

STUDY SERIES
(*Survey Methodology* #2014-04)

**Success of Applying Census Residence Rule
to Resolve Duplication**

M. Mandy Sha¹
Emily Peytcheva²
Ryan King

¹ RTI International

Center for Survey Measurement
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: June 24, 2014

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Success of Applying Census Residence Rule to Resolve Duplication

M. Mandy Sha¹ and Emily Peytcheva, RTI International
Ryan King, U.S. Census Bureau

¹Corresponding author:

Mandy Sha, Program for Research in Survey Methodology, Survey Research Division, RTI International, 230 West Monroe Street, Chicago, Illinois, USA, Email: msha@rti.org

Abstract

The U.S. Census Bureau aims to count each person living in the United States just once and only once, yet some people are duplicated or listed at more than one residence, for various reasons. Since Census 2000, computer matching is employed to identify potential duplicates. The Targeted Coverage Follow-up (TCFU) Interview has been developed to collect information about the living situations of the suspected duplicates. The information is to be used to resolve where those suspected duplicates should be counted, using the Census Residence Rule that determines people's usual residence by where they lived or stayed most of the time or on Census Day (April 1st) if there was no usual residence. In a large-scale cognitive testing conducted in 2010 and 2011, the TCFU questionnaire was tested with actual 2010 Census participants, including 226 suspected duplicates or a proxy from their household. Duplication was revealed and coded for 116 cases for analysis in this paper. We examined the completeness of the address and dates in the TCFU to determine whether the Census Residence Rule can be applied automatically without clerical intervention. We also looked at the success of applying the Residence Rule by different types of duplications. The findings suggested that there is possibility for automation in resolving duplication as part of the TCFU administration, while clerical effort is likely warranted to resolve inconsistencies in confirmed duplicates where the respondents did not always provide complete information (i.e. the address itself and the dates lived there) about the duplicate address, or when they were not certain of their whereabouts on the Census Day.

Key Words: Census residence rule, duplication, U.S. Census

Introduction

Starting in Census 2000, computer matching of person records identified suspected duplicates in the census. These were people who were suspected to have been counted twice at more than one housing unit (HU), or who were counted in a HU and also at a non-HU facility or shelter (i.e., group quarters (GQ), see complete definition of GQs in Census Bureau, no date). To resolve computer-identified duplication, the Census Bureau must apply the Census Residence Rule to the suspected duplicates' living situations and determine whether to remove a record from the household roster. According to the 2010 Census Residence Rule (Census Bureau 2010):

- Count people at their usual residence, which is the place where they live and sleep most of the time.
- People in certain types of facilities or shelters (i.e., places where groups of people live together) on Census Day [April 1, 2010] should be counted at the facility or shelter.
- People who do not have a usual residence, or cannot determine a usual residence, should be counted where they are on Census Day.

The Coverage Follow-up (CFU) questionnaire was designed to interview HUs with a potential coverage issue, either undercoverage or overcoverage. This includes interviews at HUs with suspected duplicates to resolve where they should be counted. The CFU interview was not specific to the type of coverage situation, but asked at the household level if any person had stayed at another place for reasons such as (but not limited to) college housing, custody, or a second home. This approach needed improvement because the respondents often did not indicate having another place where they (or the suspected duplicates) lived (Childs et al., 2009). In response, the Census Bureau developed the Targeted Coverage Follow-up (TCFU) questionnaire, which involved asking questions naming each duplicate, instead of broad household level questions as in the CFU. Specific, targeted questions were also asked of duplicates who were counted in a GQ. According to Childs et al. (2011), the TCFU asked targeted questions about the suspected duplicates based on (1) whether the phone numbers collected at each address of the suspected duplication matched, (2) the age of the people who were duplicated, and (3) whether the match was a HU-HU match or a HU-GQ match. A cognitive test was conducted in 2009 to examine whether respondents understood the experimental TCFU questions, as well as whether the questions functioned as intended to reveal other addresses (Childs et al., 2011).

In 2010, the TCFU, revised after the 2009 cognitive test, was tested in a large-scale qualitative study with computer-identified suspected duplicates from the 2010 census. One of the primary objectives of this research was to cognitively test the TCFU questionnaire for issues related to privacy, confidentiality, comprehension, and exhaustiveness of the questions and to gain first-hand knowledge from individuals who were identified as suspected duplicates on all possible living or other arrangements that lead to duplication and to understand what sensitivities are related to those situations. A secondary goal of the study was to evaluate the extent to which the TCFU collects sufficient information about addresses and patterns of movement to be able to apply the Census Residence Rule and determine where a person should be counted. Patterns of movement would be determined through asking dates of their time at each address provided or mentioned in the TCFU; how often they lived or stayed at those places during those dates; and whether they were at those places on Census Day. Additionally, the possibility of confirming residency status of the duplicate through automated processing was of interest.

Using cognitive interview data collected from this large-scale TCFU qualitative study, this paper examines the completeness of the address and date information of the confirmed duplicates to

determine whether the Census Residence Rule can be applied automatically without clerical intervention.

Methods

A total of 226 cognitive interviews with suspected duplicates or their proxies (anyone in the HU that is not a suspected duplicate) were available from the TCFU qualitative study. The sample covered 18 HU and 9 GQ living situations that had been organized in three main groups:

- Phone match cases: a person was reported in two different HUs that provided the same phone number on each census questionnaire. The TCFU was designed for the interviewer to confirm both addresses with the respondent, a previously forbidden procedure due to confidentiality concerns.
- Nonphone match cases: a person was reported in two different housing units that provided different phone numbers on each census questionnaire. The TCFU asked about only one of the HU addresses (the address the respondent was contacted at) and was designed to draw out the second address if the respondent chose to reveal it.
- GQ cases: a person was reported in a GQ as well as in a HU. The TCFU asked only about the HU addresses (the address the respondent was contacted at) and was designed to draw out the GQ address if the respondent chose to reveal it. The GQs covered on this study included military quarters, college group housing, jail or correctional facilities, juvenile facilities, group homes, nursing homes, workers residential facilities, and homeless shelters or soup kitchens.

There may be one or more duplicates in a household; the Phone and Nonphone match cases were categorized further into the quantity and makeup of the household members who were duplicated. The details can be found in Peytcheva, Sha, Gerber et al. (2012).

As mentioned in the introduction, the TCFU collects additional addresses for each duplicate person in the household. After respondents were asked about an address at which they stayed during 2010, they were next asked about dates when they stayed at that address. As with real-life studies, respondents did not always provide complete information in the questionnaire. Some addresses or the associated dates were incomplete or they were uncertain about their whereabouts on April 1, 2010. Among the 226 cognitive interviews, duplication was confirmed for 155 people across two rounds of testing (39 interviewed in the first round and 116 in the second round). We limit our analyses to the second round of testing because starting in round 2, each interviewer followed an elaborate coding scheme and coded whether dates and addresses were complete or partial. Dates were considered complete when exact dates were provided or when the information could still be used for applying the Census Residence Rule (e.g. “January

to May”). The interviewers also coded what respondents told them about their whereabouts on Census Day, which was April 1. We used the 116 confirmed round 2 duplicates for our analysis.

At the beginning of the cognitive interview, the TCFU questionnaire was administered (mimicking a real interview), followed by retrospective probing in order to evaluate respondent understanding, their living situation, and issues with privacy and confidentiality. For each address that the respondent reported staying at during 2010, they were asked about the dates when they stayed at that address. Because the TCFU data was collected during the cognitive interview, they can be used to examine the completeness of address and date data from the 116 confirmed duplicates.

A coder used the addresses and dates collected in the TCFU only to apply the Residence Rule, and determine which address was the one where the duplicated person stayed most of the time. This address was called the “primary” address. If there was insufficient dates or frequency of stay information in the TCFU, the address stayed at on Census Day was coded as the primary address (except GQ cases where Census Day was always used according to the Residence Rule). Another variable, “true” address, was created to reflect where the respondent should be counted based on all the information gathered in the TCFU *and* the cognitive portion of the interview, which included extensive probing to make sure there was detailed information about each individual living situation. The difference between the “primary” and the “true” address was of main interest to us as it is informative about the extent to which data from the TCFU only is sufficient to automatically apply the Residence Rule. In other words, we define success in applying the Residence Rule using TCFU data only if the primary and true addresses were the same.

In addition, we suspected that successful application of the Residence Rule would be associated with the circumstances surrounding the duplication. Therefore, we examined the success of applying the residence rule using various information about the case, such as type of match, type of respondent, number of duplicates, and reasons for duplication.

Our research questions were:

1. How complete were the address and date data from the confirmed duplicates and how well could the Census Residence Rule be applied using that information?
2. What was the success rate applying the Census Residence Rule for each of the following:
 - Type of match (Phone, Nonphone or GQ match)
 - Type of respondent (suspected duplicate vs. proxy)
 - Number of duplicates in the household;
 - Reason for duplication (e.g., second home, place respondent moved from, etc.),

3. What was the possibility for automation of the verification procedure as part of the TCFU administration?

Results

Completeness of duplicate address and associated dates in the TCFU

As shown in Table 1, the 116 confirmed duplicate cases were divided across interview types as follows: 28 cases were “Phone match” interviews, 55 cases were “Nonphone match” interviews and 33 cases were “GQ” interviews (see the beginning of the Methods section for a description of each category).

Table 1. Number of confirmed duplicates by the type of match

Type of case	Number of confirmed duplicates
Phone match	28
Nonphone match	55
GQ	33
Total	116

For all 28 Phone match cases, complete duplicate addresses were available in the TCFU data because the TCFU was designed to confirm both sides of the duplicate addresses in a Phone match situation (This was when the duplicate was reported in two different HUs who gave the same phone number on each census questionnaire). All of the addresses were confirmed. In comparison, fewer complete duplicate addresses were reported in the Nonphone match and GQ cases. As shown in Table 2, out of the 55 Nonphone match cases, 32 (58%) provided complete addresses in the TCFU. Among GQ cases, only about a quarter of duplicate addresses (8 of 33) were complete.

Table 2. Completeness of duplicate address

	Complete addresses	Partial Addresses	Not reported in TCFU	Total
Phone match	28 (100%)	0 (0%)	0 (0%)	28 (100%)
Nonphone match	32 (58%)	18 (33%)	5 (9%)	55 (100%)
GQ	8 (24%)	24 (73%)	1 (3%)	33 (100%)
Total	68 (59%)	42 (36%)	6 (5%)	116 (100%)

In terms of completeness of dates associated with the duplicate addresses (regardless of the completeness of those addresses), Table 3 shows that 57 of the 116 (49%) confirmed duplicates

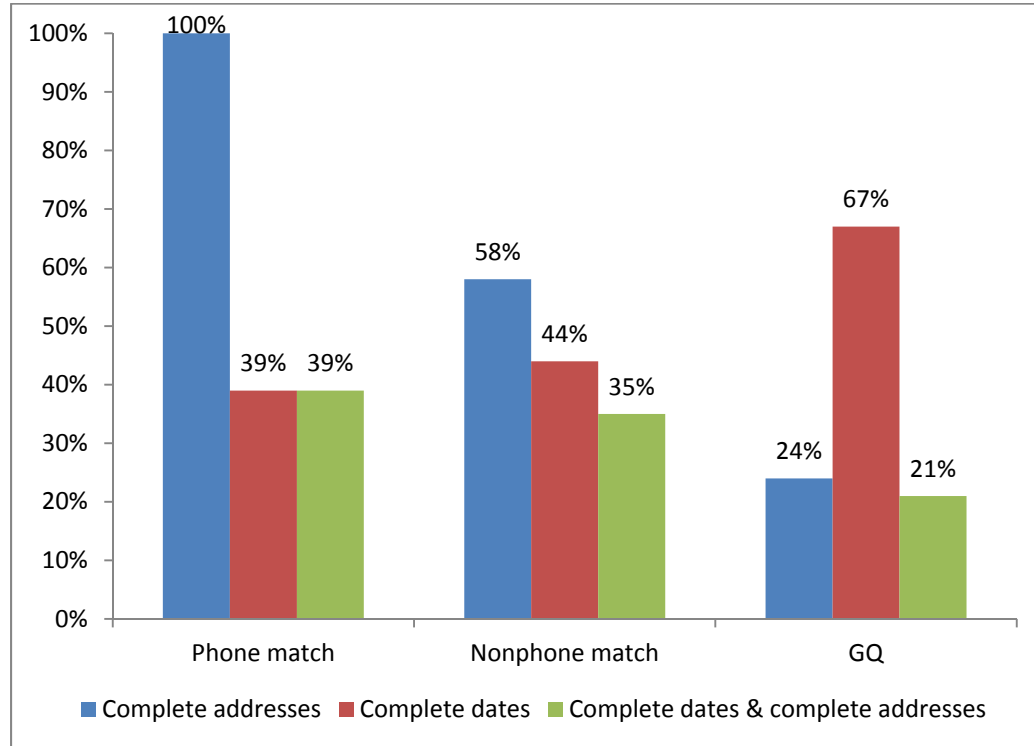
provided complete dates for each of the duplicate addresses. When the duplicate addressees were examined, the addresses for 68 of the 116 cases (59%) were complete. When the completeness of *both* the reported duplicate address and date information were examined, 37 of the 116 confirmed duplicates (32%) qualified. Looking at these results by interview type, we found that all Phone match cases that provided complete dates also provided a complete addresses (but not the other way around). A total of eight GQ cases provided complete addresses and seven of those also provided complete dates for those addresses.

Table 3. Completeness of dates associated with duplicate addresses

	Complete addresses	Complete Dates	Complete Address and Date	Number of Confirmed Duplicates
Phone match	28 (100%)	11 (39%)	11 (39%)	28
Nonphone match	32 (58%)	24 (44%)	19 (35%)	55
GQ	8 (24%)	22 (67%)	7 (21%)	33
Total	68 (59%)	57 (49%)	37 (32%)	116

Figure 2 visualizes the numbers reported in Table 3. It illustrates that the Phone match and Nonphone match cases looked similar in terms of complete dates provided for both duplicate addresses – 11 out of the 28 Phone match cases (39%) and 24 out of the 55 Nonphone match cases (44%) provided complete dates. The GQ cases appeared very different from the rest – 22 out of the 33 respondents (67%) provided complete dates for both addresses. However, only a little more than 20% of GQ cases had both a complete date and a complete address together, which was almost less than half of cases in Phone match (39%) and Nonphone match (35%) cases.

Figure 2. Completeness of addresses vs. dates vs. both dates & address



Success of applying the Residence Rule using TCFU information

As indicated in the Methods section, address and dates in the TCFU were used by a coder to determine the “primary” residence, and a “true” residence was discerned from both TCFU and cognitive interview debriefing details. Based on the TCFU information only, we were unable to determine the “primary” residence for 9% (11) of the 116 cases because there was insufficient information on the address, the dates, or the frequency of stay in the TCFU, but the cognitive interview debriefing had enough details to determine all (116) of the “true” residences. We found the “primary” and “true” residence were the same in 88 cases, which constituted the majority (76%) of the 116 confirmed duplicates.

By interview match type

The success of using TCFU data to apply the Census Residence Rule differed by match type. As shown in Table 4, the Phone match interviews yielded the highest number of successful identification of the address where the respondent should be counted – in 27 out of the 28 cases (or 96%), the “primary” residence determined using the TCFU data was consistent with the “true” residence. The majority of the Nonphone match cases were also successful in applying the Residence Rule – 46 out of the 55 cases (84%).

Among the 33 GQ cases, only in 15 (45%) cases the “primary” and “true” residence concurred. Many GQ persons reported not considering themselves living or staying in a GQ, as they had another permanent residence. However, the Residence Rule states for many GQ situations that if a person is there on April 1, then they should be counted at the GQ.

Table 4: Success of applying the Residence Rule by interview type

Interview Type	“Primary” = “True”		Grand Total
	No Success	Success	
Phone Match	1 (4%)	27 (96%)	28 (100%)
Nonphone Match	9 (16%)	46 (84%)	55 (100%)
GQ	18 (55%)	15 (45%)	33 (100%)
Total	28 (24%)	88 (76%)	116 (100%)

By reasons of duplication

Table 5 shows that owning another property and moving from one address to another were the most frequently cited reasons for duplication among Phone match cases and applying the Census Residence Rule was straightforward. The top reasons for duplication among Nonphone match cases were tri-fold and included; moving from one address to another, owning another property and staying at a relative’s address. Among these three reasons, cases where the duplication occurred because of another property ownership were most likely to be successful in determining where the duplicates should be counted when applying the Census Residence Rule (92% of the cases), however the other two also had high success rates (77% for moving, and 75% for staying at a relatives). Among the GQ cases, College, Jail and Nursing homes had the highest number of confirmed duplicates in our sample but College was the most successful in terms of resolving duplication – 6 in 7 were resolved.

Table 5: Success of applying the Residence Rule by reason for duplication and by match type

“Primary” = “True”				
Type	Reasons for duplication	No success	Success	Total
Phone match	Completed wrong form		2	2
	Convenience address (e.g. mail, school)		1	1
	Moved		6	6
	Other property, such as a Second Home	1	14	15
	Relatives household		4	4
	Subtotal	1	27	28
Nonphone match	Convenience address (e.g. mail, school)		4	4
	Custody		5	5
	Moved	4	13	17
	Other property	1	12	13
	Relatives household	4	12	16
	Subtotal	9	46	55
GQ	Military Quarters	2	0	2
	College	1	6	7
	Jai or Correctional Facilities	5	2	7
	Juvenile Facilities	1	0	1
	Group Homes	1	0	1
	Nursing Homes	4	3	7
	Workers Residential Facilities	2	2	4
	Homeless Shelters or Soup Kitchens	2	2	4
	Subtotal	18	15	33
Grand Total		28	88	116

By type of respondent

Table 6a shows the counts of cases for which we were able to successfully apply the Residence Rule by type of respondent. In 79% (50 of 63) of the cases when the respondent was the duplicate, it was successful to use TCFU data to determine where he or she should be counted. The success rate for proxy respondents was lower, but nevertheless not drastically different, at 72%. When match type is examined, we found that regardless of the type of respondent, almost all Phone match cases and the majority of Nonphone match cases were successful. On the other

hand, we had much lower success with GQ cases when the respondent when interviewing either the duplicate or proxy.

Table 6a. Success of applying the Census Residence Rule by self vs. proxy

Interview Type	“Primary” = “True”		Grand Total
	No Success	Success	
Self	13 (21%)	50 (79%)	63 (100%)
Proxy	15 (28%)	38 (72%)	53 (100%)
Total	28 (24%)	88 (76%)	116 (100%)

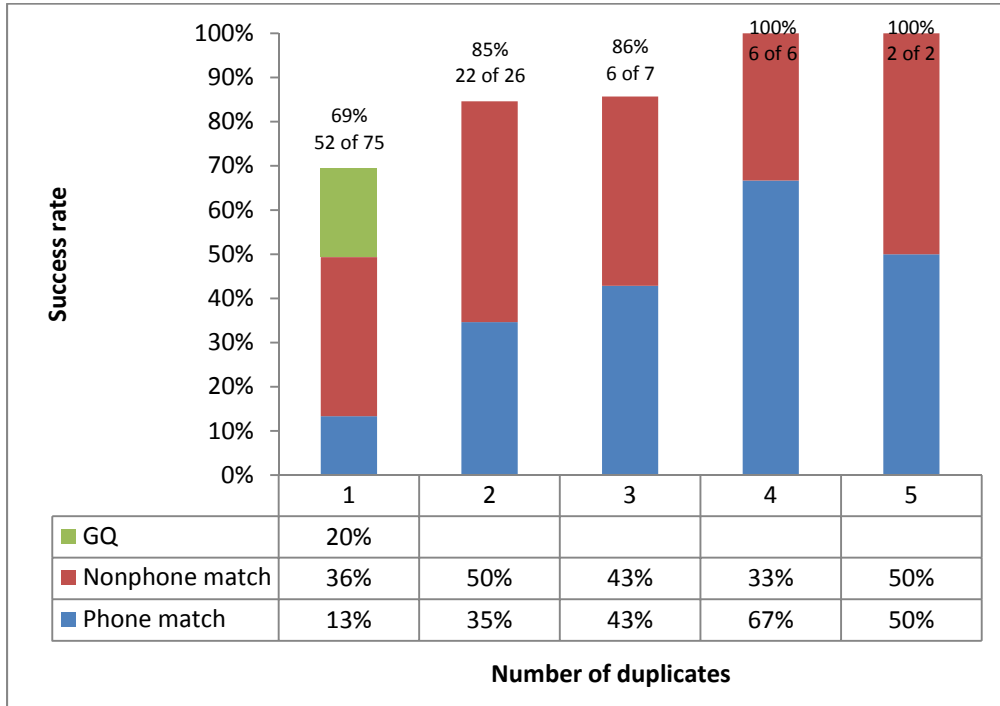
Table 6b. Success of applying the Census Residence Rule by self vs. proxy per match

Interview Type	“Primary” = “True”		Grand Total
	No Success	Success	
Self	13 (21%)	50 (79%)	63 (100%)
Phone match	0 (0%)	23 (37%)	23 (37%)
Nonphone match	5 (8%)	22 (35%)	27 (43%)
GQ	8 (13%)	5 (8%)	13 (21%)
Proxy	15 (28%)	38 (72%)	53 (100%)
Phone match	1 (2%)	4 (8%)	5 (9%)
Nonphone match	4 (7%)	24 (45%)	28 (53%)
GQ	10 (19%)	10 (19%)	20 (38%)
Total	28 (24%)	88 (76%)	116 (100%)

By number of duplicates in the household

Figure 3 shows that when there were more than one duplicate in the household, our chances of resolving the duplication were higher. Success rate was at 100% when there were 4 or 5 duplicates in the household, as compared to 69% when there was only one duplicate. The table below Figure 3 shows that there is no clear pattern by match type when there were more than two duplicates in the household and the number of cases was small. When there were one or two duplicates in the household, in general, there were more Nonphone match cases than Phone match cases. The GQ match cases by definition only have one duplicate, and we include them for illustration purpose.

Figure 3. Success of applying the Residence Rule by number of duplicates



Discussion

Overall, we were successful in applying the Residence Rule in 88 of the 116 confirmed duplicates (76%). This was determined by confirming that their “primary” residence identified from their TCFU responses only was consistent with the “true” residence that were discerned from both TCFU and the cognitive interview debriefing. This result showed that respondents’ true address generally aligned with what Census Bureau would conclude based on information provided in the TCFU.

In addition, match type (Phone, Nonphone, and GQ matches) seemed to be closely associated with the success of applying the Residence Rule in the TCFU data. All except one of the phone match interviews resolved duplication using TCFU data only, followed by success in 84% of Nonphone match interviews, and in less than half of the GQ cases. The GQ cases were the least successful type. Potential TCFU automation would likely be successful in resolving duplication in phone match cases, where the same phone number appeared on both census questionnaires. For example, the TCFU may be revised to provide an immediate opportunity for the Phone match respondent to acknowledge that a permanent move took place (with probing from the recruiter/interviewer to confirm that the move was indeed permanent) and then skip the rest of the TCFU interview because that was the reason for duplication. However, automation in Nonphone match interviews might be less straightforward and may require clerical intervention

due to only contacting one household or due to the possibility of receiving discrepant information should both households be contacted.

We found that the completeness of address and dates information was also dependent on interview type. The Phone match and Nonphone match cases collected more complete duplicate addresses (100% and 58%, respectively) than GQ cases (24%), but the opposite was true for complete dates. GQ cases had 67% reporting complete dates, followed by Nonphone match cases (44%) and Phone match cases (39%). However, a cross-tabulation of complete dates and complete addresses showed that both Phone and Nonphone match cases almost doubled the number of GQ cases that were able to associate complete dates and complete addresses together. The finding that GQ cases were more knowledgeable about the dates associated with duplicate addresses than the address itself may not be surprising. Being institutionalized could be a memorable event in one's life and respondents might be more likely to remember when and for how long they stayed in a GQ. They might also be able to describe or name the institution they stayed in, thus providing partial address information, but did not necessarily know its exact address. Given that Census Day whereabouts was crucial to resolving duplication when a GQ address was involved, future research may consider revising the TCFU to provide more emphasis on dates surrounding the Census Day, rather than "most of the time". Clerical work would be required to establish a complete address to verify duplication, but could be facilitated when a name or description was provided.

Furthermore, the number of duplicates in the household also drove the success of applying the Residence Rule using TCFU, but there was no clear pattern when a specific type of respondent (duplicate or proxy) was interviewed. These findings were intuitive. When there were more than one duplicate in the household, it reasoned that the chances of success of resolving duplication were higher because the duplication would have likely occurred for a family or people who knew each other. For example, moving from one residence to another and owning a second home were both top reasons for Phone and Nonphone match duplication. The respondents would possibly know more about where the duplicates were throughout the year. Perhaps not surprisingly, success in applying the Residence Rule was high (nearly 80%) when the respondent was the duplicate him or herself determining where he or she should be counted. The success rate for proxy respondents was also high, at 72%, which suggested that they might be knowledgeable enough about the duplicates and could be more readily relied upon when the duplicate him or herself was not available. This may apply to duplication situations where success was already expected to be high, such as Phone match cases or when several duplicates reside in the same household. In determining if a change to the Residence Rule is warranted, additional research should examine data quality provided by proxy versus duplicate respondents, such as privacy concerns, refusals, and number of probes needed to elicit address and dates.

Limitations

This study has several limitations. First, there were some recall issues because some respondents

were being screened in the spring of 2011, which was more than a year after Census Day. Second, because of small cell sizes, our analysis did not examine success by specific living situations, even though some of the living situations were quite complex (e.g. Nonphone matches and partial household matches involving adults and children). Third, because the recruitment and interviewing were designed to be conducted in English only on this TCFU qualitative study, the success in applying the Census Residence Rule could not be generalized to duplicates in non-English language dominant households. This is clearly an area for future research.

Acknowledgements

Disclaimer: This paper is released to inform interested parties of research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

The authors acknowledge Jennifer Hunter Childs and Sarah Heibel of the Census Bureau who co-designed the TCFU questionnaire and the 2010 TCFU qualitative research study. We also thank the RTI-RSS team for collecting the cognitive interview data.

References

- U.S. Census Bureau (2010). Census Residence Rule. [WWW page]. URL https://www.census.gov/population/www/cen2010/resid_rules/resid_rules.html
- U.S. Census Bureau (no date). Decennial Management Division Glossary: Group Quarters [WWW page]. URL <http://www.census.gov/dmd/www/glossary.html#G>
- Childs, J. H., Fernández, L., Clifton, M., & Meyers, M. (2009). CFU CPEX experimental question cognitive testing: Undercount, overcount, and duplicate experimental question sequences. Study Series, Survey Methodology #2009-17. Washington, DC: Statistical Research Division, U.S. Census Bureau.
- Childs, J. H., Sorokin, A., & Jurgenson, N. (2011). 2009 targeted coverage follow-up (TCFU) cognitive testing. Washington DC: Statistical Research Division, U.S. Census Bureau.
- Peytcheva, E. A., Sha, M., Gerber, E., Cook, S. L., Schoua-Glusberg, A., King, T. R., Kenward, K., Childs, J., Heibel, S., & King, R. (2013). *Qualitative interviewing with suspected duplicates and cognitive testing of the Targeted Coverage Follow-up (TCFU) interview* (Survey Methodology # 2013-09). Research and Methodology Directorate, Center for Survey Measurement Study Series. U.S. Census Bureau. <http://www.census.gov/srd/papers/pdf/ssm2013-09.pdf>