Results and Products from the Census Coverage Measurement Program for the 2010 U.S. Census

Scott Konicki[1]

U.S. Census Bureau, Washington, D.C. 20233, United States of America

E-mail: Scott.M.Konicki@census.gov

Abstract

To assess the coverage of persons and housing units in the 2010 Census, the United States Census Bureau conducted a post-enumeration survey called the Census Coverage Measurement survey. In this paper, I present an overview of the coverage estimates and products from this evaluation of the 2010 decennial census. As with previous post-enumeration surveys conducted by the Census Bureau, we estimated the net coverage of the decennial census count using dual system estimation. New for the 2010 Census, we provided estimates of the components of census coverage. We described how the census count was composed of correct and erroneous enumerations and whole-person census imputations. We also presented the population estimate as the sum of correct enumerations and omissions. These component estimates allowed us to quantify the offsetting errors that resulted in a negligible net overcount. The Census Coverage Measurement estimates for the United States were released in a series of twelve reports in May 2012, seventeen months after the initial 2010 Census results were announced. In December 2012, we published single-page graphical summaries of the census coverage estimates for states and large places and counties on the Census Bureau website. The 2010 coverage measurement data will serve as an invaluable source of information for various research and planning activities in preparation for future censuses.

Keywords: coverage evaluation, dual system estimation, post-enumeration survey

1. Introduction

In 2010, the United States conducted its 23rd decennial census of population and housing. As with recent decennial censuses, the United States Census Bureau carried out a post-enumeration survey to evaluate the coverage of the 2010 Census. The post-enumeration survey for 2010, called the Census Coverage Measurement (CCM) survey, measured the coverage of the household population and housing units outside of Remote Alaska enumeration areas. Group quarters and persons residing in group quarters were excluded from the coverage analysis. The CCM program provided estimates of net coverage and the components of census coverage as an evaluation and to inform planning for future censuses. Results from the CCM survey were not used to adjust the 2010 Census counts for any purpose.

This paper presents an overview of the results from the 2010 CCM survey and discusses the products released by the program. I also provide a brief summary of the survey operations and the estimation methodology. Because the components of census coverage are a new feature of the 2010 CCM program, I focus on these estimates rather than the net coverage estimates. While the CCM survey was also conducted in Puerto Rico, this paper will only discuss the results for the 50 states and the District of Columbia.

---

[1] Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

2. CCM Operations and Estimation Methodology

This section provides a high-level description of the major survey operations and methodology for the census coverage estimation. More information on these topics can be found in Viehdorfer (2011) and Mule (2008), respectively.

2.1. Sampling, Data Collection, and Matching Operations

The initial CCM sample of block clusters was selected in 2009. A block cluster is one or more contiguous census blocks and averages 30 housing units. From August to December 2009, field workers listed the addresses in the sample areas. Using information from this independent listing, the CCM interview sample was identified. The final sample contained about 170,000 housing units in the United States. Interviewers conducted an independent enumeration of the persons in the sample housing units beginning in August 2010. To maintain operational independence, this enumeration did not begin until after decennial census field activities were completed. Persons and housing units from the CCM sample were matched to the census enumerations to provide the data for coverage estimation. Followup interviews were conducted in early 2011 to help resolve questionable cases.

2.2. Net Coverage Estimation

Like previous post-enumeration surveys, the CCM program used dual system estimation to generate population estimates of persons and housing units. The two samples used were called the Population (P) sample and the Enumeration (E) sample. The P sample consisted of the persons and housing units independently enumerated by the CCM survey. The E sample was a sample of the census enumerations in the same areas as the P sample. After matching and followup, the P sample provided information on persons and housing units missed by the census and the E sample provided information on erroneous census inclusions.

For the 2010 CCM, we used logistic regression modeling to produce synthetic population estimates. In previous post-enumeration surveys, post-stratification, which is a special case of logistic regression modeling, was used to produce the population estimates. The use of more general logistic regression modeling allowed us to include additional independent variables and continuous variables without having to include unnecessary higher-order interactions as when forming post-stratification cells.

Estimates of the net coverage are the difference of the population estimate and the census count for a given estimation domain. A positive estimate denotes a net undercount, and a negative estimate denotes a net overcount. The percent net coverage is the net coverage estimate divided by the population estimate, expressed as a percentage.

2.3. Components of Census Coverage Estimation

For the first time in production, the 2010 CCM provided estimates of the components of census coverage. For persons, the components of the census count were correct enumerations, erroneous enumerations, and whole-person census imputations. For housing units, the components of the census count were correct and erroneous enumerations. We also described the person and housing unit population estimates as the sum of correct enumerations and omissions.

The estimates of correct and erroneous enumerations were design-based estimates using the matching results of the persons and housing units in the E sample. A census record was considered correctly enumerated if the record corresponded to a person or housing

unit that should have been counted in the correct geographic area. The correct geographic area changed based on the level of geography being evaluated. For example, consider a person who was enumerated in California, but should have been counted in New York. This person is a correct enumeration for the national-level estimate because he was enumerated somewhere in the United States in the coverage universe. However, the census record in California is an erroneous enumeration for the state-level estimates. Other types of erroneous enumerations included duplicate enumerations and persons or housing units that should not have been enumerated at all. For duplicate enumerations, information from the CCM interview was used to determine which enumeration was correct. We classified the other enumeration(s) as erroneous due to duplication.

Whole-person census imputations were census records for which all of the characteristics were imputed. The CCM did not assess whether individual whole-person census imputations were correct or erroneous. The whole-person census imputations were tallied, not estimated, and thus the results were not subject to sampling error. Omissions were persons or housing units that should have been enumerated but were not. To estimate omissions, we subtracted the correct enumerations from the population estimate.

3. Estimates of Net Coverage

As with previous census coverage evaluations, the CCM program produced net coverage estimates showing undercounts or overcounts. We provided person net coverage estimates for the nation, major demographic groups, certain census operational areas, states, and large places and counties. Similarly for housing units, we provided net coverage estimates for the nation, characteristics of housing units, certain census operational areas, states, and large places and counties. Table 1 and Table 2 provide examples of some net coverage estimates for persons and housing units, respectively. For persons at the national level, we estimated a net overcount of 0.01% that was not significantly different from zero. We measured differential net coverage by race, among other characteristics. For total housing units, we estimated a 0.60% net undercount.

Table 1. Net Coverage Estimates for Persons (in Thousands)

Race Alone or In Combination[1]	Census Count	Net Undercount	Standard Error	Percent Net Undercount	Standard Error (%)
U.S. Total	300,703	-36	429	-0.01	0.14
White	225,547	-1,206*	319	-0.54*	0.14
Black	40,153	846*	209	2.06*	0.50
Asian	16,969	0	88	0.00	0.52
American Indian and Alaskan Native	5,056	8	36	0.15	0.71
Native Hawaiian or Pacific Islander	1,189	12	25	1.02	2.06
Some Other Race	21,448	355*	69	1.63*	0.31

A positive estimate denotes a net undercount, and a negative estimate denotes a net overcount.
The census count excludes persons in group quarters and persons in Remote Alaska.
[1]A person can be included in more than one classification.
*Estimate is significantly different from zero.

Table 2. Net Coverage Estimates for Housing Units (in Thousands)

Occupancy Status	Census Count	Net Undercount	Standard Error	Percent Net Undercount	Standard Error (%)
U.S. Total	131,676	790*	266	0.60*	0.20
Occupied Units	116,699	36	160	0.03	0.14
Vacant Units	14,977	755*	174	4.80*	1.06

A positive estimate denotes a net undercount, and a negative estimate denotes a net overcount.
The census count excludes housing units in Remote Alaska.
*Estimate is significantly different from zero.

4. Estimates of Components of Census Coverage

We provided component estimates for the same estimation domains as the net coverage estimates. Table 3 provides the national-level component estimates for persons. We estimated that 284.7 million (94.7%) of the census person records were correct enumerations and 10.0 million records (3.3%) were erroneous enumerations. The majority of the erroneous enumerations were due to duplication (8.5 million, not shown in table). The remainder of the census count was the 6.0 million (2.0%) whole-person census imputations. We estimated that 16.0 million people were omitted from the 2010 Census. These omissions represented 5.3% of the CCM population estimate. Many of the omissions may have been accounted for by the imputations. From these component estimates, one can see how the omissions balanced with the erroneous inclusions and whole-person census imputations, resulting in a net overcount that was practically zero.

Table 3. Component Estimates for Persons (in Thousands)

Component of Census Coverage	Estimate	SE(EST)	Percent	SE(%)
Census Count	300,703	0	100.0	
Correct Enumerations	284,668	199	94.7	0.07
Erroneous Enumerations	10,042	199	3.3	0.07
Whole-Person Census Imputations[1]	5,993	0	2.0	0
Estimate of Population from CCM[2]	300,667	429	100.0	
Correct Enumerations	284,668	199	94.7	0.1
Omissions[3]	15,999	440	5.3	0.1

1. These imputations represent people for whom we did not collect sufficient information.
2. This number is the CCM estimate of people who should have been counted in the CCM universe.
3. Omissions are people who *should have been* enumerated in the United States, but were not.

Table 4 provides the national-level component estimates for housing units. Of the 131.7 million census housing units, the CCM estimated that 128.2 million (97.3%) were correct enumerations and 3.5 million (2.7%) were erroneous enumerations. The majority of the housing unit erroneous enumerations were due to reasons other than duplication (2.3 million, not shown in table), such as nonresidential units (e.g., group quarters or commercial) and nonexistent units (e.g., empty lots or demolished). We estimated that 4.3 million housing units were omitted from the 2010 Census.

Table 4. Component Estimates for Housing Units (in Thousands)

Component of Census Coverage	Estimate	SE(EST)	Percent	SE(%)
Census Count	131,676	0	100.0	
Correct Enumerations	128,184	147	97.3	0.1
Erroneous Enumerations	3,492	147	2.7	0.1
Estimate of Housing from CCM[1]	132,467	266	100.0	
Correct Enumerations	128,184	147	96.8	0.2
Omissions[2]	4,283	230	3.2	0.2

1. This number is the CCM estimate of housing units that should have been included in the CCM universe.
2. Omissions are housing units that *should have been* enumerated in the United States, but were not.

4.1. Census Operational Outcomes

In addition to the estimation domains mentioned previously, we also provided component estimates based on results of certain census operations. Because operational outcomes are characteristics of the census records that we could not measure in the P sample, we did not generate dual system estimates for census operational outcomes. Therefore, this section does not discuss estimates of net coverage or omissions. The 2010 CCM program

marked the first time that coverage estimates for census operational outcomes were measured. Here, I present some of these estimates for persons.

The CCM program estimated the components of census coverage for persons by the mail return date of the housing unit in which the person was enumerated. The estimates of correct and erroneous enumerations were similar for persons in households that did return a mail questionnaire, regardless of the date of return. The percent of whole-person census imputations were low for mail return cases and increased slightly as the return date extended further beyond Census Day (April 1, 2010). For persons in the mail return universe who did not have a valid mail return, the rates of erroneous enumerations and whole-person census imputations were higher than those persons with a valid return.

The Nonresponse Followup (NRFU) field operation primarily involved census enumerators interviewing housing units in areas that received a mailback questionnaire but did not respond by mail. We estimated the components of census coverage for the NRFU operation by the month of completion and the type of respondent. The percent of NRFU cases that were whole-person census imputations rose as the months went along, from 2.9% for April to 17.5% for July and August. Compared to household member respondent cases, NRFU cases completed by a proxy respondent had a slightly higher rate of erroneous enumerations (6.7% versus 5.0%) and a much higher rate of whole-person census imputations (23.1% versus 1.6%). Types of proxy respondents included neighbors and persons who moved into the housing unit after Census Day.
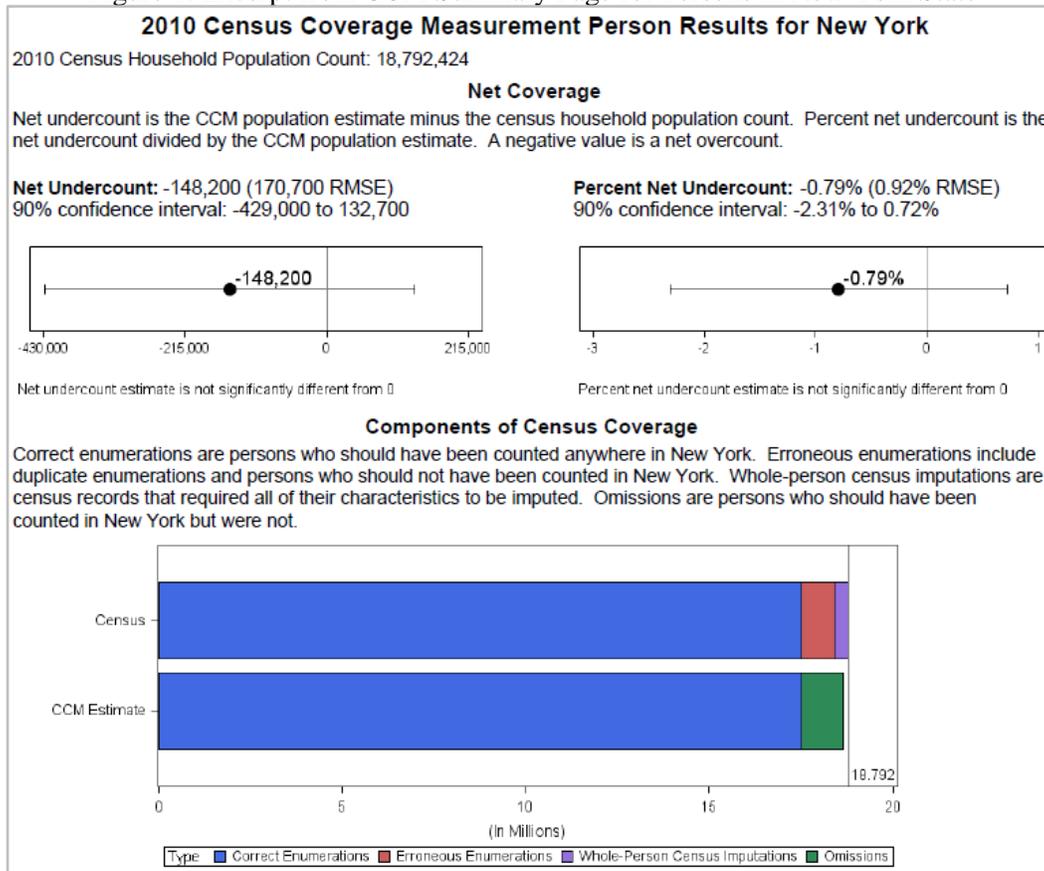
5. Products from the 2010 CCM Program

The 2010 CCM estimates were publically released at a news conference and technical briefing on May 22, 2012, seventeen months after the initial release of the 2010 Census population counts on December 21, 2010. The CCM release consisted of twelve reports providing the coverage estimates for the United States. This report series also included results of estimation procedures such as the handling of missing data and the details of the logistic regression modeling. A single report containing all of the coverage estimates for Puerto Rico was released a month later.

In December 2012, CCM staff provided single-page graphical summaries of the census coverage estimates for states, places, and counties. We used SAS® software's Output Delivery System to create 1,830 Portable Document Format (PDF) files – a person and housing unit coverage summary for each of the 915 governmental entities evaluated by the CCM program. The summaries included 90-percent confidence intervals for the coverage estimates and stacked bar charts showing how the component estimates add to the census count and the CCM population estimate. While these summaries did not include new estimates, the documents provided an easily accessible rendering of the CCM data that was available for governmental entities.

Figure 1 is an excerpt from the summary page for persons in New York State. The net coverage results and the stacked bar chart for the components are shown. The summary pages for states as well as places and counties with a census person count of at least 500,000 also contained confidence interval plots for the individual component estimates. For places and counties with a census person count between 100,000 and 500,000, the summary pages included the net coverage results and a note explaining that component estimates are not available for the given governmental entity. All of the estimation reports and graphical summary pages can be found on the CCM results website at www.census.gov/coverage_measurement/post-enumeration_surveys/2010_results.html

Figure 1. Excerpt from CCM Summary Page for Persons in New York State

2010 Census Coverage Measurement Person Results for New York

2010 Census Household Population Count: 18,792,424

Net Coverage

Net undercount is the CCM population estimate minus the census household population count. Percent net undercount is the net undercount divided by the CCM population estimate. A negative value is a net overcount.

Net Undercount: -148,200 (170,700 RMSE)
90% confidence interval: -429,000 to 132,700

Percent Net Undercount: -0.79% (0.92% RMSE)
90% confidence interval: -2.31% to 0.72%

Net undercount estimate is not significantly different from 0

Percent net undercount estimate is not significantly different from 0

Components of Census Coverage

Correct enumerations are persons who should have been counted anywhere in New York. Erroneous enumerations include duplicate enumerations and persons who should not have been counted in New York. Whole-person census imputations are census records that required all of their characteristics to be imputed. Omissions are persons who should have been counted in New York but were not.

6. Conclusion

Unlike previous post-enumeration surveys, innovations for the 2010 CCM survey allowed us to break down the net coverage of the census into its components. We quantified the offsetting errors of omission and erroneous inclusion that resulted in a negligible net overcount. With the component estimates, we were able to provide evaluations of various census operations and their impact on the final count of persons and housing units. These analyses provided useful information beyond the demographic and geographic estimates of net coverage. Making full use of the Census Bureau website, we disseminated background material in a variety of media formats and published our series of estimation reports and graphical summary pages. Throughout the following decade, data and results from the 2010 CCM program will be used to help plan the 2020 Census and its coverage evaluation survey.

References

Mule, T. (2008), "2010 Census Coverage Measurement Estimation Methodology," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-E-18, U.S. Census Bureau.

Viehdorfer, C. (2011), "The Design of the Coverage Measurement Program for the 2010 Census – Revision #1," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-B-07-R1, U.S. Census Bureau.