

Household Contact and Probing to Collect a Complete Address List and Household Roster in the Census Coverage Measurement Program¹

Graciela Contreras and Lora Rosenberger
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract

The goal of the Census Coverage Measurement (CCM) Program for the 2010 Census was to provide estimates of net coverage error and components of census coverage, for both housing units and persons living in housing units. To measure the coverage of housing units, the Independent Listing operation created a list of housing units or potential housing units that could exist as of Census Day or by the Person Interview timeframe in selected block clusters, completely independent from any census operations. Each interviewer made up to three attempts to contact a household member to ask if there were additional units at the sample address, such as garage or basement apartments, that may have been missed without talking to a household member. To measure coverage for persons, the Person Interview, which listed all household members living at the sample address any time between Census Day and the day of the Person Interview, included six questions probing for any people that may be tenuously attached to the housing unit. This paper focuses on the final disposition of the housing units and persons collected by these additional questions and probes in the CCM Program.

Key Words: coverage measurement, independent listing, person interview

1. Background: Relevant Census Coverage Measurement Operations

The purpose of the 2010 CCM program was to evaluate coverage error in the 2010 Census to determine what needs improvements in future censuses, meaning 2020 and beyond. The CCM was designed to measure the coverage of housing units and persons in the United States (U.S.) and Puerto Rico, excluding Remote Alaska; and excluding group quarters and persons residing in group quarters. The CCM provided estimates of the net coverage error and the components of census coverage, including omissions and erroneous enumerations. Since the CCM was an evaluation, it did not affect the results of the 2010 Census.

The CCM was conducted independent of the 2010 Census. Seven separate sets of operations cover the entire CCM process:

- Sample Design
- Independent Listing
- Initial Housing Unit Matching and Followup
- Person Interview
- Person Matching and Followup
- Final Housing Unit Matching and Followup, and
- Estimation

¹ Disclaimer: This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.

This paper focuses on the Independent Listing Operation, Person Interview (PI) Operation, and both the initial and final Housing Unit Matching and Followup Operations.

1.1 Independent Listing Sample and Housing Unit Coverage Question

The first CCM field operation, Independent Listing, was conducted from August 28, 2009 to December 12, 2009, in the United States and Puerto Rico. The primary sampling unit in this field operation was a block cluster, which consists of one or more geographically contiguous census blocks. Block clusters were formed to balance statistical and operational efficiencies for the CCM Program. A stratified sample of block clusters was selected for each state or state equivalent (e.g., Washington D.C.).

Listers canvassed each block cluster assigned, and independent from any similar 2010 Census Operation, listed in paper registers called Independent Listing Books (ILBs) all housing units and units that could become housing units by the time of the CCM PI, conducted from August 14, 2010 to October 16, 2010. Listers also map spotted each unit they listed and updated paper maps by adding and deleting streets, as needed.

The 2010 Census Address Canvassing operation, which updated the Census Bureau's housing units list in all census block clusters, only made one attempt to verify address information with a household member, and then listed by observation. In contrast, CCM Independent Listers were required to go to each household up to three times to try to interview a respondent before they could conduct the listing by observation. Figure 1 shows the coverage question listers were instructed to ask at each household. This question was asked during Independent Listing to get better coverage of hard to see or locate housing units, such as basement apartments, converted garage apartments, or other places people might live that could be missed by observation alone. We analyze the data collected from the coverage question in this paper to aid in planning future CCMs.

Figure 1: Additional Housing Unit Coverage Question

(13) At (address), are there any basement or garage apartments, trailers, or other residences, even if no one is living there now?

1 Yes – **How many?** _____

Is/Are the other residence(s) attached to the main residence (e.g., basement apartment) and/or detached (e.g., mobile home, carriage house)?

a Attached – *Change item 12a to multiunit, then go to 14a.*

b Detached – *SKIP to 16 then go to the next line number and record other residence.*

2 No – *SKIP to 16*

In September 2009, an initiative to reduce nonsampling error in the CCM program was implemented. The sample size for operations after Independent Listing was decreased and resulting surplus funds from the reduced workload were put towards approaches to reduce the nonsampling error. As a result of the sample reduction, in addition to the pre-specified subsampling of small block clusters, which occurred before Initial Housing Unit Computer Matching, the sample block clusters were reduced by dropping whole block clusters from the initial sample.

1.2 Initial Housing Unit Matching and Followup Operations

After the new housing unit sample was identified, Initial Housing Unit Computer Matching occurred. The file of housing unit addresses listed by CCM was computer matched against the list of housing units addresses updated during the Census Address Canvassing and Group Quarters validation operation within each sample block cluster and one ring of surrounding blocks. The results of computer matching were preprocessed with clerical matching software to assign initial match codes of match, possible match, nonmatch, or duplicate to the CCM and census addresses. During Before Followup Clerical Matching, clerical matchers looked for matches to the computer nonmatches and reviewed the matches, possible matches, and duplicates from computer matching to determine if any coding corrections were needed. Any addresses that could not be directly matched between the two address files went to Initial Housing Unit Followup (IHUFU).

During IHUFU, interviewers collected additional information for these unresolved addresses to help in a resolution of match codes during After Followup Clerical Matching. During After Followup clerical matching, clerical matching staff used the results of the IHUFU operation to attempt to assign a match code to addresses that went to IHUFU. Once the IHUFU operation wrapped up, the PI field operation occurred.

1.3 Person Interview Operation

The CCM PI was conducted by personal visit using a Computer-Assisted Personal Interview (CAPI) instrument on a laptop computer. The purpose of the PI was to obtain information about the residents of the sample housing unit at the time of the interview. This included nonmovers and people who may have moved into the selected housing unit since Census Day (April 1, 2010), known as in-movers. In addition, PI collected information about people who were living at the sample unit around Census Day to identify persons who moved out of the sample housing unit between Census Day and the time of the CCM interview (out-movers). The PI also collected information to determine if there were any other alternate addresses, where any of the people listed may have been counted in the Census and information necessary to geocode the alternate addresses for future operations. It also collected information to determine where each current resident was living on Census Day and the new address where each out-mover currently lives.

Similar to the housing unit data the person data collected during the PI was computer matched and persons were identified as matched, possibly matched, or not matched. Then in the clerical matching operation, final residence status codes were determined identifying where the people should be counted on Census Day. Cases that remained unresolved were sent to the CCM Person Followup operation, in which interviewers ask for additional details for these unresolved persons. The results gathered from the followup operation were reviewed by the matching staff and final residence codes were

assigned to all persons collected in the PI operation. The final match code results assigned to the tenuously attached people collected from the additional probe questions during the PI operation are presented in this paper.

1.4 Final Housing Unit Sample and Operations

The CCM Final Housing Matching and Followup operations consist of computer processing, clerical matching, and Final Housing Unit Followup. In Final Housing Unit Clerical Processing any changes to units resulting from census operations since the Initial Housing Unit operations were flagged for a clerical review. Census addresses that were added to a block cluster or its surrounding blocks were coded as new nonmatches to be reviewed during clerical matching.

During Before Final Housing Unit Clerical Matching, staff attempted to match addresses, search for duplicate addresses, and determine the Census Day housing unit status and enumeration status for the CCM and census addresses that had undetermined statuses after final computer processing. Cases that remained unresolved following this operation were eligible for Final Housing Unit Followup (FHUFU). During FHUFU, interviewers collected additional information for these unresolved addresses so that it might allow a resolution of match codes to be made. Clerical matching staff used the results of the FHUFU operation to attempt to assign a match code to these addresses. The result of this operation was a set of files containing Final Housing Unit match codes for the CCM addresses and the census addresses in the sample block clusters and their surrounding blocks. The final match code results assigned to the potential additional housing units collected from the coverage question during Independent Listing are presented in this paper.

2. Notes on Data Used

The data used for this research paper was collected from the paper Independent Listing Books and the results discussed in this research paper are taken directly from what was data captured off the questionnaire. Staff at the National Processing Center clerically identified addresses identified by lister's responses to the coverage question as potential additional added housing units. The Independent Listing Books were paper questionnaires and minimal edits were performed on the coverage question, so inconsistencies in the data sometimes existed. The results in this research paper exclude Puerto Rico.

3. Methods and Results

3.1 Number of Potential Additional Housing Units in Independent Listing

During Independent Listing, 834,223 U.S. housing units were listed. The housing units were classified as single family homes, multiunits (apartments), mobile homes not in a park, mobile homes in a park, or as other (for example, occupied camper, tent, van, boat, etc.).

There were 4,219 housing units identified as potential additional housing units added to the CCM Independent Listing due to the responses to the coverage question. These units are thought to be potential "hidden" or hard to locate housing units. Out of these housing units, 2,554 potential additional housing units remained in the CCM sample after

sampling operations following Independent Listing. Table 1 shows the percentages of additional housing units remaining in sample or sampled out after Independent Listing by housing unit type based on the total number of housing units listed.

Table 1: Distribution of Potential Additional Housing Units Resulting from the Coverage Question Remaining in Sample or Sampled Out after Independent Listing

	Type of Housing Units (Percent)				
	Single Family Home	Multiunits	Mobile Home not in a Park	Mobile Home in a Park	Other
Remaining in Sample after Independent Listing	0.57	0.16	4.21	0.41	17.81
Sampled Out after Independent Listing	0.55	0.16	4.47	0.00	8.61
Total	0.57	0.16	4.31	0.24	13.81

Source: Independent Listing Verified File

Note: The percentages in this table were calculated over the total number of each type of housing unit listed per sample indicator. These will not add up to 100 percent.

As shown in Table 1, less than one percent of the potential additional housing units that remained in the sample after Independent Listing were listed as single family homes, multiunits, or mobile homes in a park. Over 4 percent of the potential additional housing units that remained in sample were listed as mobile homes not in a park. These units are mobile homes or trailers that are not managed by an organization or company. Over 17 percent of the potential additional housing units remaining in sample were listed as “other.” These “other” type of housing units can vary from tents, vans, boats, campers as long as they were occupied at the time of Independent Listing. These “other” types of units and trailers not in a park have been identified as hard to locate without the aid of a respondent.

The units remaining in sample after Independent Listing underwent the Initial Housing Unit matching and followup operations. The housing units that remained unresolved or had any changes since the Initial Housing Unit operations or census housing units that were added as new nonmatches then underwent the Final Housing Unit matching and followup operations. Each of the potential additional housing units that remained in the sample received a final housing unit match code.

We reviewed the final housing unit match code status for the potential additional housing units resulting from the coverage question, to help aid future listing operations, identify how well the coverage question did at identifying housing units the Census may have missed. We classified the final housing unit match codes into the following categories: matches, nonmatches, or not housing units. Matches consist of addresses identified as the same between the CCM and Census address lists. Nonmatches consist of CCM addresses not matching the census addresses, addresses that were incorrectly listed in the sample block cluster and those that cannot be assigned that match code with confidence. A unit

was coded as not a housing unit if it did not exist as a housing unit on Census day or if it was a geocoding error², in which case it did not exist within the sample block cluster.

Table 2 shows the distribution of the final housing match codes by type of housing unit.

Table 2: Final Housing Unit Match Status Code Distribution

Final Housing Unit Match Code Status	Type of Housing Units (Percent)					
	Single	Multiunits	Mobile Home not in a Park	Mobile Home in a Park	Other	Total
Nonmatch	16.39	2.46	4.61	0.00	0.61	24.08
Match	38.32	6.05	16.91	0.92	0.92	63.11
Not a Housing Unit	7.17	0.51	4.41	0.00	0.72	12.81
Total	61.89	9.02	25.92	0.92	2.25	100.00

Source: Final Housing Unit Matching, Review, and Coding System Output File

Over 24 percent of the remaining potential additional housing units were nonmatches, where single-family homes make up the majority, at over 16 percent. This nonmatch percentage is of interest because this indicates the additional housing units are those that were listed in the CCM program that were not listed during the census operation. This may be due to the question that the listers had to ask of each respondent.

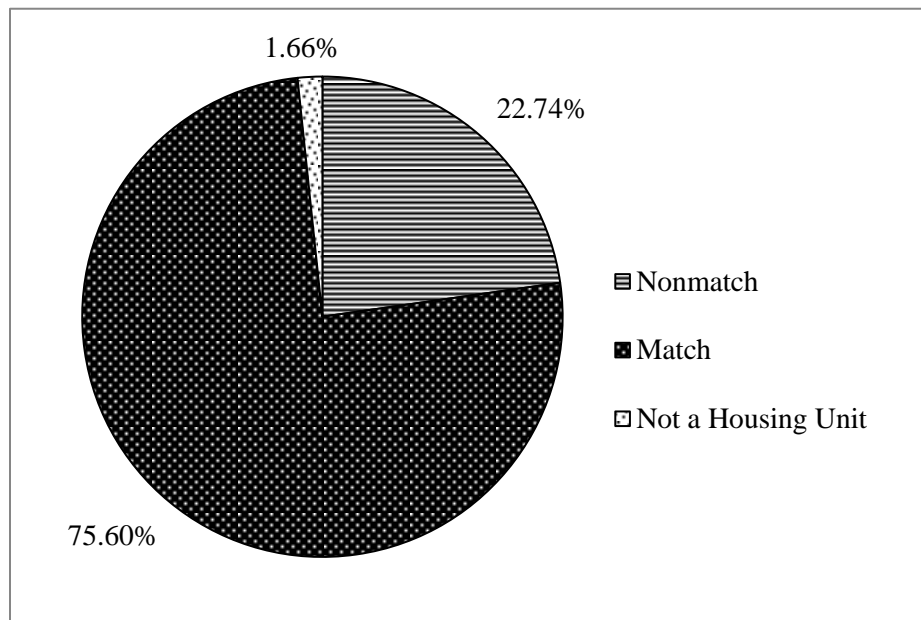
Over 63 percent of the additional housing units were matches, and 12.81 percent were not housing units. More than one-third of the additional housing units were single family and matches. The housing units that matched were also listed by census. When matches occur, it is thought these additional housing units could have been listed without the aid of the coverage question as they could be seen or listed by observation during the census operation. Alternatively, these housing units were not listed initially but added at a later census operation. More research is needed to identify the effectiveness of asking the coverage question during Independent Listing and if additional housing units may be identified in other CCM operations.

Some of these potential additional housing units were sent out in the PI operation (conducted from August to October 2010) and may have had residents occupying the household during Census Day. The final residence codes of the people in these housing units are of interest. A total of 1,328 people were collected in these additional housing units. Figure 2 shows the breakdown of people collected in these additional housing units by final housing unit match code. Over 75 percent of the people collected were in matched households (meaning the housing units matched census housing units) while over 22 percent were in non matched households. In addition, over one percent were not a housing unit. These may be attributed to a housing unit definition issue. The unit may have been a housing unit during the PI operation but not on Census Day. People that were collected may have moved into the housing unit during the PI operation but the unit was not a housing unit on Census Day. The people collected may have also been those of the

² A geocoding error means that the census unit was erroneously geocoded to the sample block cluster on the Census Unedited File, it actually exists outside of the block cluster and one ring of surrounding blocks.

unresolved kind where they could not be coded at the housing unit during Census Day (not enough information).

Figure 2: Percentage of People in Additional Housing Unit by Final Match Code



3.2 Tenuously Attached Persons to Housing Units

The CCM Person Interview operation attempted to include anyone who could possibly be counted at the sample address on Census Day, including those that may not consider the sample address their main residence. The instrument probed for the types of people that may stay at the address that are typically missed or may not have been thought of as part of the household. It also collected a roster in situations where the sample address may be transient, based on the type of unit, such a houseboat, recreational vehicle, or a long-term seasonal address, to ensure collection of any people staying there that had no other place to live.

Listed below is the main question wording for each roster and probe in the instrument. The six questions in bold are aimed at collecting those people that may be tenuously attached to the housing unit.

ROSTER 1 (Transient): *What are the names of the people staying here and have no other place where they usually live?*

ROSTER 2 (Main): *We'll start by making a list of everyone who lives or stays here now. Let's start with you. (Anyone else?)*

PROBE 1: *Is there anyone who has another place to live but stays here often?*

PROBE 2: *Is there anyone who is staying here until they find a place to live?*

PROBE 3: *Are there any babies, foster children, or other children who stay here that you didn't mention yet?*

PROBE 4: *Have I missed any relatives or unrelated people who live or stay here?*

ROSTER REVIEW: *I am going to show you the list of people I recorded. Is everything spelled correctly? Is the list complete?*

OUTMOVER: *Was there anyone else living or staying here during March or April who is no longer living here?*

OUTMOVER REVIEW: *I am going to show you the list of people who moved out or passed away. Have I spelled everything correctly?*

WHOLE HOUSEHOLD of OUTMOVERS ROSTER: *What are the names of the people who lived here on April 1, 2010? (Anyone else?)*

Table 3 shows the breakdown of where each person was collected. Note that in this table, the additional people that we collected may or may not be from the additional housing units in the previous sections.

Table 3: Person Interview Distribution of People by Roster/Probe where Added

Roster/Probe Where People were Added	Number of People Added	Percent of People Added
Roster 1: Semi-Transient People Roster	240	0.06
Roster 2: People living here now (Main Roster)	396,346	93.65
Probe 1: Stays often	7,909	1.87
Probe 2: No other place to live	1,058	0.25
Probe 3: Other children	1,010	0.24
Probe 4: Other relatives, people, etc.	489	0.12
Roster Review	277	0.07
Outmover Roster	9,153	2.16
Outmover Review	15	0.00
Whole Household of Outmovers Roster	6,745	1.59
Total	423,242	100.00

Source: Person Interview and Reinterview Output

As expected, most people (93.65 percent) were added in the main roster (Roster 2), but 2.53 percent of the people were added through the first roster question and probes 1 – 4. An additional 0.07 percent was added through the roster review question, which was asked towards the end of the interview to include someone the respondent may have forgotten to list in the household.

These additional probes were not well liked by either the respondents or the interviewers. They were considered repetitive and tedious. Often interviewers and respondents would shorten or cut off the questions. In the future, we believe that we need to do some probing to get the respondent to list people they may not consider to be a part of the household. But we need to further review how we probe to reduce resistance and make it more acceptable by interviewers and respondents. The next table shows some of the key characteristics for people collected from each of the probes.

Table 4: Review of PI Person Characteristics by Probe where Person was Included†

Column	1	2	3	4	5	6	Total People	
Characteristics of People	People that have no alternate address	People that are Non-White	People that are Hispanic	Peak Age Range for People	People counted at Sample Address at end of PI	People counted at Sample Address at end of CCM Person Matching		
All People	Count	343,652	143,451	81,162	N/A	334,159	361,489	423,242
	Percent/Age Range	81.20	33.89	19.18	None	78.95	85.41	100.00
Probe 1: Stays Often	Count	545	3,075	1,600	2,430	0	2,487	7,909
	Percent/Age Range	6.89	38.88	20.23	15-24	0.00	31.45	1.87
Probe 2: No other place to live	Count	325	483	226	264	315	394	1,058
	Percent/Age Range	30.72	45.65	21.36	20-29	29.77	37.24	0.25
Probe 3: Missing Children	Count	490	546	265	744	426	480	1,010
	Percent/Age Range	48.51	54.06	26.24	0-9	42.18	47.52	0.24
Probe 4: Missing relatives	Count	182	277	134	102	162	197	489
	Percent/Age Range	37.22	56.65	27.40	20-29*	33.13	40.29	0.12
Roster Review	Count	189	134	72	73	174	186	277
	Percent/Age Range	68.23	48.38	25.99	0-9	62.82	67.15	0.07
All Probes	Count	1,731	4,515	2,297	2,880	1,077	3,744	10,743
	Percent/Age Range	16.11	42.03	21.38	20-29	10.03	34.85	2.54

Source: Person Interview Final Version Output and Results of CCM Final Person Matching

†People are not unique to a row and people can be counted in more than one row. Hence, columns do not add to the total.

*Probe 4's peak age range contains only 20.86 percent of the people while all other peak age ranges contain at least 25 percent of the people.

A total of 423,242 people were collected in the PI operation. Column 1 of Table 4 shows the percent of people collected from the various probes that did not provide an alternate address. This ended up being 16.11 percent of all people interviewed. Some of the people from the probes may have not been listed in the main roster due to not understanding who should be listed. However, we think that people added at some of the probes might also

result from the people being tenuously attached to the household and likely having another address. The Probe 1 percentage for column 1 is lower than the other probes, as we expected because this probe specifically refers to people who have someplace else to live or stay and most could probably just been listed at the other address. The percentages for the remaining probes seem to be high percentages even though they are lower than the overall percent for all people reported. Some more research into people listed in the probes that have no other addresses reported should be done to see why they were not listed in the main roster.

Column 2 is the percent of probe people that are non-white and Column 3 is the percent of people who are Hispanic. Typically, minorities have a more complex household structure (Martin 2007 and Schwede et. al. 2005). We thought that probe people would be added given these complex household structures. We observed that 42.03 percent of people added through probes were non-whites and 21.38 percent were Hispanic. The probe people have a higher percentage of minorities than when all people are considered.

Column 4 looks at the peak age range for each probe. One of the populations typically missed within Census is young males (Martin 2007). We hoped each of the probes would collect this age range (except for Probe 3, which targeted babies and children). Results indicate that most of the people collected by the specific probes generally collected more of the people we were targeting.

The overall goal of the probes was to find people who should be counted at the sample address. However, it was more efficient to add people early who had a chance of being counted in the end. Column 5 shows how the person was coded at the end of PI post-processing. There was not one person added from Probe 1 who was listed automatically as being counted at the sample address. Column 6 is the final residence status the probe people received in CCM after all CCM Person Followup and matching operations. We can see that the person matching and followup activities were able to increase the percent of probe people counted at the sample address. In particular, these operations confirmed residence status at the sample address for 31.45 percent of the people from Probe 1 that the PI post processing could not confirm. As one can see, 34.85 percent of all probe people collected should have been counted at the sample address. These are people that were not listed in the main roster but collected through the probes and they could have been missed during the census operation at that sample address. For more information on the PI operation and results, please refer to the “2010 Census Coverage Measurement Person Interview Operation Assessment Final Report.”

4. Conclusion

The housing unit coverage question aided in the collection of 4,219 potential “hidden” housing units. Only 2,554 of these housing units remained in sample for subsequent CCM operations. Most of these units were of the “other” type. These types of housing units may have been missed if data collection was completed by observation, without the help of a respondent, since they can vary from being a tent, van, boat, or camper. Close to a quarter of these housing units were nonmatches, meaning that these were listed in our CCM program and not listed or captured during the Census operations. Over half of these additional housing units were matches to Census housing units. By having listers ask the potential basement or garage apartments coverage question of a respondent, the CCM survey was able to get a list of these hard to find units that may have been missed

otherwise. It helped ensure better housing unit coverage for subsequent CCM operations and to evaluate coverage error.

As far as coverage for persons in a household, during the PI operation, the additional probe questions for tenuously attached persons were not well liked by the respondents. Still, over two percent of the people collected in the PI operation were found through these additional probes. Out of these additional people, a high percent of them were younger and a minority, that is the people that were targeted. After the matching operations and all householders had been coded, we determined that over 30 percent of all the people collected through the additional probes should have been counted at the sample address. These people could have been missed during the census operations at that sample address.

These additional coverage and probing questions aided in the collection of additional housing units and people in households that may have been missed if not as asked at all during the CCM survey. Since we had a small number of housing units and persons collected by the probes (less than 3 percent) we cannot say for certain that these questions ensure complete coverage. It would be beneficial to explore these questions further in order to determine if they are necessary, considering the interview time and additional burden on the respondent to conduct all the probes. A cost analysis could be conducted to determine if it is cost effective to have interviewers out in the field, knocking on people's doors several times to get a respondent to provide answers to these questions.

Acknowledgements

We would like to thank Kyra Linse and the National Processing Center analysts for helping with this analysis.

References

Linse, Kyra & Argarin, Ashley C. (2012). "2010 Census Coverage Measurement Person Interview Operation Assessment," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-I-16, April 2, 2012.

Martin, Elizabeth(2007). "Strength of Attachment: Survey Coverage of People with Tenuous Ties to Residences," *Demography*: Vol. 44. No2, May, 2007.

Schwede, Blumberg, and Chan (2005). "Complex Ethnic Households in America," Oxford, UK.