found in draft registration records. The study used the demographic method of comparing aggregated totals constructed by a clerical operation. The study estimated that there were 14.9% more Black males of 21–35 years of age registered for the draft than were counted in the census and 2.8% more non-Black males in the same age category (Price, 1947).

This result led to the development of census coverage evaluation methodologies, the first one being Demographic Analysis. The estimates produced by Demographic Analysis are a sum of totals for subpopulations based on aggregating administrative records from different record sources, such as birth and death records, to form an estimate of the total population that can be compared to the total from a census. The 1950 Census was the first census to have its coverage evaluated using Demographic Analysis (Coale, 1955). Demographic Analysis has been used to evaluate the coverage of every U.S. Census at the national level since 1950 and is still used today although the method and data sources have improved over the years. Demographic Analysis does not produce estimates for subnational geographic areas such as states and has limited race results since it uses historical data sources.

The need for estimates of census coverage for geographic and demographic subgroups led to the development of two other methods. One is the Post-Enumeration Survey (PES) used by the USA and several other countries (Mulry, 2014). The other is the reverse record check, developed by the Statistics Canada which relies on an administrative record system that is updated on an ongoing basis between censuses (Statistics Canada, 2007, 2021). One of the innovations in the estimation of the coverage of the 2010 Census came from the PES estimation procedure incorporating the Demographic Analysis results for some hard-to-count subgroups. Section 6 gives the details.

3 Post-Enumeration Survey in 1950, 1960, and 1980

This section provides an overview of the Census Bureau's initial attempts to implement the PES to evaluate the coverage of the 1950, 1960, and 1980 Censuses. A more detailed discussion appears in Mulry (2012).

A Post-Enumeration Survey (PES) is a survey conducted after a census for the purpose of evaluating the coverage of the census. The first PES in the USA was conducted after the 1950 Census and was motivated by the undercount of draft-age males discovered in the 1940 Census. A PES uses two systems, which may be samples. The Census Bureau's implementation uses samples where one is a sample of the population, called the P sample, and the other is a sample of census enumerations, called the E sample. The basic strategy is that enumerators conduct interviews at the addresses in the P sample that include collecting the current household roster along with characteristics and where each person resided on Census Day plus a roster of the people living at the address on Census Day. Then a clerical operation matches the people on the P sample roster at each address in the P sample to the Census enumerations in two phases. In the first phase, those in the P sample that match to a census enumeration at the reported Census Day address receive a status of Match. When the matching operation cannot decide, the person receives a status of Unresolved, and the form is sent back to the field for interviewers to collect more information. The E sample enumerations also receive one of three statuses, Correct Enumeration, Erroneous Enumeration (if person was not a resident at the address on Census Day), or Unresolved. When P sample and E sample people receive an Unresolved status, their forms are sent for further fieldwork to determine each person's Census Day address. If the interviewer conducting the second interview is unable to determine where the person lived on Census, the person retains the status of Unresolved. Each census enumeration that retains an Unresolved status receives an imputed probability of being a Correct Enumeration, and P sample people with an Unresolved status receive an imputed probability of being a Match.

The methodology for collecting and processing the data that the PES collects has evolved over the years. The changes include almost all aspects, such as how the samples are selected, how the P and E sample interviews are implemented, the use of technology, and the estimation approach. Section 3.1 contains a short discussion of the Census Bureau's first attempts in 1950 and 1960 to conduct a PES, and Sect. 3.2 discusses the implementation of the 1980 PES to evaluate the 1980 Census. The

Census Bureau did not conduct a PES after the 1970 Census. The Census Bureau did implement a PES after the 1990 Census and subsequent censuses, and Sects. 4, 5, and 6 contain discussions of these implementations.

3.1 PES in 1950 and 1960

The first attempt to implement a PES was aimed at evaluating the coverage of the 1950 Census (Marks et al., 1953). The P and E samples each had about 25,000 housing units and were selected in a manner that resulted in the chosen areas overlapping as much as possible to reduce the expense of the data collection. The strategy was for the P sample interview to be of much higher quality than the census interview so that the error could be estimated by comparing the results of the P sample to the census results in the E sample. When the P sample results did not agree with the census for the same housing unit, interviewers were sent to collect information to resolve the discrepancies so that errors in the P sample could be identified. Then the corrections could be incorporated into the results of the clerical matching operation.

The strategy relied on these procedures discovering the truth in the sample areas. Then an estimate of the population size could be formed by multiplying the total census count by the ratio defined by the total number of people in the P sample housing units divided by the total number of people in the census in the same housing units as shown below:

$$True\widehat{Population} = (Census\ Count)\,x\,\frac{\text{number of people in P sample in P} - \text{sample housing units}}{\text{number of people in census in P} - \text{sample housing units}} \tag{1}$$

Unfortunately, the results failed to meet the Census Bureau's quality standards. The PES estimate of population size was lower than the estimates derived from demographic methods (Coale, 1955). The PES estimate of undercoverage was 2.1 million persons, which was 1.4% of the enumerated population, while the demographic method estimated the undercoverage to be 5.4 million which was 3.6% of the enumerated population. The Census Bureau's analyses found that the "minimum reasonable estimate" of undercoverage was 3.7 million which was 2.5% of the enumerated population. Subsequent analyses performed in preparation for evaluating the coverage of the 1960 Census found weaknesses in both the PES data collection and estimation and also in some of the assumptions used in producing the demographic estimates (Marks & Waksberg, 1966). Another concern about the 1950 PES was that some PES interviewers did not follow instructions completely. The interviewers were given a sealed census roster for each address. The interviewer's instructions were to open the envelope after completing the PES interview and compare the new roster with the census roster. Then, while still on the doorstep, the interviewer could identify differences and ask questions to identify errors in one or both rosters. However, there were reports that some interviewers did not ask for a household roster on Census Day but only opened the census roster and verified it.

A second attempt to implement a PES was aimed at evaluating the coverage of the 1960 Census. However, the design had P and E samples that were selected independently and retained the assumption that the P sample interview would be more accurate than the census responses in the E sample. The outcome of the 1960 PES also was not satisfactory. Reminiscent of the results from the 1950 PES, the 1960 PES estimates of population size were lower than the national-level estimates derived from demographic methods (Marks & Waksberg, 1966).

3.2 PES 1980 and Dual System Estimation

The Census Bureau introduced a new design for a PES to evaluate the 1980 Census. The 1980 PES implemented a new estimation method called dual system estimation, which led to a new design for sample selection.

3.2.1 Dual System Estimation

A major part of the new design was using dual system estimation (DSE) which did not require the assumption that the data collected for the P sample was without error (Chandrasekar & Deming, 1949). The method had been used in programs sponsored by the United Nations (UN) that focused on estimating population size in other countries. Implementing the DSE, which is another name for capture-recapture, required only that the P sample be a second enumeration of the population as opposed to being a near-perfect enumeration that was required for the estimation approach used in the 1950 and 1960 PESs. The estimation approach used post-stratification, not the log-linear form of the estimator used in some applications of capture-recapture methods.

Data collection for the P and E samples must satisfy four basic assumptions (Chandrasekar & Deming, 1949). One is that selection for inclusion in the P sample is independent on selection for inclusion in the E sample. This assumption means that the census and the P sample could not share data or information. For example, a census interviewer who also worked on the data collection for the P sample had to work in areas that were not included in the interviewer's census assignments. Second, the probability of being included in the census is not correlated with being included in the P sample. Third, each individual is unique, and records for the individual can be identified on both lists without error. And fourth, there are no spurious events in the E sample list or the P sample list, which for the Census Bureau's PES means that there are no sample records that are duplicates, nonexistent, or not in the population of interest (Mule, 2008).

When the four assumptions hold, the following two ratios of expected values are equal. The ratio on the left is based on the E sample and the ratio on the right is based on the P sample. In some capture-recapture applications, the ratio on the right in Eq. (2) is called the *detection probability*:

$$\frac{E\left(\widehat{Number\ of\ correct\ census\ enumerations}\right)}{E\left(\widehat{Population\ size}\right)} \sim \frac{E\left(\widehat{Number\ of\ matched\ people}\right)}{E\left(\widehat{Number\ of\ survey\ enumerations}\right)} \tag{2}$$

A *correct census enumeration* is one where the person is enumerated at the address where the person lives and sleeps around Census Day, which is April 1 of the census year. The enumeration also is required to be *data-defined*, which means that the record has enough information to identify the person uniquely. An enumeration is classified as data-defined if it has two or more characteristics, one of which may be a name. However, sometimes a data-defined enumeration cannot be uniquely identified, such as when an enumeration with the minimum information has characteristics that are common in their area. A *matched person* is one that has a record in the P sample that can be matched to the person's census enumeration.

Using Eq. (2) and algebra, an estimator of the population size can be constructed as follows:

$$\widehat{Population\ size} = (\widehat{Number\ of\ correct\ census\ enumerations})\frac{(\widehat{Number\ of\ survey\ enumerations})}{(\widehat{Number\ of\ matched\ people})} \tag{3}$$

One aspect of using samples is the need to include both small and large subpopulations, such as race and Hispanic ethnicity groups, and geographic areas such as states and metropolitan areas. Therefore, the sample selection probabilities will be higher for smaller population groups than for the larger groups. The estimation needs to account for the variation in the selection probabilities by incorporating sampling weights equal to the inverse of the selection probabilities.

The formula for the DSE based on samples uses the same formula as in Eq. (3) with the addition of a ratio adjustment of the estimated number of correct enumerations to the number of data-defined enumerations. However, the inclusion probabilities are not equal throughout the population, which affects whether Eqs. (2) and (3) hold. The remedy is to partition the population into groups where the inclusion probabilities are believed to be equal or nearly so. The groups are called *poststrata* (indexed by *j*), and the post-stratified estimator for an area *C* is shown below. The *data-defined census enumerations* are those that have enough information for the matching operation to identify them if they are in the P sample. An enumeration is classified as data-defined if it has two or more characteristics, one of which may be a name. Enumerations that are not data-defined remain in the census but are excluded from the E sample. Therefore, the matching is a three-step procedure where the first step determines if the census enumeration is data-defined, and for those that are, the second step identifies the ones that are correct enumerations, and the third step determines if the P sample person matches to a census enumeration.

The formula for the post-stratified DSE estimate of the population size for an area C, $\widehat{TOTAL_C}$, when using J poststrata is as follows:

$$\widehat{Total}_C = \sum_{j \in J} CEN_{Cj} \left[\frac{DD_j}{CEN_j} \frac{\widehat{CE}_j \Big/ \widehat{ETOT}_j}{\widehat{M}_j \Big/ \widehat{PTOT}_j} \right] \tag{4}$$

where

CEN_{Cj} = number of census enumerations in poststratum j in area C
CEN_j = number of census enumerations in poststratum j
DD_j = number of data-defined enumerations in the census in poststratum j
$\widehat{ETOT_j}$ = estimated number of data-defined enumerations in the E sample in poststratum j
$\widehat{CE_j}$ = estimated number of correct enumerations in the E sample in poststratum j
$\widehat{M_j}$ = estimated number of P sample people in poststratum j that match a census enumeration in the correct location
$\widehat{PTOT_j}$ = estimated number of people in the P sample in poststratum j

3.2.2 1980 PES

The E and P samples for the PES in the 1980 Post-Enumeration Program (PEP) were both nationwide samples that were selected in completely different ways. For the 1980 PES, the P sample used the April and August waves of the current population survey (CPS) which is an ongoing nationwide survey that measures unemployment and is conducted separately from the census. The combination of the two CPS waves included about 124,000 housing units with about half from each wave. The P sample questions appeared on a supplementary questionnaire that was administered after the CPS questions and asked who resided at the address on Census Day. The E sample was constructed by selecting 10 housing units from each enumeration district in the USA which resulted in a sample size of about 110,000 (Fay, 1988). Interviewers visited each housing unit in the E sample and verified that each person listed on the census questionnaire for the address was a resident on Census Day. If the people listed on the census questionnaire had moved, the interviewer sought information about them from neighbors and at the post office (Mulry, 2012).

The clerical matching of the two samples to the census to determine who should be on each list was cumbersome and time consuming. The census file and both sample files needed to be available for the matching. The matching for those who moved between the census and the PES interviews was exceptionally time consuming.

The results of the 1980 PES showed some undercount, but there was a controversy over the best way to construct the estimate of the net undercount (U. S Census Bureau, 1980). Some statisticians inside and outside the Census Bureau were not confident that the implementation of the 1980 PES satisfied the assumptions underlying the DSE. There was a concern that the estimates based on the DSE were affected by correlation bias, so analyses assessed the impact of some of the assumptions by constructing 12 sets of estimates. In the end, the preferred set of PES estimates of net undercount were 1.0% for the USA, 5.7% for Blacks, 4.5% for non-Blacks, and 0.0% for others ((U. S Census Bureau, 1980), p. 9–10). Another concern was that the estimates of net undercount at the national level based on the PES were lower than the estimate from Demographic Analysis which was 1.2% for the USA. Other estimates of net undercount from Demographic Analyses were 4.5% for Blacks and 0.8% for non-Blacks (Long et al., 2003). The estimated net undercount prompted a call to adjust the 1980 Census for the undercount using the 1980 PES data. The Census Bureau opposed adjusting the 1980 Census using PES data and stated so in an announcement. Detroit, New York City, and the State of New York filed a lawsuit asking that the Census Bureau be ordered to adjust the 1980 Census for undercount. These lawsuits were consolidated to the court hearing the New York case. The judge ruled that the Census Bureau's decision was not arbitrary and capricious. Therefore, in the end, the 1980 Census was not adjusted (U.S. Census Bureau, 2021).

4 1990 PES

In the aftermath of the 1980 Census, the Census Bureau decided to prepare in a manner that would enable an adjustment of the results of the 1990 Census if such an adjustment was deemed necessary. The preparations included a research and testing program during the decade leading up to the 1990 Census. The program incorporated test censuses during the decade and a dress rehearsal in 1988, each including a PES. The testing program facilitated refining PES data collection, processing, and estimation methodology.

One of the components of the research and testing program for the 1990 Census was the development of computer matching software that could be used in matching the P sample records to E sample records. The goal was to improve the quality of the matching and to produce the matching results faster than was possible with clerical matching. The development of the new matching software leveraged methodology developed by Fellegi and Sunter (1969) for matching records (Jaro, 1989; Kelley, 1986). In addition, clerical staff conducted a quality control operation on a sample of the computer matching results to assure accuracy.

Another component of the research and testing program was the development of methods to assess the quality of the PES estimates which could be used to evaluate their suitability for adjusting the 1990 Census for undercount. See Hogan (1993) for an overview of the 1990 PES methodology and Belin et al. (Belin et al., 1993) for a discussion of the new approach to imputing enumeration status using hierarchical modeling.

4.1 Dr. Stokes's Contributions to Interviewer Quality Control

Dr. Stokes has made significant contributions to the study of interviewer variance and bias. Interviewer effects on data collected in censuses and surveys can be substantial. Interviewer variance was a major reason the 1970 Census started the collection of census data by mail instead of personal interview (Stokes & Mulry, 1987).

Her interest in interviewer effects and quality control started when she worked at the Census Bureau early in her career and continued during her career as an academic. Initially, her work at the Census Bureau focused on optimizing the design of quality control samples to detect interviews fabricated by interviewers (Biemer & Stokes, 1989).

Fabrications of interviews during the data collection for the estimation of census coverage error is particularly important. A reason is that one of the assumptions underlying the DSE in Sect. 3.2.1 states that the E sample list and the P sample list used in estimation do not contain spurious records, such as fabricated records. When the assumption of no spurious events in data holds, the relationship in Eq. (2) that underlies the DSE holds. Interviewing quality control is therefore essential.

Another reason that the detection of fabricated interviews is important is because the quantity being measured is very small. A relatively small number of errors have the potential for a substantial impact on the estimate. For the past eight censuses, the Census Bureau has measured an error in the census count. For example, the Census Bureau estimates that the 1990 Census count for the population was 1.6% too low and the count for Blacks was 4.4% too low. This type of difference in accuracy is called the *differential undercount*. The differential undercount is important because key uses of the census data are for fixed-sum distributions such as the apportionment of Congress, the drawing of districts for state legislatures, and the federal fund allocation programs.

When Dr. Stokes started research in nonsampling error measurement in surveys, one of her concerns was that estimation of the correlated component of response variance usually assumed a normal distribution whereas most survey data were categorical. The paper "Estimation of the Correlated Component of Response Variance for Categorical Variables" (Stokes & Mulry, 1987) subsequently showed that the assumption could cause substantial underestimation of the sample size during the design of a study to measure the effect of interviewers.

Through continuing research on interviewer effects, Dr. Stokes made significant contributions during the consideration of adjusting the 1990 Census numbers for undercount. In addition, she provided the technical expertise for the evaluation of the effect of interviewer fabrication on the quality of the estimates of undercount. This role was the culmination of the research she conducted under contract with the Census Bureau.

Her work during the evaluation of the 1990 PES focused on assessing the assumption of no fabrication of interviews in the PES data. Despite an elaborate quality control program for the interviewing of the 1990 PES, some fabrication of interviewers was detected during the research studies leading up to the 1990 Census (Stokes & Jones, 1989). One result of the research was that the single-person households were the ones most likely to have fabricated interviews. The rationale was that since there was only one household member, these were the most difficult addresses to find someone at home. Thus, interviewers would make several attempts to make contact, but if they were unsuccessful, out of frustration, the interviewer would use the name on the mailbox and fill in the rest of the information. Therefore, one recommendation at the end of the quality control evaluation was that one-person households be checked at a higher rate than households with more than one person.

4.2 Outcome of the 1990 PES

Estimates of the net undercount in the 1990 Census were not used to adjust the census counts although there was litigation that reached the Supreme Count. In 1999, the U.S. Supreme Court ruled in an opinion written by Chief Justice William Rehnquist that the census numbers used for the apportionment of seats in the House of Representatives could not be based on samples because the Constitution required using the "actual enumeration" from the census (Department of Commerce vs United States House, 1999). However, the opinion did not prohibit adjusting the census numbers for other uses. See Prewitt (2012) for a brief discussion of the implications of the decision.

The 1990 PES estimated the net undercount of the U.S. population to be about 1.6% or about 4.0 million people. The estimate of the net undercount for Whites was about 1.8 million people while the net undercount rate for Blacks was about 1.4 million. But because the Black population was far smaller than the White population, the percent net undercount rate of 4.4% for Blacks was higher than for the 0.9% undercount rate for Whites (U.S. Census Bureau, 2021b).

5 2000 Census Accuracy and Coverage Evaluation

After the controversy over the possibility of an adjustment of the 1990 Census counts, the Census Bureau decided to create a process for deciding in March 2001 whether the coverage error in the 2000 Census numbers warranted an adjustment for use in redistricting. The Supreme Court made its decision that prohibited adjustment for redistricting in 1999, but the planning and research for the 2000 Census had started several years earlier. These preparations continued after the Supreme Court decision in 1999 because the census numbers may have needed an adjustment for use in other Census Bureau programs. The work included developing a process and predefined criteria for deciding whether adjustment was appropriate. An evaluation program, called the 2000 Census Accuracy and Coverage Evaluation (ACE), was designed to collect and analyze data that would inform the decision. The ACE included a post-enumeration survey and other analyses.

Dr. Stokes served on the panel "Measuring a Changing Nation: Modern Methods for the 2000 Census" that was convened by the National Academy of Sciences. The panel reviewed the Census Bureau's plans for the 2000 Census and the results of Census Tests. These plans included incorporating applications of new technology in several operations. The Census Bureau sought review and advice concerning the performance of the new technology used in the collection and processing of census data and coverage measurement data in the Census Tests conducted in preparation for the 2000 Census.

Advances in technology enabled innovations in the Census Bureau's collection, processing, and analysis of the census and ACE data to be completed in time to make an adjustment decision in March 2001. Much of the technology had been available previously but had not developed to the point where census planners could count on it for implementation and processing on the large scale and short time frame required to collect data from the 115.9 million housing units in the USA (Woodward & Damon, 2001). In 2000, all census response forms, both mail and Nonresponse Followup (NRFU) operation, were scanned by optical character and mark recognition technologies and converted to electronic format for processing (Kline, 2004).

The ACE interviewers used laptop computers when collecting the PES data. Addresses for each P sample block cluster were loaded into the laptop for the interviewer assigned to the area. Interviewers then used the laptop computers to collect data from respondents. The laptops contained the entire questionnaire, and interviewers were able to transmit the collected data electronically to the processing center. The laptops enabled faster processing and analyses of the data than was possible for the previous PES implementations that used paper questionnaires followed by a keying operation.

The interviewing quality control operation also used laptops. An advantage was that the original census responses for an address could be loaded into the interviewers' laptops. After conducting a quality control interview, the interviewer was able to push a button, and the laptop would present a comparison between the census household roster and the Census Day roster provided during the quality control interview. If there were differences in the two rosters, the quality control interviewer was able to ask questions to resolve any issues while still with the respondent.

The biggest surprise from ACE was the discovery that the estimated number of duplicate enumerations in the 2000 Census was much higher than expected. In addition, mail returns that were thought to be the best responses were included in some of the duplicates that were detected. Another finding was that the duplication occurred more frequently among household members under 30 than over 30. Examples of duplicate enumerations include the following: (1) college students being counted at both their college address and their parents' address, (2) children whose parents are divorced being counted at both parent's addresses, and (3) people who move around Census Day (April 1) being counted at both their old and new addresses since census data collection for NRFU goes into summer.

The discovery of the problem with duplication occurred during implementation of the process the Census Bureau had set up to arrive at a decision on whether to adjust the 2000 Census numbers issued for a purpose other than redistricting the seats in the House of Representatives. Further investigation found that a substantial number of erroneous enumerations had gone undetected in the processing of the ACE.

The Census Bureau continued to study whether to incorporate an adjustment to the census numbers that would be used in producing the intercensal estimates and other census products. The focus was on creating another revision, called ACE Revision II, that would be based on additional research concerning the level of duplication and the possibility of an adjustment for correlation bias in the DSE. At this point, a research project, called the Statistical Administrative Records System (StARS), created with the Census Bureau's newly developed administrative records database methodology, had progressed to the point of being useful in detecting census duplicates without fieldwork. StARS was able to create a database that covered the U.S. population by merging federal administrative records. Linking E sample and P sample records to StARS aided in identifying duplicates and other enumeration errors.

In the end, ACE Revision II estimates included several adjustments. The research with StARS and a clerical matching project produced an estimate of 5.8 million duplicates that was the basis of one of the adjustments of ACE Revision II (Mulry et al., 2006). Because the Demographic Analysis estimates produced a ratio of males to females that was higher than observed in the ACE, an adjustment for correlation bias was included in the ACE Revision II estimates. The correlation bias adjustments were created separately for Blacks and non-Blacks within three age categories: 18–29, 30–49, and 50 and over. However, an adjustment was not included for non-Black males 18–29 years of age because the data did not support the estimation in this category. More details about the adjustment may be found in (Bell, 1993, 2001). In addition, errors found during several evaluation studies were corrected in the data used in forming the ACE Revision II estimates.

The 2000 ACE Revision II estimates were the first PES estimates that measured a net overcount in a census of the USA. Figure 1 displays the net coverage estimates based on the PES and Demographic Analysis methodologies for the 1980, 1990, and 2000 censuses conducted in the USA.

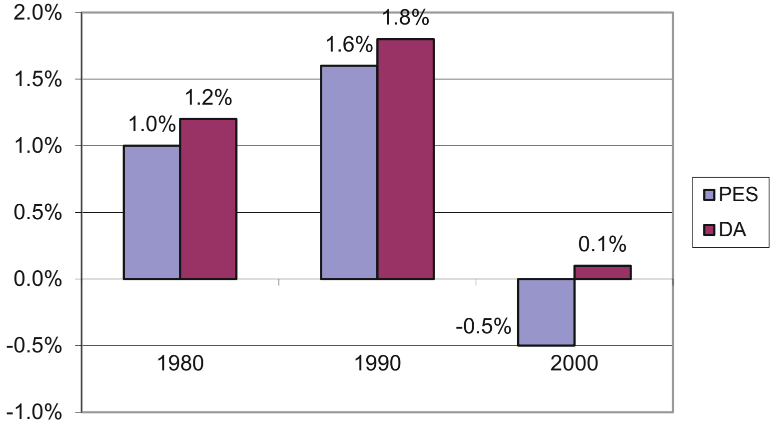

Fig. 1 Percent net undercount estimates from post-enumeration surveys and Demographic Analysis for the 1980–2000 censuses. (Source: Long, Robinson, and Gibson (2003))

6 Innovations in the 2010 PES Methodology

The Census Bureau pursued several innovations in the 2010 implementation of the PES used to evaluate the 2010 Census. The program was called the Census Coverage Measurement (CCM). The enhancements in the CCM involved producing estimates of the components of census coverage error, preparing to include a correction for correlation bias in case one was needed, and using logistic regression instead of post-stratification in the construction of the DSE to produce the estimates of census coverage error. The 2010 CCM focused on measuring the coverage of people in housing units. The CCM also evaluated the coverage of housing units, but those estimates are not discussed in this document.

6.1 Components of Census Coverage Error

One innovation used the PES data to form national-level estimates of the components of census coverage error, namely, the total number of erroneous enumerations and the total number of people missed by the census. Creating these estimates required data processing that differed in some ways from the processing for forming the DSE. For example, a different imputation procedure was needed to compensate for missing data when forming estimates of the erroneous enumerations and the people missed.

The definitions of the four components of census coverage error for persons in housing units are listed below (Mule, 2008). The estimation of the correct, erroneous, and omission components included all the data-defined enumerations in the E sample and did not require that they have a name. This section contains a high-level discussion of the approach to the estimation of the four components of census coverage error. For more details about the estimation method, see Mule (2008). Bell and Cohen (2009) also discuss the 2010 PES. Table 2 contains the estimates for the components of coverage error.

1. *Correct Enumerations.* Estimates of the number of correct enumerations in the final census count were produced at the national level using E sample data, which was a national sample of data-defined enumerations in census housing units. An enumeration was considered correct for component estimation if it was for a person who was counted once and only once in the U.S. housing unit universe. One rule was that if the person was supposed to be enumerated in a housing unit and was included in a housing unit anywhere in the USA, then that person was considered correctly enumerated. If such a person or unit was included multiple times, one of the enumerations was designated correct, and the others were classified as erroneous.

   The estimation approach used a two-stage ratio adjustment to reduce the variability of the estimates and ensure that the sum of estimates for selected subgroups added to the total. The first stage was a ratio adjustment to the E sample weights that was done by identifying cells, which were formed by using characteristics such as race/ethnicity, tenure, age/sex groupings, and then ratio adjusting the sum of the sampling weights in each cell to the total number of data-defined enumerations in the census. The second stage adjustment was applied to each of the first stage cells by a ratio adjustment to the total number of data-defined census enumerations within the cell.
2. *Erroneous Enumerations.* The E sample also was used to produce an estimate of the number of erroneous inclusions in the final census count using the same estimation approach that was used to estimate the number of correct enumerations. Erroneous inclusions consist of duplicate numerations and enumerations of people who should not have been counted in a housing unit. In addition, enumerations for persons born after Census Day and persons who died before Census Day are considered erroneous. The CCM processing identified whether the person should have been counted in the (1) same county but outside of the PES sample block cluster search area, (2) different county in the same state, or (3) different state. The erroneous enumeration estimates used the similar two-stage ratio adjustment.
3. *Whole-Person Census Imputations*. The CCM program tabulated and reported the number of whole-person imputations in housing units directly from the census. The CCM program did not evaluate whether these imputations were correct or erroneous. Whole-person imputations are the result of one or more steps that may include imputing whether a housing unit is occupied, the household size, and the characteristics of the household members.
4. *Omissions*. The CCM program created estimates of the number of omissions of people in housing units from the census. The estimation of the number of

omissions relies on the two following relationships for net error in the census count:

$$Net\ Error = TruePopulation - Census \tag{5}$$

$$Net\ Error = Omissions - Erroneous\ Enumerations \tag{6}$$

Note that Eq. (5) can be rewritten as the following:

$$Omissions = Net\ Error + Erroneous\ Enumerations. \tag{7}$$

Substituting Eq. (5) for *Net Error* in Eq. (7) and some algebra yields the following:

$$Omissions = (TruePopulation - Census) + Erroneous\ Enumerations. \tag{8}$$

Finally, substituting an estimate of the *TruePopulation* size and an estimate of the number of *Erroneous Enumerations* from the PES yields an estimator for *Omissions* as follows:

$$\widehat{Omissions} = True\widehat{Population} - Census + Erroneous\widehat{Enumerations} \tag{9}$$

6.2 Correction for Correlation Bias

Another innovation in the 2010 PES addressed the vulnerability of the DSE to correlation bias, which arises when probabilities of a person or group of people being included in the census and the PES sample are correlated. The remedy was the incorporation of an adjustment for correlation bias. A version of the ratio adjustment for correlation bias first appeared in a revision of 2000 PES estimates. The adjustment was based on the ratio of males to females for Blacks and non-Blacks based on Demographic Analysis estimates of the 2010 population size derived from birth records, death records, and estimates of immigration and emigration (Konicki, 2012; Mulry, 2014).

The correlation bias correction provides a remedy to a violation of the first assumption underlying the DSE (see Sect. 3.2), which requires that inclusion in the census is not correlated with inclusion in the P sample. However, heterogeneity in inclusion probabilities for the census or P sample or both does occur across subgroups. Some people, such as adult males ages 20–35, are hard to count and therefore have lower inclusion rates in both the census and the P sample (Mulry, 2014 p. 50–51).

6.3 Logistic Regression Instead of Post-stratification

Implementations of the PES from 1980 through 2000 used post-stratification in forming the DSEs that were used to evaluate census coverage error. The post-stratified DSE has the disadvantage of requiring an adequate number of observations to produce a reliable estimate of the population defined by a post-stratum. This requirement limits the number of subpopulations for which DSEs can be used to produce census coverage error estimates.

Research during the 1990s demonstrated that PES data collected for forming a post-stratified DSE also could be used in logistic regression models to produce the estimated probabilities needed for constructing a different form of the DSE (Haberman et al., 1998; Alho et al., 1993). This finding enabled creating estimates of population size for subgroups formed using the independent variables in the models and thereby facilitated the construction of estimates of census coverage error for these subgroups. Because the DSEs formed using logistic regression enabled constructing estimates of census coverage error for many more subgroups than were possible when using the post-stratified DSE, the Census Bureau opted to pursue implementing this approach in the 2010 PES.

The form of logistic regression estimator for the DSE uses three separate logistic regression models: one model that predicts the probability of a record being data defined, second that predicts the probability of an E sample record being a correct enumeration, and a third that predicts the probability of a P sample record matching a census record in the search area of its sample block. Then the following formula provides a PES estimate of the population size in poststratum j in area C_j:

$$\widehat{Total}_{Cj} = \sum_{i \in Cj} \left[\pi_{dd,i} {}^{\pi_{CE,i}} \Big/ {\pi_{NM,i}} \right] \tag{10}$$

where

$\pi_{\mathrm{dd,i}}$ = probability of the i-th record being data defined.
$\pi_{\mathrm{CE,i}}$ = probability of the i-th record in the E sample being a correct enumeration.
$\pi_{\mathrm{M,i}}$ = probability of the i-th record in the P sample matching a census enumeration in the search area of its sample block cluster.

Post-stratification requires partitioning the samples into groups that are large enough to form reliable estimates, which possibly suppresses the variability of the estimated probabilities of inclusion in the E and P samples because every observation in a poststratum receives the same estimated inclusion probability. Using the three separate logistic regression models to estimate the probabilities of being data-defined, a correct enumeration, a nonmatch permits more variability and possibly reduces the risk of violating the assumption that the probability of being included in the census is not correlated with being included in the P sample. This is the second on the list of assumptions underlying the DSE given in Sect. 3.2.

6.4 Consultation with Dr. Stokes and Other Experts About 2010 PES Methodology

The Census Bureau sought review and advice about the 2010 PES estimation plans from outside experts on capture-recapture methodology and dual system estimation. They did this by engaging the Committee of Professional Associations on Federal Statistics (COPAFS) to arrange and conduct a meeting of experts, titled the Census Coverage Measurement (CCM) Workshop. At the meeting, a Census Bureau staff presentation of plans preceded a discussion of the topic that included comments on the proposed plans. The papers that Census Bureau staff prepared for the meeting are available at https://www.census.gov/programs-surveys/decennial-census/about/coverage-measurement/pes.html.

Dr. Stokes was invited to the meeting in recognition of her expertise in capture-recapture estimation methodology and for her contributions concerning the design of the quality control operation for the Census Bureau's 1990 PES fieldwork (Stokes & Jones, 1989; Biemer & Stokes, 1989). Her assignment was to review the plans for the imputation for the estimates of two of the 2010 Census components of coverage error, erroneous enumerations, and correct enumerations. The issue was which of the two proposed methods to use for estimating the probability that a census enumeration was correct. The cell method would assign the correct enumeration rate observed for a cell to each enumeration in the cell. The logistic regression method would instead assign each enumeration a probability estimated from the model. Dr. Stokes recommended the logistic regression approach because the method for selecting independent variables for the model was more straightforward. Although the 2010 PES used the cell method, the plans for the 2020 PES imputation include using logistic regression models.

Dr. Stokes leveraged her expertise to provide useful comments on many other aspects of the plans for the 2010 PES. One suggestion grew out of a discussion of the proposed plan to fill in missing characteristics in 2010 Census enumerations by using the characteristics for the person that could be found in the 2000 Census records. Dr. Stokes suggested going a step further and to consider the 2000 Census to be administrative records and use the 2000 Census records to enumerate some households when the household at an address appears to have the same family structure and the people are ten years younger (U.S. Census Bureau, 2009). The Census Bureau adopted a variation of this proposal in the 2020 Census by using administrative records to enumerate 5.6% of the addresses in the USA (Mulry et al., 2021).

6.5 2010 PES Estimates

The Census Bureau incorporated suggestions from the experts into the plans for the 2010 PES. The results of implementing the new methodology in the 2010 PES

Table 1 National estimates of net undercount by census year from PES

		Net undercount		Percent net undercount	
Year	Census count (thousands)	Estimate (thousands)	Standard error (thousands)	Estimate (%)	Standard error (%)
2010	300,703	−36	429	−0.01	0.14
2000	273,587	−1332*	542	−0.49*	0.20
1990	248,710	3994*	488	1.61*	0.20

Source: The 2010 estimates are from Mule (2012) and the 2000 and 1990 estimates are from Kostanich (2003)
The 2000 and 2010 Census counts exclude persons in group quarters and persons in Remote Alaska
A negative net undercount or percent net undercount estimate indicates an overcount
An asterisk (*) denotes a (percent) net undercount that is significantly different from zero
The standard error estimates are model-based and based on the PES

Table 2 Estimates of the components of 2010 Census Coverage

Components of census coverage	Estimate (thousands)	Standard error (thousands)	Percent (%)	Standard error (%)
Census count	300,703	0	100.0	0
Estimates from PES				
Population size	300,667	429	100	0
Correct enumerations	284,668	199	94.7	0.1
Omissions	15,999	440	5.3	0.1
Net undercount = (PES estimate - census count)	−36	429	−0.01	0.14

Source: Mule (2012). A negative net undercount indicates an overcount

appear in Table 1, which includes the estimates of net undercount in the 2010, 2000, and 1990 censuses in the USA. The estimates of the components of census coverage, correct enumerations, erroneous enumerations, and omissions based on the 2010 PES are shown in Table 2.

7 Current Research

Census Bureau staff currently are looking at ways of improving DSE and census coverage error estimates. Data from administrative records appears to be a fertile ground for research in this area. For the 2020 Census, one innovation was the use of data from federal and third-party sources of administrative records (ARs) to create high-quality household rosters for use in enumerating some households in the 2020 Census Nonresponse Followup (NRFU) operation. The main goal of using ARs in this process was to reduce the cost of the NRFU fieldwork while maintaining its high quality. The use of AR information reduces the number of contact attempts by NRFU enumerators at addresses that were in NRFU because a self-response was not received. AR rosters were used to enumerate addresses only if a self-response was not submitted for the address during the self-response period and if one contact attempt by a NRFU enumerator failed to resolve the status of the address. See Mulry et al. (2021).

In recent years, statistical agencies in other counties have examined the potential for improving DSE estimates for subgroups and their entire populations by incorporating "known" totals from administrative record systems (Bryant & Graham, 2015; van der Heijden et al., 2018, 2020). This is feasible in countries where administrative record systems have high coverage of the population. However, some of these countries have minority groups that are poorly covered by their administrative record systems; thus, these countries are looking for ways to improve estimates for their minorities. Because the USA does not have a single source of administrative records that covers the entire population, the Census Bureau's research is focusing on ways of using these approaches where the "known" totals are from Demographic Analysis. Even though Demographic Analysis estimates are available only at the national level, the intent of the research is to gain knowledge about the strengths and weaknesses of the administrative records for future applications.

As part of the 2020 Census Program for Evaluation and Experiments (CPEX), the Census Bureau is conducting the Administrative Record Dual System Estimation Study that is building on the use of administrative records in the 2020 Census. This project seeks to determine whether administrative records and third-party data for the U.S. population can replace the data collected in PES fieldwork in DSEs. In particular, the project is examining whether the use of administrative records as the second system produces census coverage estimates that are close to the survey-based results. Using administrative records could alleviate the need to conduct the field data collection, develop clerical matching software, and pay the clerical matching personnel costs to produce the DSEs for census coverage estimates. This has the potential to reduce the cost substantially.

The Administrative Record Dual System Estimation Study builds on methods used in other countries for deriving estimates of the population size using files created by linking registers. The linking of registers may not produce a file that covers the entire target population. Van der Heijden et al. (2018) discuss an application of the expectation maximization (EM) algorithm of log-linear models that estimates the part of the population missed by the registers. A novel application creates estimates of the causes of accidents where the cause is recorded in both the police and hospital registers, but the police reporting is more accurate. The paper shows how one can use the EM algorithm to produce estimates. Van der Heijden et al. (2020) describe an application of EM in a census context by using multiple registers to estimate the size of the New Zealand Maori population. The Administrative Record Dual System Estimation project is experimenting with the use of this EM methodology to estimate the size of race and Hispanic-origin populations in the USA. Since race and Hispanic origin are available from responses to the 2020 Census and from historical administrative records, estimates of the size of subpopulations can be compared with the estimates produced by the evaluation.

8 Summary

Dr. Lynne Stokes brought her unique skill set to bear in devising methods for estimating bias due to interview fabrication in the dual system estimator used for estimating census net undercount. She gained an in-depth knowledge of capture-recapture estimation when she worked at the Fish and Wildlife Service, as demonstrated in her paper "The Jolly-Seber Method Applied to Age-Stratified Populations" (Stokes, 1984). When she moved to the Census Bureau, she learned about survey research methodology and the challenge of designing quality control samples to detect interview fabrication as demonstrated in her co-authored paper "The Optimal Design of Quality Control Samples to Detect Interviewer Cheating" (Biemer & Stokes, 1989).

Dr. Stokes applied her background to use interviewing quality control data and the evaluations of the 1990 PES to estimate the number of residual fabrications remaining in the data after the quality control operation identified and corrected some fabrications. In addition to estimating the bias at the national level, she also constructed bias estimates for geographic and demographic subpopulations. Her work on the quality of the PES data was critical to deliberations regarding the adjustment of the 1990 Census.

More importantly, the method that Dr. Stokes used in estimating the residual fabrication errors convinced the Census Bureau of the effectiveness of the quality control operation to the point that it became an accepted practice. The Census Bureau did not construct the estimate of the residual fabrication in the interview data for any of the subsequent PESs. The basic approach to the PES interviewing quality control has remained the same even though technological advancements have enabled enhancements in the operation.

Dr. Stokes has demonstrated a flair for adapting methods developed for one application to other uses. For example, she generalized her work on estimating the amount of residual fabrication in a survey data set to the problems of quality acceptance sampling in manufacturing. The paper she co-authored with her colleague Betsy Greenberg at the University of Texas entitled "Estimating Nonconformance Rate after Zero-Defect Sampling with Rectification" (1992) generated substantial interest among engineers from semiconductor manufacturing settings who adapted the method to their projects. Drs. Stokes and Greenberg next expanded their research topic to include the possibility of misclassification error in quality control operations. This type of error may cause a good batch to fail or a bad batch to pass. In their paper entitled "Repetitive Testing in the Presence of Inspection Errors," Drs. Stokes and Greenberg (2012) formulated a rule about how many times to repeatedly test a batch before considering it to fail. The rule for repetitive testing is used in manufacturing and has numerous potential applications in surveys.

The evaluations of U.S. censuses continue to use the PES methodology to evaluate the coverage of the decennial census. These implementations of the PES have built on the quality control methodology that Dr. Stokes developed for the 1990 PES.

References

Alho, J., Mulry, M. H., Wurdeman, K., & Kim, J. (1993). Estimating heterogeneity in the probabilities of enumeration for dual system estimation. *Journal of the American Statistical Association, 88*(423), 1130–1136. https://doi.org/10.1080/01621459.1993.10476386

Belin, T., Diffendal, G. J., Mack, S., Rubin, D. R., Schafer, J. L., & Zaslavsky, A. L. (1993). Hierarchical logistic regression modeling for imputation of unresolved enumeration status in undercount estimation. With discussion and rejoinder. *Journal of the American Statistical Association, 88*(423), 1149–1159. with discussion and rejoinder 1160-1166. https://doi.org/10.1080/01621459.1993.10476388

Bell, R. M., & Cohen, M. L. (2009). *Coverage measurement in the 2010 census*. National Academy of Sciences.

Bell, W. R. (2001). *ESCAP II: Estimation of correlation bias in 2000 A.C.E. using revised demographic analysis results*. Executive Steering Committee for A.C.E. Policy II, Report No. 10. dated October 13, 2001. Washington, DC: U.S. Census Bureau.

Bell, W. R. (1993). Using information from demographic analysis in post-enumeration survey estimation. *Journal of the American Statistical Association, 88*(423), 1106–1118. https://doi.org/10.1080/01621459.1993.10476381

Biemer, P. P., & Stokes, S. L. (1989). The optimal design of quality control samples to detect interviewer cheating. *Journal of Official Statistics, 5*, 23–39. https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/the-optimal-design-of-quality-control-samples-to-detect-interviewer-cheating.pdf

Bryant, J. R., & Graham, P. (2015). A Bayesian approach to population estimation with administrative data. *Journal of Official Statistics, 31*(3), 475–487. https://doi.org/10.1515/JOS-2015-0028

Chandrasekar, C., & Deming, W. E. (1949). On a method for estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association, 44*(245), 101–115. https://doi.org/10.1080/01621459.1949.10483294

Coale, A. J. (1955). The population of the United States in 1950 classified by age, sex, and color-a revision of census figures. *Journal of the American Statistical Association, 50*(1), 16–54. https://doi.org/10.1080/01621459.1955.10501249

Department of Commerce vs United States House. 98–404. Supreme Court of the U.S. 1999. https://www.law.cornell.edu/supremecourt/text/525/326.

Fay, R. E. (1988). *The coverage of the population in the 1980 census*. PHC 80-E4. Evaluation and Research Reports.1980 Census of Population and Housing. Washington, DC: U.S. Census Bureau.

Fellegi, I. P., & Sunter, A. B. (1969). A theory for record linkage. *Journal of the American Statistical Association, 64*, 1183–1210. Alexandria, VA: American Statistical Association. https://www.tandfonline.com/doi/abs/10.1080/01621459.1969.10501049?msclkid=bc6f6f01cfe511ec908375c560ae084b

Greenberg, B., & Stokes, S. L. (1992). Estimating nonconformance rate after zero-defect sampling with rectification. *Technometrics, 34*(2), 203–213. https://www.jstor.org/stable/1269236

Haberman, S., Jiang, W., & Spencer, B. (1998). *Development of methodology for evaluating model-based estimates of the population size for states*. NORC Working Paper Series No. WP-2021.03 (with minor updates in 2021). Chicago, IL: NORC. https://www.norc.org/PDFs/Working%20Paper%20Series/WPS_HABERMAN_2021.03.pdf

Hogan, H. (1993). The 1990 post-enumeration survey: Operations and results. *Journal of the American Statistical Association, 88*(423), 1047–1060. https://doi.org/10.1080/01621459.1993.10476374

Jaro, M. (1989) Advances in Record-Linkage Methodology as Applied to Matching the 1985 Census of Tampa, Florida. *Journal of the American Statistical Association, 84*, 414–420. http://dx.doi.org/10.1080/01621459.1989.10478785

Kelley, R. P. (1986). *Robustness of the Census Bureau's record linkage system*. Proceedings of the American Statistical Association, Section on Survey Research Methods. 620–624. http://www.asasrms.org/Proceedings/papers/1986_116.pdf?msclkid=89715673cfe411ecb1eed28538e496ec.

Kline, D. (2004). *Census 2000 data capture*. Census 2000 Testing, Experimentation, and Evaluation Program. Topic Report No. 3, TR-3. U.S. Census Bureau. Washington, DC. http://www.census.gov/pred/www/rpts/TR3.pdf.

Konicki, S. (2012). *2010 Census coverage measurement estimation report: Adjustment for correlation bias*. DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-11. Washington, DC: U.S. Census Bureau.

Kostanich, D. (2003). *Technical assessment of A.C.E. Revision II*, DSSD A.C.E. Revision II Memorandum Series #PP-61. Washington, DC: U.S. Census Bureau. https://www.nrc.gov/docs/ML1233/ML12335A672.pdf

Long, J. F., Robinson, J., & Gibson, C. (2003). Setting the standard for comparison: Census accuracy from 1940 to 2000. In *2003 proceedings of the American Statistical Association, section on government statistics* (pp. 2515–2524). American Statistical Association.

Marks, E. S., Mauldin, W. P., & Nisselson, H. (1953). The post-enumeration survey of the 1950 census: A case history in survey design. *Journal of the American Statistical Association., 48*(262), 220–243. https://www.jstor.org/stable/2281284

Marks, E. S., & Waksberg, J. (1966). Evaluation of the 1960 census through case-by-case checking. In *1966 proceedings of the American Statistical Association, social statistics section* (pp. 62–70). American Statistical Association.

Mule, T. (2012). *2010 census coverage measurement estimation report: Summary of estimates of coverage for persons in the United States*. DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-01. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/g-series/g01.pdf

Mule, T. (2008). *2010 census coverage measurement estimation methodology.* DSSD 2010 Census Coverage Measurement Memorandum Series #2010-E-18. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/ccm-workshop/2010-e-18.pdf

Mulry, M. H. (2012). Post-enumeration survey. In M. J. Anderson, C. Citro, & J. Salvo (Eds.), *Encyclopedia of the U.S. Census* (2nd ed., pp. 339–343). Sage/CQ Press.

Mulry, M. H. (2014). Measuring undercounts for hard-to-survey groups (Chapter 3). In R. Tourangeau, N. Bates, B. Edwards, T. Johnson, & K. Wolter (Eds.), *Hard-to-survey populations* (pp. 37–57). Cambridge University Press. https://doi.org/10.1017/CBO9781139381635.005

Mulry, M. H., Mule, T., Keller, A. K., & Konicki, S. (2021). *Overview of Administrative Records Modeling in the 2020 census*. 2020 Census Program Memorandum Series: 2021.10. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/administrative-record-modeling-in-the-2020-census.pdf

Mulry, M. H., Bean, S. L., Bauder, D. M., Wagner, D., Mule, T., & Petroni, R. J. (2006). Evaluation of census duplication using administrative records. *Journal of Official Statistics, 22*, 655–679. Statistics Sweden, Stockholm, Sweden. http://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/evaluation-of-estimates-of-census-duplication-using-administrative-records-information.pdf

Prewitt, K. (2012). Decennial censuses: Census 2000. In M. J. Anderson, C. Citro, & J. Salvo (Eds.), *Encyclopedia of the U.S. Census* (2nd ed., pp. 166–169). Sage/CQ Press.

Price, D. (1947). A check on underenumeration in the 1940 census. *American Sociological Review, 12*(1), 44–49.

Spencer, B. D. (1985). Statistical aspects of equitable apportionment. *Journal of the American Statistical Association, 80*, 815–822. https://doi.org/10.1080/01621459.1985.10478188

Statistics Canada. (2021). *Coverage of the 2016 census: Level and trends*. Ottawa, Ontario: Statistics Canada. https://www150.statcan.gc.ca/n1/en/pub/91f0015m/91f0015m2020003-eng.pdf?st=MOftGoal.

Statistics Canada. (2007). *2006 census technical report: Coverage*. Ottawa, Ontario: Statistics Canada. https://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/coverage-couverture/cov-couv_index-eng.cfm.

Stokes, S. L. (1984). The Jolly-Seber methodology applied to age-stratified populations. *Journal of Wildlife Management, 48*(3), 1053. https://doi.org/10.2307/3801468

Stokes, S. L. & Greenberg, B. (2012). Repetitive testing in the presence of inspection errors. *Technometrics, 37*(1), 102–111. https://doi.org/10.1080/00401706.1995.10485893

Stokes, S. L., & Jones, P. M. (1989). Evaluation of quality control procedure for the post enumeration survey. *1999 Proceedings of the Survey Research Section*, Annual Meeting of American Statistical Association. Alexandria, VA: American Statistical Association. 696–698. http://www.asasrms.org/Proceedings/papers/1989_127.pdf

Stokes, S. L., & Mulry, M. H. (1987). On the design of interpenetration experiments for categorical data items. *Journal of Official Statistics, 4*, 389–402. Statistics Sweden, Stockholm, Sweden https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/on-the-design-of-interpenetration-experiments-for-categorical-data-items.pdf

U. S Census Bureau. (1980). Chapter 9. Research, evaluation, and experiments. In *1980 procedural history*. U.S. Census Bureau. https://www2.census.gov/prod2/decennial/documents/1980/proceduralHistory/Chapter_09.pdf

U.S. Census Bureau. (2009). *Transcription of the 2010 census coverage measurement workshop, Jan 12–13*. U.S. Census Bureau.

U.S. Census Bureau. (2021a). Directors 1790 – 1820. In *Census then now*. U.S. Census Bureau. https://www.census.gov/history/www/census_then_now/director_biographies/directors_1790_-_1810.html

U.S. Census Bureau. (2021b). *1980 Overview*. U.S. Census Bureau. https://www.census.gov/history/www/through_the_decades/overview/1980.html

van der Heijden, P. G. M., Cruyff, M. J. L. F., Smith, P., Bycroft, C., Graham, P., & Matheson-Dunning, N. (2020). Multiple system estimation using covariates having missing values and measurement error: estimating the size of the Maori population in New Zealand. *arXiv*. arXiv:2007.00929 [stat.AP].

van der Heijden, P. G. M., Smith, P., Cruyff, M., & Bakker, B. (2018). An overview of population size estimation where linking registers results in incomplete covariates, with an application to mode of transport of serious road casualties. *Journal of Official Statistics, 34*(1), 239–263. https://doi.org/10.1515/JOS-2018-0011

Woodward, J., & Damon, B. (2001). *Housing characteristics: 2000*. Report No. C2KBR/01–13. Washington, DC: U.S. Census Bureau. https://www.census.gov/prod/2001pubs/c2kbr01-13.pdf#:~:text=According%20to%20Census%202000%2C%20there%20were%20115.9%20million,2000%2C%20the%20United%20States%20housing%20inventory%20increased%20by