Managing Quality on a Large Qualitative Research Study with Complex Respondent Recruitment Criteria[1]

M. Mandy Sha[1], Katherine Kenward[2], Julie Feldman[1] and Sarah Heimel[3]

[1]RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709
[2]Research Support Services, 906 Ridge Avenue, Evanston, IL 60202
[3]U.S. Census Bureau, 4600 Silver Hill Road, Suitland, MD 20746

Abstract

Qualitative methods, such as cognitive interviewing, are often used to pretest questionnaires. Commonly observed limitations include a small number of respondents, purposively chosen recruitment criteria, and interviewer variability. Managing quality associated with these operational aspects is particularly important to the success of large qualitative research studies with complex recruitment criteria. In 2010-2011, Census Bureau conducted a large-scale cognitive testing study of the Targeted Follow-up (TCFU) questionnaire, a detailed instrument designed to resolve instances of people who appear to have been duplicated in the Census. This study faced the challenge of completing 226 cognitive and 50 ethnographic interviews over a total of 6 months and two rounds, using a team of 13 highly-qualified interviewers in five states and 7 recruiters experienced with cold-calling. The majority of the interviews were conducted in the field with actual 2010 US Census participants recruited from 27 living situations (both households and Group Quarters), without the respondents knowing that they were duplicated. Stringent requirements were put in place to protect respondent confidentiality and privacy mandated by Title 13. This paper presents a description of the methods to manage quality and estimates their impact. We measure quality by the success of representing a diversity of living situations and duplications, the quality of the TCFU information collected, and interviewer and recruiter productivity. Our efforts to manage quality centered on making continuous improvement on recruitment strategy-setting and monitoring, interviewer selection and training, and communications and logistics management.

Key Words: quality, cognitive testing, Census coverage, project management

1. Introduction

In 2010-2011, the Census Bureau conducted a large-scale qualitative study to investigate why people could be duplicated (counted twice) in the decennial Census. The results are to be used to inform methods to minimize duplication in the 2020 Census.

The Targeted Coverage Follow-Up (TCFU) questionnaire was designed by the Census Bureau to interview suspected duplicate households and resolve duplication, without violating privacy, sensitivity, and Title 13 concerns.[2] Through two rounds of

[1] 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.

[2] See Childs, Sorokin & Jurgenson (2011) for details about the TCFU and prior duplication research.

interviewing over 6 months, a total of 226 cognitive interviews were completed to test the TCFU. Interviewers used a protocol guide that had standard probes but also probed spontaneously based on respondents' reactions and responses. An additional 50 semi-structured, ethnographic-style interviews were also completed to gain in-depth understanding of the duplicate living situations. The interview participants were actual respondents in the 2010 Census who were identified to be suspected duplicates by a computer algorithm. Some participants were reporting for household members who were suspected duplicates. While the decennial Census is mandatory, participation was voluntary and paid in this research study. Interviews were completed within an hour in Illinois, North Carolina, New York, and their neighboring states, as well as California. All interviewers and recruiters were highly-qualified and experienced.[3]

While large-scale quantitative surveys often interview thousands of respondents and can employ hundreds of interviewers, it is not as common for qualitative research studies to include a large sample and team of interviewers. This is because qualitative interviews are often semi-structured and data has to be coded and organized using transcripts or detailed interview summaries. They also do not require a representative sample. For cognitive interviews, Willis (2005) describes the logistics of cognitive testing to involve 9 to 15 respondents to be repeated for at least two rounds. Respondent recruitment usually depends on quota sampling recruited through advertisements or "word of mouth" and the selection focused on representing a great cross-section of the population. He also recommends involving "several interviewers" and using interviewers who "have a history of cognitive interviewing experience… [and] questionnaire design experience". In contrast, the TCFU research study was a large-scale effort in terms of the number of respondents interviewed, the complexity of the respondent recruitment, and a large team of highly-skilled interviewers and recruiters. In addition, although all Census Bureau research studies uphold confidentiality standards, the TCFU study required both confidentiality and privacy of the respondents to the most stringent standards required by the Title 13 law, and not generate negative public reaction. This paper describes the efforts to manage quality on this large qualitative study and provides an estimate of the impact of these efforts on quality.

2. Methods

According to the Project Management Institute (PMI), quality is "the degree to which a set of inherent characteristics fulfills requirements." (PMI 2008). Fulfilling requirements meant that the project produced what it was created to produce for the Census Bureau, which was to provide a qualitative picture of the issue of duplication. The study protocols needed to address the following characteristics that affect quality:

1) Completion of all work associated with almost 300 interviews by September 30, 2011 to inform the TCFU operational study
2) Respondents were recruited from a list sample of suspected duplicates from the 2010 Census
3) Narrow respondent selection criteria (see the 27 living situations listed in Table 3-1 in *Section 3 Results*)
4) Highly-skilled recruiters who are experienced with cold-calling
5) Highly-qualified interviewers who collect quality TCFU information
6) Familiarity with the 129-page TCFU questionnaire

[3] See Peytcheva, Sha, Gerber et al (2012) for greater details about the study design and findings.

7) Ease of navigating in the duplicated persons list (there are two "sides" of a household or GQ pair, causing duplication)
8) Most interviews were conducted in private locations in the field, not in a laboratory environment
9) Protection of confidentiality and privacy stipulated by Title 13: respondents cannot know who was duplicated, or even that duplication occurred. Revealing the information could potentially violate Census Bureau's confidentiality promise for the decennial census, as it is possible that duplication was because of someone else's census form completion.
10) Stringent standards: the list sample was secured in a laptop with no USB port or DVD drive, no internet capability, and double-password protected; personally-identifying information (PII) was only transmitted via paper and telephone; specific criteria were used when talking to all types of sample members
11) Control cost

In the following sections, we describe our efforts to manage quality on this study. We discuss three main factors that contribute to quality for this study: respondent recruitment, interviewing and protection of confidentiality and privacy, and communications. All methods were governed by the project management plan created by the RTI project manager. Although we present the methods as a whole, it is important to note that we made continuous improvements on the methods based on regular recruiter and interviewer debriefings, results from audits, and the priorities communicated by the Census Bureau research team. The improvements focused on preventing risks from being realized, implementing corrections, and minimizing rework.

Respondent Recruitment

Unlike quota sampling, we could only recruit from the list of suspected duplicates in the 2010 Census furnished by the Census Bureau, without telling respondents the reason they were contacted. We used the following methods to conform our performance to this requirement:

Large Team of Highly-Skilled Recruiters

Because we had to contact many households to render about 300 interviews by September 2011, we assembled a team of seven recruiters who were experienced at cold-calling and managed by two co-leads. All of the recruiters had a college degree, except one who completed some college. Four of the seven recruiters also had extensive experience interviewing and recruiting in the field.

Although the recruiters were quite experienced, we provided them with project-specific training so that their performance could be measured against the protocol. First, recruiters were trained to review their recruitment assignment to familiarize themselves with the 'story' behind the case. Once recruitment calls were made, all contacts were recorded on a paper record of call. Next, recruiters used a recruitment guide to administer screening questions. We chose to allow recruiters to formulate their own introduction by referring to a list of bullet points with information that had to be covered such as the voluntary nature of the study, etc. The screening questions were scripted because Census Bureau provided very specific guidelines about how to verify that we had reached the intended household without letting them know that they were duplicated in the census. In the

second round, Census Bureau required that recruiters screen out households that had made a permanent move between March and August of 2010 (the decennial census field period). At one-third of the recruitment period when recruiters became familiar with the screening guidelines, we allowed them to tailor the questions if the flow of the recruitment call naturally led to it.

Second, we trained interviewers on how to navigate with ease the duplicated persons list by overcoming the "constraints" placed on this process to protect the confidentiality and privacy of the 2010 Census participants. The list resided on Census Bureau-issued, secure laptops. PII contained in the list could not be printed or emailed, and information could only be shared by telephone or on paper via traceable mail. There were up to 74 columns in the laptop that provided crucial information for recruitment, for which recruiters needed to learn to navigate and apply to the respondent screening process. The files were organized by distance to the interviewers and by the 27 living situations.

Recruiters were also trained to identify potential "false matches" and bring them to the attention of managers for possible exclusion prior to initiating the screening call. These were households that appeared to have been misidentified during the computer matching process. Recruiters were trained to pay special attention to name variations, relationships of the duplicate to other family members on each side, and distance between sides. One of the recruitment co-leads sent cases of suspected false matches to Census Bureau for determination.

Strategic Recruitment Sample Selection

The RSS recruitment co-lead created an algorithm to select the recruitment sample and assign them to the recruiters. Because 27 living situations (including both Housing Unit and Group Quarters) needed to be represented in the interviews, they were the first selection criteria.

Based on Census Bureau's requirements of including less studied situations, the next criteria was to give preference to these duplicates, such as:

- Households with more than one duplicates
- As much as possible, a variety of duplicate age ranges within each case
- Greater geographic distance between the duplicate households for a case
- Two sides of a duplicate pair which were geographically diverse in mileage
- Complex household compositions beyond the typical nuclear family duplications
- Specific living arrangements, such as Snowbirds (seasonal movers)

We had to analyze the sample closely to identify these types of cases and to monitor recruitment progress to ensure that we did not over-recruit one type of household while the resources could have been redirected to difficult-to-reach households. This was quite challenging because the complex recruitment criteria created permutations of sample selection possibilities. The recruitment co-leads tracked recruitment progress on a daily basis and monitored the permutations regularly.

Advance Letters

Our risk analysis in the project management plan showed that allaying respondent concerns regarding legitimacy of the study would be a high impact risk. To increase the study's credibility and reassure respondents of the Census Bureau's commitment to

confidentiality and privacy, the Census Bureau sent out advance letters that introduced RTI and RSS as trusted partners. To impress upon respondents the importance of the study, the letter was signed by the Director of the Census Bureau. Once a case was selected for assignment to a recruiter, an advance mailing letter was generated and was sent to one side of a case (in some cases, both sides) prior to starting the telephone recruitment, although the cold-calling method complemented the recruitment when necessary. To assure confidentiality of the duplicated respondents, the salutations in the letters and on the envelopes were generic.

Interviewing and Protection of Confidentiality and Privacy

We assembled a team of 13 highly-qualified interviewers to complete the large number of interviews. More than half were credentialed and published authors, had a mean of 14 years of experience (ranging from 3 to 27 years), and five were experts on census coverage issues. Similarly with the recruiters, all interviewers received project-specific trainings so that their performance could be measured against the protocol. The trainings familiarized the interviewers with the 129-page TCFU questionnaire, how to prepare for the interviews (there was a checklist of 21 mandatory and recommended items for an interview), staying safe in the field, and protocol to protect confidentiality and privacy.

Communications

Because of the large team assembled for this study, we knew that planning for communications would be crucial to quality. Among the seven recruiters, two recruitment co-leads, and 13 interviewers, there were up to 231 lines of communication. For the management teams, there were up to 36 lines of communication among the 3-person Census Bureau research team and the six RTI-RSS leadership team members who had direct contact with the Census Bureau.[4] We describe our methods below for communications.

<u>Among Recruiter and Interviewers</u>

The 13 interviewers and nine recruiters/co-leads were not co-located. Information distribution among them was mostly electronic or by telephone. In specific, interviewer availability, interview appointment time, and Title 13-protected PII had to be communicated swiftly and accurately to prevent rework. We used the following communication channels:

1) Interactive, regularly scheduled meetings between recruiters and their lead
2) Web-accessible calendar to track interviewer availability used by recruiters for appointment setting
3) Standard email form for non-PII appointment information, followed up by a telephone call for PII
4) Reminder calls to respondents about upcoming appointments

When we started to experience a higher number of "no shows" (respondents who scheduled appointments, agreed on the location and then did not show up), we provided recruiters and interviewers with a protocol about rescheduling these interviews. The protocol covered the level of effort (number of times called) and who should follow up between the recruiter and the interviewer. No shows and rescheduling could mean

[4] Lines of communication is calculated using this formula: N (N-1)/2 where N=the number of team members. Communication hierarchy also existed among RTI-RSS management team and the interviewers, as well as between RTI and RSS. They used the same principles described in this section for the two most complicated hierarchies and therefore are not elaborated in this paper.

rework, thus requiring additional time and cost that could have been avoided. Interviewers were required to alert the recruitment co-leads immediately when it happened.

Among Census Bureau and RTI-RSS Project Management Team

The Census Bureau research team consisted of three researchers who shared equal decision-making authority. RTI-RSS management team included a project manager, two recruitment co-leads, two lead researchers, and associate project manager. To facilitate communication, the RTI project manager was always the only person who communicated with the Census Bureau research team, except when an issue needed to be addressed by a specific team member. We used the following communication channels:

a) Regular meetings focused on resolving open issues and approving changes

Census Bureau monitored the technical performance of the recruiters and interviewers by having a weekly telephone conference call with the RTI-RSS leadership team. At least 48 hours prior to the meeting, the RTI project manager sent an agenda that included a status report and a list of open issues. During the meeting, status updates were kept at a minimum (because they were in the agenda already reviewed by the Census Bureau) and participants focused most of the time on closing open issues and approving or rejecting changes to the project scope. All decisions were documented in written meeting minutes that were approved by the Census Bureau team.

b) Use of appropriate communication methods at the right time

We also convened ad-hoc meetings and emails to follow up on decisions that could not be resolved during the weekly meeting. Whenever there was a complex issue or when multiple emails could not reach a consensus, the RTI project manager would schedule an interactive meeting to ensure immediate feedback and clarifications. For example, Census Bureau had high expectations about what should be included in the final report and provided detailed written comments after reviewing the first draft. The RTI project manager set up a series of short meetings to discuss the expectations and confirm the intent of the comments. At times, only one of the 3-person Census Bureau research team could attend a meeting, but the representative always brought the others up to date.

c) Pre-agreement on ways to organize the interview data

The interview data was recorded in the form of interview summary reports using electronic word processing files that were stricken of PII. Depending on the respondents' living situation, an interview summary could range between five to 30 pages. Although one official summary report exists for each of the 276 interviews, we estimate at least 600 files were exchanged between Census Bureau, RTI, and RSS by the end of the project. One of Census Bureau's requirements was that the lead researchers review each summary prior to submission to the Census Bureau. Multiple versions of the summaries were created internally when interviewers revised the reports based on lead researcher feedback. Census Bureau also required revisions to some of the summaries, resulting in additional exchanges.

At the onset of the project, Census Bureau and RTI-RSS agreed on ways to organize this large amount of data, including the following methods:

- Use consistent file naming convention. Each file name embedded critical identifiers such as round, cognitive or ethnographic interview, unique Case ID,

identifier for the living situation, and file version number, for example: TO001_C2_SC80964_2d_v2.docx.

- Implement version control. At the end of the filename, a version number was indicated, e.g., _v1 for the first version.
- Regularly scheduled electronic and FedEx delivery. On a weekly basis, the lead researchers submitted password protected electronic files via email to the Census Bureau. To reduce file size and to add another layer of security, individual files were compressed in a password-protected zip file. After the electronic delivery was received by the Census Bureau, hardcopy materials were submitted via traceable mail. The hardcopy materials included answers to the TCFU questionnaire, written consent forms, audio tapes, and other PII-bearing materials.
- Use industry recognized software to organize data.[5] Because of the large amount of information produced, the electronic files were entered into a software known for organizing qualitative data. This way, data coding could be expedited and managed centrally.

3. Results

In this section, we measure quality achieved by how much the method-driven results met the requirements of the Census Bureau.

Representing All Required Living Situations

The team of recruiters contacted over 3,000 households during the active recruitment period, with a success rate of about 9%.[6] As shown in Table 3-1, at least one cognitive interview was conducted with a household proxy or a duplicate him or herself for each cell of the 27 living situations.[7] Recruiters and interviewers were able to meet or surpass the targets in all of the cells except for two Type 1 partial household match cells 1g (all duplicates are children) and 1i (all other combination of ages) and a Type 3 cell 3j (religious GQ) because of problems in the sample for recruitment. There were fewer than 25 households in the active sample for recruiters to contact or incomplete telephone numbers were listed.

[5] The software was used for the 226 cognitive interview summaries only. For the 50 ethnographic interviews, the lead researcher decided to not use the same software.

[6] The 9% success rate is calculated using the number of cases successfully recruited and divided by the total number of cases initiated for recruitment. We did not exhaust attempts for all cases initiated for recruitment simply because the quota for a cell had been filled.

[7] The 27 living situations included 18 HU and 9 GQ types that were organized in three main groups: (1) Type 1 ("phone match" cases): a person was counted in two different HUs and gave the same phone number on each census questionnaire; (2) Type 2 ("nonphone match" cases): a person was counted in two different housing units, and gave different phone numbers on each census questionnaire; and (3) Type 3 ("GQ" cases): a person was counted in a GQ as well as in a HU. Recruiters only contacted the HU. Type 1 and Type 2 cases were also identified as "whole household match" when every person listed on one census questionnaire is also listed on the second questionnaire, or as "partial household match", when at least one person on the questionnaire does not match to a person on the other questionnaire. Subgroups included adults, seniors, children, or a combination of ages. There may be one or more duplicates in a household pair.

Table 3-1. Final Status of Respondent Recruitment

Cases	Dupli-cate	Proxy	Total
Type 1 Cases – Phone Match	**37**	**5**	**42**
1a: Whole household (HH) match: All duplicates are adults	9	0	9
1b: Whole HH match: All duplicates are seniors	6	0	6
1c: Whole HH match: The matches include children	5	0	5
1d: Whole HH match: Any other combination of ages	4	0	4
1e: Partial HH match: All duplicates are adults	6	1	7
1f: Partial HH match: All duplicates are seniors	2	3	5
1g: Partial HH match: All duplicates are children	0	1	1
1h: Partial HH match: Other age combinations with children	4	0	4
1i: Partial household match: All other combination of ages	1	0	1
Type 2 Cases – Nonphone Match	**71**	**54**	**125**
2a:Whole HH match: All duplicates are adults	17	0	17
2b:Whole HH match: All duplicates are seniors	6	1	7
2c: Whole HH match: The matches include children	10	0	10
2d: Whole HH match: Any other combination of ages	4	0	4
2e: Partial HH match: All duplicates are adults	16	10	26
2f: Partial HH match: All duplicates are seniors	10	14	24
2g: Partial HH match: All duplicates are children	0	19	19
2h:Partial HH match: Other age combinations with children	5	8	13
2j: Partial HH match: Any other combination of ages	3	2	5
Type 3 Cases – Group Quarters (GQ)	**22**	**37**	**59**
3a: I: Military GQ	1	3	4
3b: J: College GQ	1	11	12
3c: K: Jail GQ	2	9	11
3d: L: Juvenile GQ	0	3	3
3e: N: Group Home GQ	5	2	7
3f: M: Nursing Home GQ	5	3	8
3g: O: Homeless GQ	5	3	8
3h: P: Workers GQ	2	3	5
3j: Q: Religious GQ	1	0	1
Total	**130**	**96**	**226**

Quality of the TCFU Information Collected

Of all 226 cognitive interviews, a little over half (120 or 53%) of the TCFU information collected included a complete address and dates that could be used to confirm an exact match. That is, they were confirmed duplicates in the 2010 Census. We proactively sent the remaining 106 or 47% of the interviews for Census Bureau data review. In these interviews, either respondents did not confirm or mention the other address that was listed in the recruitment file, or they reported an incomplete address or dates in the TCFU. For these 106 cases that did not produce an exact match, the Census Bureau research team was able to determine their duplication situation based on the information collected in the interview. As shown in Table 3-2, the review determined that about one third or 35 cases had a verified dup address and were confirmed duplicates.

A duplication pair was labeled a "false match" or "suspected false match" when Census Bureau determined during the data review that the two computer-matched people in the recruitment file were likely different people. In addition to using TCFU information

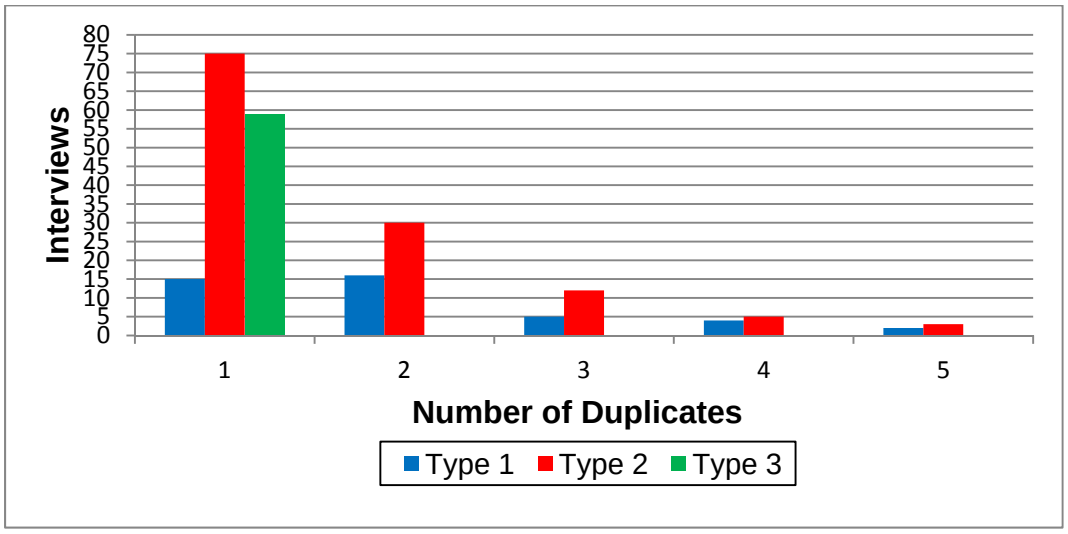
collected during the interviews, the recruiters also attempted to identify them during the recruitment process (before interviews took place) using the parameters established by the management team. In Round 2, of the 34 cases submitted for Census review, 27 were determined by Census Bureau to be false matches and eliminated from being interviewed.

Table 3-2. Case Outcome of Census Bureau Data Review[8]

Outcome	Count and Percentage
Verified duplicates	35 (33.0%)
False match and suspected false match	32 (30.2%)
HU mix-up	1 (0.9%)
Not revealed duplication	38 (35.8%)
Total	**106 (100%)**

Another TCFU quality indicator is whether the completed interviews included diversifying the characteristics of the duplicates, which was a priority indicated by the Census Bureau research team. We were able to recruit and interview households with multiple duplicates and households with duplicates of various ages. Figure 3-3 shows that completed interviews included households with multiple duplicates, but the majority were not. This was mainly because of restrictions in the available sample. For example, during Round 2 where the majority of the interviews were completed, there were only five households with six duplicates in the recruitment sample for the recruiters to attempt. However, the final distribution of households with multiple duplicates fared better than its distribution within the sample available to recruiters, meaning the recruitment was successful in representing multiple duplicate households.

Figure 3-3. Number of duplicates in completed TCFU interviews[9]



[8] The first two categories in Table 3-2, "verified duplicates" and "false match and suspected false match" are explained in the text. The category "HU mix-up" is different from a false match in that people are listed at the wrong address because of mail delivery or interviewer error in matching the form to the correct address to conduct the interview. The category "not revealed duplication" is used when there was not enough information from the interview to know whether respondents actually were connected to the other side of address or if it was a false match.

[9] Type 3 cases were always limited to one duplicate.

Figure 3-4 shows the age distribution for all of the duplicates covered in the completed interviews. Three age categories had the highest number of dups represented: children (aged 0-4), young adults (aged 20-24), and elderly (aged 75-79). In addition, during Round 1, the majority of duplicate addresses for a case were 12 miles or less apart. In Round 2, only ten percent of the complete interviews were 12 miles or less apart—the majority of cases had much greater distances. The improvement from Round 1 recruitment can be attributed to implementing the preventive action of making more strategic assignments by carefully analyzing the list sample. Table 3-5 shows that our recruitment was able to account for households whose duplicate address was both at a close or greater distance.

Figure 3-4. Duplicates age distribution in completed TCFU interviews

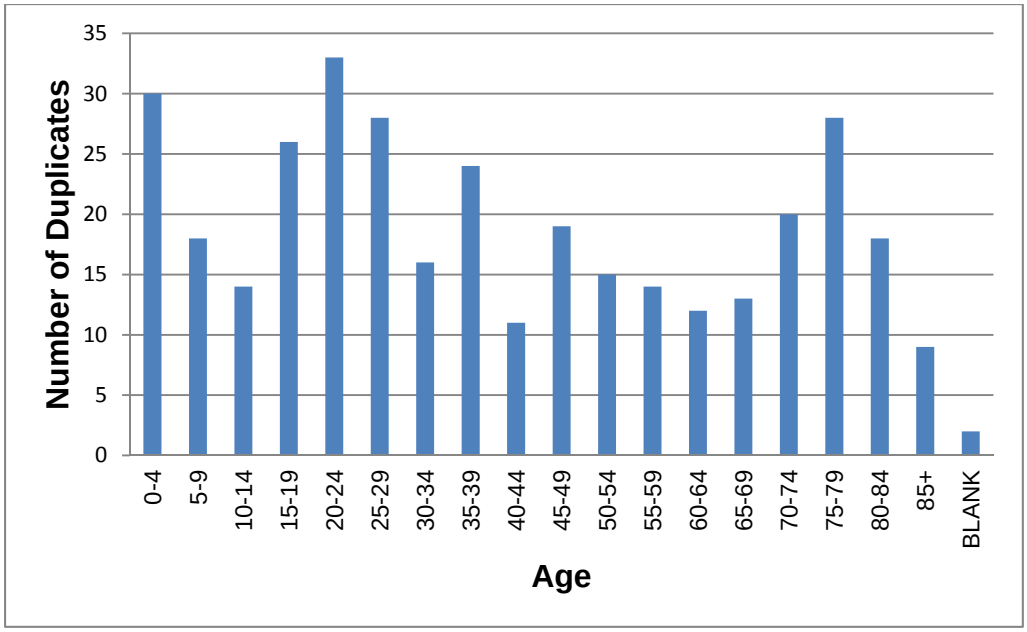


Table 3-5. Simple Statistics for Mileage between Addresses

Measures	Type 1 (n=42)	Type 2 (n=125)	Type 3 (n=59)
Mean (SD)	191.8 (419.3)	113.4 (316.0)	84.3 (181.0)
Variance	175,779	99,862	32,778
Mode / Range	6 / 1,504	0 / 2,460	0 / 966
Quantiles			
100% Max	1,505	2,460	966
95%	1,186	544	603
90%	902	181	181
75% Q3	52	94.42	78.26
50% Median	10	24	11.11
25% Q1	5.33	8	5
10%	2.26	3.15	2
0% Min	1	0	0

Additional duplication situations were represented in the cognitive interviews and in the ethnographic interviews. We interviewed and reported about complex household compositions beyond the typical nuclear family duplications, such as intergeneration relationships. We explored distant kinship and non-kinship duplication by including more "interesting" cases, such as where a duplicate was listed as a cousin on one census form and an uncle in the duplicate households on the other census form. We were also able to examine less studied living arrangements and "interesting" duplication situations. For example, we were able capture Snowbirds by recruiting for households whose duplicate address appeared in states traditionally associated with Snowbirds such as Florida, Arizona, and Southern California.

Recruiter and Interviewer Productivity

The main reason to assemble a large team of recruiters and interviewers was to recruit and complete almost 300 interviews quickly. The goal was achieved. To understand how much the recruiters and interviewers contributed to that goal, we ranked the recruiters and interviewers by their productivity, i.e. number of cases recruited or completed. We also compared their strengths and weaknesses to measure the degree of the individual productivity. Even though many factors contribute to productivity, this simple analysis showed interesting patterns.

Table 3-6 shows the strengths and weaknesses of the three most productive recruiters who recruited eighty percent of the interviews. Recruiter A was responsible for recruiting the highest number of completed interviews because of a flexible schedule and this recruiter also had superior performance recruiting hard-to-reach populations and superior cold-calling techniques. This recruiter was also able to work independently. However, navigating with ease in the computer sample, recognizing sample anomalies, and following administrative procedures were not strengths. As a result, Recruiter A received several remedial trainings. In comparison, Recruiter B showed strengths in all of the evaluation dimensions, but did not perform as outstanding of a job as Recruiter A on recruiting hard-to-reach populations. Recruiter C performed at the same superior level with Recruiter A in terms of recruiting hard-to-reach population and did not require remedial trainings. Because she did not have a flexible schedule that allowed her to work more hours to recruit respondents, her productivity ranked third.

Table 3-6. Strengths and weaknesses of top performing recruiters

Recruiter	Flexible schedule	Superior performance recruiting hard-to-reach population	Superior cold-calling techniques	Worked independently	Navigated with ease in the computer sample	Able to recognize sample anomalies	Did not need remedial trainings	Adhered to all administrative procedures
A	•	•	•	•				
B	•		•	•	•	•	•	•
C		•		•	•		•	

We repeated the analysis for the interviewers. As expected during the interviewing process, there were cancellations and "no shows", yet some interviewers were able to complete more assigned interviews than the others. As shown in Table 3-7, having a flexible schedule, being comfortable in the field, and consistently making appointment reminder calls to respondents were the three most common traits demonstrated by interviewers who were more productive.

Similarly with the recruiters, a small cadre of interviewers were responsible for the majority of the completed interviews - Twenty percent of the interviewers completed about eighty percent of the interviews. The most productive interviewer completed 60 or twenty-two percent of all interviews. The least productive interviewer completed three or only one percent of all interviews. While all 13 interviewers were quite experienced, five interviewers were at the expert level in terms of their knowledge about census coverage issues. Interestingly, they were not among the most productive group and completed a combined 40 interviews or less than ten percent of all interviews. In addition to senior-level stature, the one trait these expert interviewers shared was not being able to offer as flexible of a schedule as the other interviewers. All interviewers conformed to the requirements to complete interviews and reports. A few interviewers occasionally did not follow project procedures to collect (non-PII) case information.

Table 3-7. Ranking of shared strengths of the most productive interviewers

Ranking	Strengths of most productive interviewers
1	Flexible schedule
2	Comfortable in the field
3	Consistently made appointment reminder calls to their respondents
4	Excellent rapport with respondents
5	Comfortable at all locations and could handle unexpected situations
6	Extensive interview preparation

Prevention, Corrective Action, and Rework

In this section, we present several examples to provide an estimate of the impact on quality when prevention and corrective actions were taken to minimize rework. These were actions to prevent risks from being realized and to correct known issues. For instance, none of our team members experienced problems identifying the accurate file to use when multiple versions of the same file had been exchanged. We were also able to easily navigate and recall the needed files both in the document repository system as well as in the analysis software. We credit this to the file handling protocol that was set up at the beginning of the project. We also proactively identified potential false matches during the recruitment phase for Census Bureau to review so that they could be excluded from the interview process. This likely contributed to the diversity of duplication situations that were included in the interviews.

In addition, we made continuous improvement after evaluating the current state of the project against the project management plan, particularly when we identified issues that needed to be resolved. First, at the beginning of Round 2, we noticed a higher level of no shows, cancellations, and rescheduling with several interviewers. While respondent attrition is expected in any study, controlling it means that we did not have to invest more resources in the same case and could direct it elsewhere. We conducted a root cause analysis to understand the factors that could be controlled and learned that these interviewers did not consistently make reminder calls to the respondents about their

appointments. As a result, some cases were lost completely and had to be replaced. Rework like this could affect quality because it stagnated interview progress and the permutations of recruitment cell fulfilment had to be reconsidered. Our corrective actions included retraining interviewers, reassigning the responsibility in some instances, or using pre-scheduled email alerts to remind the interviewers of this responsibility. In addition to using the advance letters as a way to assure respondents of the study's legitimacy, during Round 2, telephone numbers to Census Bureau regional offices were made available to respondents who wanted to verify the legitimacy of the study.

Second, during Round 1, communications became an issue between recruiters and interviewers and several corrective and preventive actions were taken. Because the list sample respondents might not live close to RTI's cognitive laboratory or the RSS facility, recruiters usually scheduled appointments in the field at a private location preferred by the respondents. This convenienced the respondents, not the interviewers who were more familiar with the cognitive laboratory setting. A few interviewers were only comfortable traveling to some locations, distance (often 50 miles one way), or during certain time of the day. A few interviewers were particularly de-motivated and several interviews were lost as a result, causing rework. We mitigated this issue and prevented future risks by sending travelers, rematching recruiters who were more sensitive to interviewer preference, and retraining interviewers on keeping the Web accessible calendar up-to-date with their availability. We would not have been able to complete all 276 interviews by the project end date without swiftly implementing these changes.

Furthermore, Census Bureau authorized the RTI manager to rebaseline the scope when there were major discoveries that threatened the project end date and budget. After learning that many respondents were duplicated because of a move, in Round 2, recruiters identified those respondents during the recruitment process and excluded them from being interviewed. As a result, we were able to include more diverse duplication situations. Using our analysis of the permutations of possible duplication situations, Census Bureau was aware which recruitment cells could realistically be filled and which might not because of a small duplicated persons sample to be attempted. Changes were approved to redirect the resources away from those cells to other priority cells designated by the Census Bureau research team. Similarly, although the preference was to use a scripted approach to recruitment, when our recruiters reported that this approach was less successful, changes were made. Census Bureau allowed recruiters greater flexibility as the recruitment period progressed and they became very familiar with the screening criteria. Census Bureau also worked with the recruitment co-leads to develop a recruitment guide that provided the flexibility of conversational approach, while listing the mandatory information that had to be given to the respondents. We believe all of these changes contributed to quality.

4. Discussion and Conclusion

Overall, we believe our methods led to a quality product. Census Bureau verified and accepted all deliverables. More importantly, the project produced what it was created to produce for the Census Bureau - providing a qualitative picture of the issue of duplication. We were able to complete 276 interviews across 27 living situations and also diversified the interview findings by including complex and less studied duplication situations. There was no breach of confidentiality, privacy, and no negative public reactions.

In terms of the quality of the TCFU information collected, the level of details in the interview summary reports were appropriate and allowed the Census Bureau to make informed decisions about the duplication situation even when respondents did not provide complete dates or address in the interview. Census Bureau requested clarifications or feedback from the interviewers for less than ten percent of the interview summary reports, but did not require any interview data to be dropped or replaced. We also proactively identified data issues that might affect quality of the TCFU interviews. For example, by alerting the Census Bureau of potential false matches at the onset of the recruitment and excluding them from the interview process, recruiters reduced the total number of false matches, and thus, uninformative cases in the study. Because suspected duplicates were identified by a computer matching program, false matches determined in the recruitment or interview stages may help the Census Bureau refine the matching algorithm. The need for reviews also informed Census Bureau that a clerical review operation is necessary for the TCFU interviews being planned for the 2020 Census.

The continuous improvement was key to control quality. We monitored the permutations of desired characteristics in the duplicated persons list and selected recruitment samples that were more likely to result in duplication situations that meet Census Bureau's research agenda. Despite about 600 electronic files being exchanged, version control problems were prevented by setting up a file handling protocol. When the number of no-shows and interview rescheduling increased, we identified the root causes and monitored the factors that were within our control, such as making appointment reminder calls. In addition, when concern about location, distance, and timing in the field became a source of tension between interviewers and recruiters and threatened interviewing progress, we addressed the issue right away. We were able to minimize the negative impacts of lost interviews and rework this way for both the recruiters and interviewers.

One source of complexity in managing quality is human influence. Although all recruiters and interviewers were highly-skilled and experienced, as required by the Census Bureau, their productivity varied. Among the three most productive recruiters, one recruiter was less successful in navigating in the recruitment sample and bringing anomalous situations to the recruitment co-leads (e.g. false matches). However, this recruiter was the most productive and recruited almost a fourth of all completed interviews. He was also successful with recruiting among the senior population who are usually difficult to reach because of trust issues. Future qualitative research studies should use experienced recruiters but tailor case assignments according to recruiters' strengths.

Among the 13 interviewers, 20 percent of the interviewers completed about 80 percent of the interviews. The gap between the number of interviews completed was large (57) by the most and least productive interviewers. Although we realize that many factors contribute to an interviewer's interview completion rate, in general, schedule flexibility, feeling comfortable in the field, and making appointment reminder calls to the respondents seem to have contributed to productivity. Five of the 13 interviewers are senior researchers who are experts at census coverage issues and were indispensable to the quality of the TCFU information collected. However, they also happened to be among the group that completed fewer interviews – these expert interviewers' availability was quite limited because they were usually engaged in several research activities that demanded their skills. We recommend future qualitative research studies that require a large team of interviewers to include both senior-level interviewers and other experienced

interviewers. This way, a large number of interviews can be completed on schedule and interviewers can coach each other to improve the quality of the data like we did.

Furthermore, the Census Bureau research team was a critical contributor to the quality of this study. They remained interested throughout the project, made fact-based decisions, and set realistic expectations based on lessons learned as reported by RTI-RSS recruiters and interviewers. When significant discoveries were made, Census Bureau worked with the RTI project manager to rebaseline the project management plan. Examples of approved changes that contributed to quality were: diversifying duplication situations to be interviewed by screening respondents about having permanently moved; redirecting resources from recruitment cells that had insufficient duplicated persons sample to recruit from; and allowing recruiters to tailor recruitment introductions.

Although this paper discusses managing quality on a large-scale qualitative research effort, the methods and measures are also applicable to small-scale efforts. To assure quality, all studies must monitor and continuously improve the process of respondent recruitment, interviewing, and communications so that the product meets the requirements of the study. As federal budget uncertainly continues, future large-scale qualitative research studies should quantify the cost of quality to demonstrate that quality does not have to be compromised while balancing the constraints of cost, time, and scope. We believe that the corrective and preventive actions on this study controlled risks and issues from jeopardizing quality and minimized rework which would have multiplied the cost.

Acknowledgements

The authors thank Jennifer Hunter Childs, Sarah Heimel, and Ryan King who designed the TCFU and this qualitative research study at the Census Bureau. We would also like to acknowledge Mandy Sha of RTI International who managed the RTI-RSS team, the recruitment co-leads Katherine Kenward (RSS) and Julie Feldman (RTI), lead researchers Emilia Peytcheva (RTI) and Eleanor Gerber (RSS), and Alisú Schoua-Glusberg of RSS. In addition, we acknowledge the team of interviewers. At RTI: Tiffany King, Sarah Cook, Tim Flanigan, Ariana Napier, Ginny Wake, Emilia Peytcheva, Mandy Sha. At RSS: Woody Carter, Betsy Strick, Eleanor Gerber, Alisú Schoua-Glusberg, Pat Marnell, Dánae Corado. The recruiters were Jon O'Connor and Mildred Hawkins of Headway/RTI, and RSS recruiters Lisa Andrews, Liliana Aguayo, Dánae Corado, Pat Marnell, and Daniela Glusberg.

References

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. (2011, September). *Qualitative interviewing with suspected duplicates and cognitive testing of the Targeted Coverage Follow-up (TCFU) interview*. Prepared for U.S. Census Bureau.

Project Management Institute (PMI). (2008). A guide to the Project Management Body of Knowledge (PMBOK guide). Philadelphia, PA: Author.

Willis, G. B. (2005). *Cognitive interviewing: A Tool for Improving Questionnaire Design.* Thousand Oaks, CA: Sage.