

Examining Components of Coverage for the 2010 U.S. Census for Several Census Operations

Patrick J. Cantwell and Vincent T. Mule, Jr.¹
U.S. Census Bureau

Abstract

In recent decades, the U.S. Census Bureau has measured the net coverage of the census of population and housing through a comprehensive post-enumeration survey. Following field work, data collection, matching, and processing, the Census Bureau released estimates of net under- or overcount for the U.S. as a whole, for large geographic areas (e.g., states, large counties and cities), and for major demographic groups, e.g., by race, Hispanic origin, gender, and age group. In the 2010 Census, for the first time on a production basis, the Census Bureau produced estimates of the *components* of census coverage: correct and erroneous enumerations, whole-person census imputations, and omissions. In this paper, we briefly define these components, and present estimates of coverage for the household population of the United States, and for several census operations, including date of mail return, month of nonresponse follow-up, and type of census response (self vs. proxy). Analysis of these results will help plan improvements to procedures for the 2020 Census.

Key Words: components of census coverage, correct enumeration, erroneous enumeration, omission, duplication, mail return, nonresponse follow-up

1. Introduction

Following the 2010 Census, the U.S. Census Bureau conducted a survey as part of the Census Coverage Measurement (CCM) program. The difference between the survey estimate of the U.S. household population and the corresponding census count was less than 0.01%. So, was the 2010 Census perfect? In many ways, no. Although the net under- or overcount in the census was essentially 0, the survey measured errors on both sides. An estimated 10 million census people were enumerated erroneously; about 85% of them were duplicates of people already counted. Another 6 million people were counted in the census, but no name was recorded and all their characteristics were statistically imputed. The records of these people are called whole-person census imputations. On the other side, the census missed about 16 million people. However, as we'll suggest later, close to 6 million of them could probably be attributed to the imputations.

One can see the advantage of studying the "components of census coverage," rather than merely the net coverage. In this paper, we'll cover three topics:

- the meaning of coverage in the census, and the difference between net coverage and components of census coverage,
- definitions for each of the components of census coverage, and estimates of these components for the U.S. as a whole, and
- estimates of components of the census count for several important census operations.

¹ The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau. Main author contact: patrick.j.cantwell@census.gov.

This paper provides results, but does not describe or explore the statistical methodology behind the procedures. For information on the methods, see Olson and Viehdorfer (2013) and Keller et al. (2013).

2. Net Error in the Census

We define the net coverage error of the census as the difference between an estimate of the population total and the census count. The estimate can be derived in various ways. For this paper, we estimate the population total based on the results of the Census Coverage Measurement (CCM) survey, a post-enumeration survey, one taken following the 2010 Census. (An alternative method of estimating the size of the population in 2010, demographic analysis, is described in Devine et al. (2012).)

Before looking at any numbers, we note that the CCM program measured coverage of the U.S. population in housing units. It made no attempt to measure the coverage of people living in group quarters, such as college dorms, jails, prisons, nursing homes, etc. The CCM also excluded a small part of Alaska we refer to as "remote Alaska."

Let's look at an estimate of the net error in the 2010 Census for the U.S. household population. The census count was 300.703 million. (If we had included the population living in group quarters and remote Alaska, this number would be 308,745,538.) The estimate from the CCM was 300.667 million, with a standard error of 0.429 million. Therefore, the census produced an estimated overcount of 36,000 people, or 0.01% of the population. This is not statistically significant. Essentially, we measured no undercount or overcount in the census. Does this mean that the 2010 Census was almost perfect? It would be difficult to make that conclusion. For one, the CCM program estimated a statistically significant undercount for renters, an overcount for owners, and other net errors for some racial or ethnic groups, such as Blacks and Hispanics.

But looking only at results for the U.S. as a whole, there were other errors beyond net error. We can gain insight into the success and problems encountered in the census by breaking the census count and the survey estimate of the population into their components, and analyzing these numbers. We will define the several components and provide results for the total U.S. household population. In subsequent sections, we'll look at several important census operations, and study their results by components.

3. Components of the Census Count

In everything that follows, it should be noted that the terms "census count" and "estimate of the U.S. population" refer to our restricted universe, that is, the U.S. household population, not including remote Alaska. Population in groups quarters is excluded. Therefore, all numbers reflect this restricted universe.

One recalls that the (rounded) household population in the 2010 Census was 300.7 million. In Table 1, we break this number into three main components. The first is called correct enumerations, representing people who should have been counted in the census. A person cannot be associated with more than one correct enumeration; if two or more census records derive from the same person, no more than one can be considered correct. Through the CCM, we estimated about 284.7 million correct enumerations in the census, or 94.7% of all census records. (Standard errors of the estimates can be found in Table 1.) This number can be further broken into two parts. The first encompasses people who

were counted in the correct state, about 283.7 million. Another 0.9 million were counted in an incorrect state.

Table 1. Components of Census Coverage for the U.S. Household Population, 2010 (in Thousands)

Component of Census Coverage	Estimate (thou.)	Standard Error (thou.)	Percent	Standard Error
Census Count	300,703	0	100.0	0
Correct enumerations	284,668	199	94.7	0.07
Enumerated in the correct state	283,720	206	94.4	0.07
Enumerated in a different state	948	31	0.3	0.01
Erroneous enumerations	10,042	199	3.3	0.07
Due to duplication	8,521	194	2.8	0.06
For other reasons	1,520	45	0.5	0.01
Whole-person census imputations	5,993	0	2.0	0
Dual-System Estimate from Survey	300,667	429	100.0	0
Correct enumerations	284,668	199	94.7	0.07
Omissions	15,999	440	5.3	0.07
Net Overcount	36	429	0.01	0.14

Source: U.S. Census Coverage Measurement survey of 2010 (Mule 2012)

Note: Does not include people living in dormitories, prisons, military barracks, etc., or Remote Alaska

The second main component of the census count is called erroneous enumerations: people who should not have been counted in the census for any of several reasons. We estimate that about 10.0 million census records, or 3.3% of the total, were included in the census in error.

The main cause of erroneous enumerations was duplication. Duplicate census records occur for various reasons. As an example, someone might rent an apartment in New York City and live there ten months each year. She might also own a house in Florida and spend January and February there. A census form was likely mailed to both addresses. If she returned a form from both places and was counted in the census at each, she would have been enumerated correctly in New York City, but erroneously in Florida. Her record at the Florida address is called a duplicate. There are other circumstances. College students might have been counted in the dormitory or somewhere near the college campus, and also on their parents' form. In cases of child custody, both parents might have listed the children on their separate forms.

Duplicates accounted for an estimated 8.5 million records, or 2.8% of all census records. The other 1.5 million erroneous enumerations were due to a variety of circumstances. Some people died before Census Day (April 1, 2010) or were born after Census Day. A person who was visiting from another country but not actually living in the U.S. should not have been counted.

The third main component of the census count is called "whole-person census imputations," and accounts for 6.0 million records, or about 2.0%. For most (but not all) of these records, we determined how many people lived in the housing unit. However, because we collected so little of the characteristic information on some or all of the people in the unit, we imputed all their characteristics and no name was recorded on the record. We believe that most of these people would have been confirmed as correct enumerations in the post-enumeration survey, if we had had sufficient information to match them to the people enumerated in the CCM operations. But without a name and the characteristic information, it would have been very difficult to match these census records to the CCM list.

It might be noted that the number of whole-person census imputations is a tally of all census records, and thus has no associated sampling error or standard error. On the other hand, the number of correct or erroneous enumerations is an estimate based on operations conducted only in the sample areas included in the CCM. Their standard errors, as displayed in the table, were computed via a delete-a-group jackknife replication using 100 groups (Imel et al. 2013).

4. Components of the CCM Survey Estimate

Earlier it was stated that the CCM estimate of the U.S. household population was essentially the same as the census count, 300.7 million people. We break the CCM estimate into two components. The first, correct enumerations, is defined the same as for the census count: someone who should have been counted in the census (only once). For this category, the estimate is the same as was seen for the correct enumeration component of the census count, that is, 284.7 million for the U.S., or 94.7%. (Because the census count and the CCM estimate are so close, correct enumerations, as a percent of the estimate of total, make up essentially the same as its percent of the census count.)

The second component of the CCM estimate is labeled census omissions. Omissions represent two types of people: (1) those who we believe really were missed in the census; and (2) those who we believe were enumerated in the census, but for whom we don't have sufficient information to verify it.

For (1), there are people who were missed because their housing unit was missed in the census. As an example, perhaps the owners of a housing unit rented their basement, but neither the basement nor the renters were captured in the census. On the other hand, a housing unit may have been captured in the census, but some or all individuals were missed. In the example above involving child custody, if neither parent included their children on the census form, the children would have been census omissions.

Category (2) above includes people for whom we collected little or no information on the census form. For each of the whole-person census imputations defined in the last section, the record has no name, and all characteristics were imputed. Thus, if that person was enumerated as part of the CCM sample, the CCM record would likely not match to a census record, and the person would appear to have been missed in the census. We believe that most of these people were correctly enumerated in the census, but without sufficient information collected.

The CCM estimate of census omissions was 16.0 million, or 5.3%. However, up to 6.0 million of them may have been accounted for by the 6.0 million whole-person census imputations.

Looking at Table 1, one can set aside the correct enumerations contained in the census count and the CCM estimate. Based on what remains, one sees that the 16.0 million census omissions were offset by 6.0 million whole-person census imputations and 10.0 million erroneous enumerations. As the errors from the two "sides" balance so closely, no statistically significant undercount or overcount remains. This was not the case for all subgroups of the U.S. total. Among people who rent, the estimate of omissions was greater than that for erroneous enumerations and imputations combined, producing an estimated undercount of renters in the census. For owners, the numbers aligned the opposite way, producing an overcount of owners.

5. Overview of Several Census Operations

Before looking at the estimates of components of census coverage for several operations conducted as part of the census, it is helpful to briefly describe the operations. For most addresses on the census list, we mailed or left a census questionnaire, and asked that a household member complete the form and mail it back. Some households were given a form as early as late February. But most received a form in the mail around the middle of March. Responses were returned from about two-thirds of the addresses in the U.S.

For those addresses from which we received no response, other procedures were conducted, beginning in April or May. Most of them fell into one of the nonresponse follow-up (NRFU) operations. In general, a census enumerator went to their address to conduct an interview. If, after several attempts, a household member could not be contacted, a proxy respondent was allowed. The proxy was someone, such as a next-door neighbor or a building manager, who could supply information about the household.

In the following sections, we analyze the components of the census count for aspects of two specific census operations to determine if there were differences in the census data collected at different periods of time or from different respondents.

In the tables and charts below, percentages of "cases" always refer to the number of *people* involved in the situation or the category described. Analogous percentages for housing units would differ, sometimes markedly. In addition, at times, the percentages don't add to 100% due to rounding.

6. Date of Mail Return

As mentioned above, mail responses were obtained from about two-thirds of the addresses in the U.S. The frequency of returns is shown in Table 2 for the relevant time periods. As the table demonstrates, the peak period of return was the second half of March, 2010.

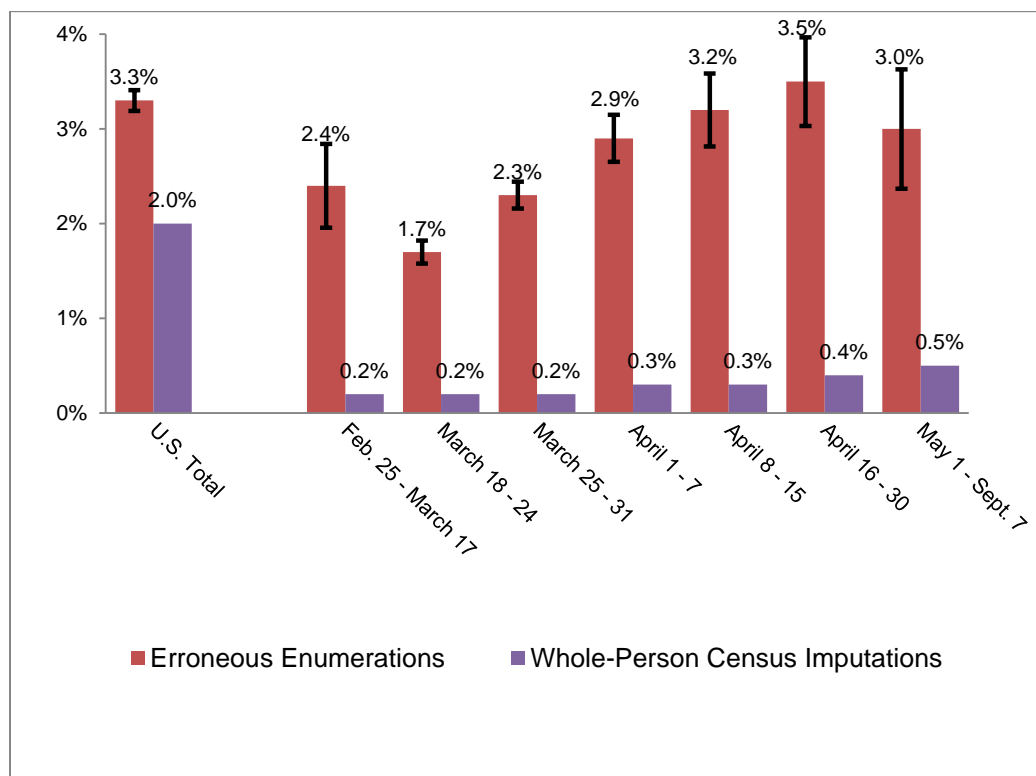
Rates of erroneous enumeration and whole-person census imputation for the time periods are provided in Figure 1. For erroneous enumerations, a 90% confidence interval is inserted at the top of the red (left) bar to indicate the sampling variability in the estimate of the rate. For imputations (purple bar to the right of each red bar), no confidence interval is shown because the percent is based on an exact tally rather than a sample

estimate. For ease of comparison, results for the entire U.S.--mail return and otherwise--is included at the left side of the chart.

Table 2. Percentage of Mail Returns by Date

Mail Return Date	Percent of All Mail Returns
February 25 - March 17	3.7%
March 18 - 24	37.9%
March 25 - 31	29.8%
April 1 - 7	14.1%
April 8 - 14	6.8%
April 15 - 30	6.0%
May 1 - September 7	1.9%

Figure 1. Components by Date of Mail Return



From the chart, one can see that rates of erroneous enumeration for most periods of mail return are smaller than that for the entire U.S. Further, looking only at the mail returns, one mostly sees a steady increase in erroneous enumerations as time progresses. Because the sample sizes are smaller for the earliest and latest periods (fewer returns in those periods), any difference there is not statistically significant. But for the majority of the returns--those arriving in March--these conclusions are statistically significant.

For whole-person census imputations, one clearly sees a much lower rate across all periods of mail return (0.2% to 0.5%) relative to the U.S. Total (2.0%). In addition, the rate of imputation increases slowly across the time periods.

7. Completion Month in the Nonresponse Follow-Up Operation

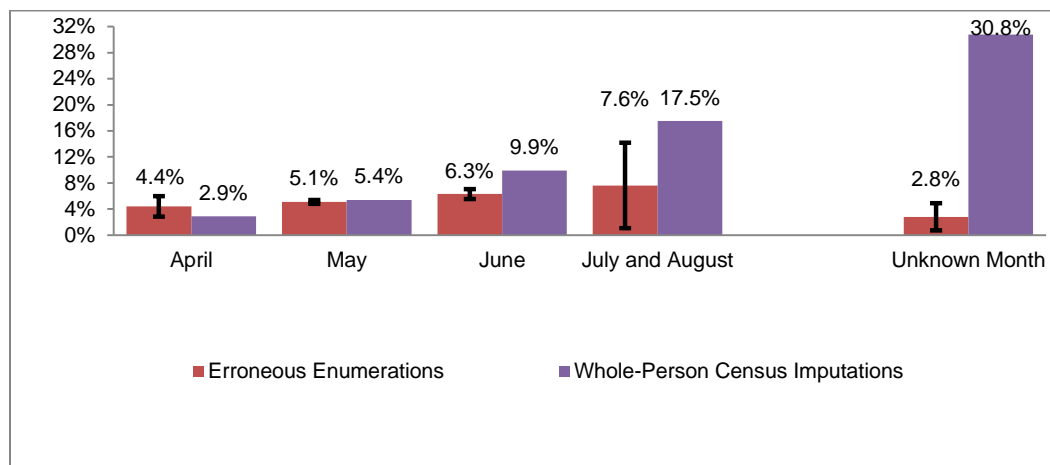
For those addresses from which no response was received, a census enumerator generally tried to procure an interview in person. Although this operation started in April and continued into August, most interviews were conducted in May (77.6%) or June (19.6%). See Table 3 for more details.

Table 3. Percentage of the Nonresponse Follow-Up Operation by Month of Completion

Month of Completion	Percent of NRFU Operation
April	2.2%
May	77.6%
June	19.6%
July and August	0.3%
Unknown Month	0.2%

Figure 2 depicts the rates of erroneous enumeration and imputation across the months. (For 0.2% of the cases, the month of interview was not captured. These results are included only for completeness.)

Figure 2. Components by Completion Month in the Nonresponse Follow-Up Operation



One sees that the rate of erroneous enumeration increases gradually as the time from Census Day increases. This increase is statistically significant from May to June; the sample sizes for the other months are too small to make valid conclusions. However, the rate of imputation does indeed steadily increase over the months.

8. Respondent Type in the Nonresponse Follow-Up Operation

When a census enumerator visited a nonresponding household, he or she tried to speak to someone living in the housing unit with knowledge about the residents. If no one was

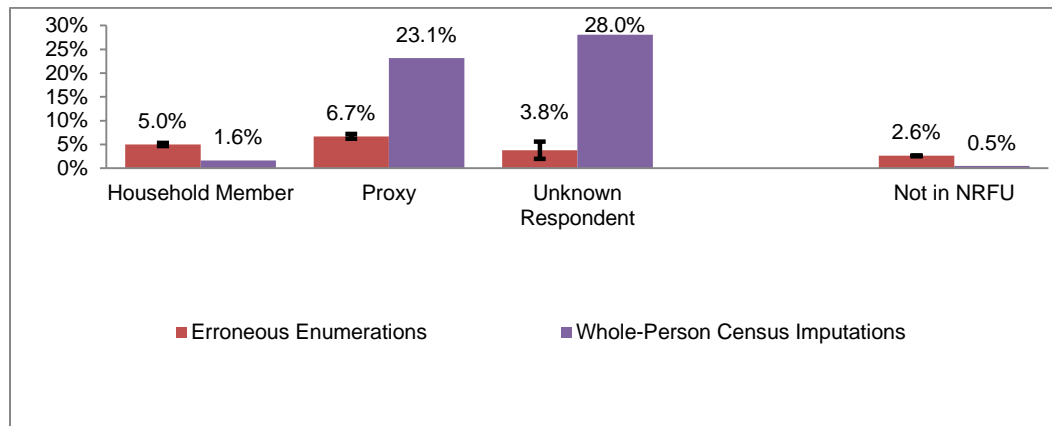
available, a response from a proxy was allowed. As mentioned above, the proxy might be a neighbor living nearby. If the unit was part of complex of apartments or condominiums, the proxy might be a building manager. In most cases, 78.4%, a household member was reached; a proxy was required in 20.8% of the cases. (The situation could not be determined in 0.8% of the cases. Once again, their results are shown for completeness.) See Table 4.

Table 4. Percentage of the Nonresponse Follow-Up Operation by Respondent Type

Respondent Type	Percent of NRFU Operation
Household Member	78.4%
Proxy	20.8%
Unknown Respondent	0.8%

Figure 3 provides the results for components by respondent type. The proportion of erroneous enumerations for proxies, 6.7%, is slightly higher than that for household members, 5.0%. What is more important here is that, when a household member is not available and a proxy is required, little or no usable information was captured in many cases. This is reflected in the rate of whole-person census imputation. The rate of imputation for proxies is 23.1%, compared to 1.6% for household members.

Figure 3. Components by Respondent Type in the Nonresponse Follow-Up Operation



It might also be noted that these error rates for household members and proxies are much higher than for cases that did not go to the Nonresponse Follow-Up Operation. The latter are primarily mail-return cases. Their error rates--2.6% for erroneous enumerations, and 0.5% for imputations--are provided in Figure 3 for comparison.

9. Summary

Rather than review the results for specific census operations discussed above, we point to the general value of analyzing the components of the census count and of the population estimate from the CCM survey. One saw that, at the U.S. level, the errors in the census count and the population estimate offset, yielding essentially no net overcount or undercount. Yet the errors in each were extensive and bear investigation.

Further, we looked below the level of the U.S. For several census operations, we divided the U.S. population by stated categories and compared the rates of erroneous enumeration and whole-person census imputation. In a similar manner, one can examine the components of the census count and the survey estimate (1) for demographic subgroups, such as by race and Hispanic origin, and (2) for geographic subdomains, such as states and large cities and counties. Results for these subdivisions of the U.S. are available in a series of memoranda found on the Census Bureau's website at http://www.census.gov/coverage_measurement/post-enumeration_surveys/2010_results.html.

10. Acknowledgments

We thank various staff at the U.S. Census Bureau for programming and analyzing the results of the CCM survey. Many are authors of the memorandum series mentioned in the previous paragraph. We also thank Debbie Fenstermaker for reviewing this paper.

11. References

Devine, J., Bhaskar, R., DeSalvo, B., Robinson, J.G., Scopilliti, M., and West, K.K. (2012). "The Development and sensitivity analysis of the 2010 demographic analysis estimates." U.S. Census Bureau, Washington, D.C. <http://www.census.gov/people/files/popworkingpapers/Development%20and%20Sensitivity%20Analysis%202010%20DA.pdf>

Imel, L., Mule, T., Seiss, M., Mulligan, J. (2013). "2010 Census Coverage Measurement Estimation Methods: Measures of Variation." DSSD 2010 Census Coverage Measurement Memorandum Series, #2010-J-06. U.S. Census Bureau, Washington, D.C.

Keller, A., Fox, T., and Davis, P.P. (2013). "2010 Census Coverage Measurement Estimation Methods: Component Estimation Methodology." DSSD 2010 Census Coverage Measurement Memorandum Series, #2010-J-05. U.S. Census Bureau, Washington, D.C.

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. U.S. Census Bureau, Washington, D.C. http://www.census.gov/coverage_measurement/pdfs/g01.pdf

Olson, D. and Viehdorfer, C. (2013). "2010 Census Coverage Measurement Estimation Methods: Net Coverage Estimation." DSSD 2010 Census Coverage Measurement Memorandum Series, #2010-J-04. U.S. Census Bureau, Washington, D.C.