Using Computer-Assisted Recorded Interviewing to Enhance Field Monitoring and Improve Data Quality[1]

Holly Fee, T. Andy Welton, Matthew Marlay, and Jason Fields
U.S. Census Bureau
4600 Silver Hill Road, Suitland, MD 20746

Proceedings of the 2015 Federal Committee on Statistical Methodology (FCSM) Research Conference

INTRODUCTION

The Survey of Income and Program Participation (SIPP) is a household-based survey designed as a continuous series of national panels. Each panel features a nationally representative sample interviewed over a multi-year period lasting approximately four years. The survey's mission is to provide a nationally representative sample for evaluating: (1) annual and sub-annual income dynamics, (2) movements into and out of government transfer programs, (3) family and social context of individuals and households, and (4) interactions among these items.

In 2014, a newly re-engineered SIPP entered production. The aim of the new design was, in part, to: (1) help reduce costs, (2) improve data processing, (3) modernize the instrument, and (4) expand and improve the use of administrative records. The new design includes: (1) an annual interview (formerly administered three times a year), (2) use of Event History Calendar (EHC) methods, (3) inclusion of key topical module contents within each wave, and (4) use of Computer-Assisted Recorded Interviewing (CARI).

CARI is a technology used to monitor interviews by recording the interactions between the field representative (FR) and respondent. The primary goal of CARI is to ensure the accuracy and quality of data collected. Additionally, CARI allows field supervisors (FSs) to assess FR performance and identify areas for improvement. Finally, CARI allows survey questionnaire designers and subject matter experts to evaluate question wording, and where applicable, identify areas in need of improvement.

CARI was first implemented in the National Survey of Child and Adolescent Well-Being (NSCAW) by Research Triangle International (RTI) in 1999 (Thissen, Fisher, Barber, and Sattaluri, 2008). Now nearly two decades later, many major survey organizations have at least one survey featuring CARI in their quality assurance program, including the University of Michigan, the National Opinion Research Center, and Westat. Each organization employs a CARI consent question similar to the Census Bureau, with CARI consent rates varying from 81-

[1] This work is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views or opinions expressed in the paper are the authors' own and do not necessarily reflect the views or opinions of the U.S. Census Bureau.

96% (Thissen, Park, and Nguyen, 2013; Arceneaux, 2007; Sirkis, 2013; Hicks, Edwards, Tourangeau, McBride, Jarris-Kojetin, & Moss, 2010).

The Census Bureau's current QA system uses telephone and in-person reinterviewing. Reinterview FRs contact respondents for reinterview within two weeks of the original interview. Reinterview FRs first attempt to contact respondents via telephone. Should a reinterview FR fail to reach the respondent by phone, the reinterview FR will visit the sampled address and attempt an in-person reinterview. For reinterview cases requiring an in-person visit or multiple phone calls, the process can be labor intensive and costly. A potential benefit of CARI is drastically reduced QA costs while still providing information equivalent to current standards. RTI's CARI QA program asserts that coders can determine interview authenticity within three minutes, generating tremendous cost savings compared to traditional reinterview (Thissen, Fisher, Barber, and Sattaluri, 2008). The Census Bureau has invested significant resources into CARI development with the objective to use CARI audio recordings as a complementary tool for validating interview authenticity and detecting possible interviewer falsification.

Only a small, but critical, subsample of questions within the instrument are selected for recording via CARI. Question selection was based on a number of considerations, but largely related to two factors: (1) placement within the questionnaire, and (2) suspect for administration error/ misinterpretation. Specifically, 12 total questions were selected – five from the front of the interview, six from the middle, and one from the end. The FR and respondent are unaware of when CARI is and is not recording the interview or what questions are and are not recorded.

At the beginning of the interview, FRs ask the household respondent for their consent to be recorded, *"This interview may be recorded for quality control purposes. Is that O.K.?"* The FR must obtain approval from the respondent before recording the interview. FRs are also prompted to ask the CARI consent question for each subsequent interview in the household. Only respondents who are 15 years old and older may complete a self-reported interview. Proxy interviews are collected for respondents who are younger than 15 years old and adults who are unavaible for a self-reported interview. Answers to CARI for proxy interviews are copied from respondents who completed self-reported interviews and who are completing a proxy interview for another household member.

Quality Assurance (QA) Component

The goal of any quality assurance program is to produce an error-free finished product. In the same manner, CARI helps to ensure that data collected contain as few errors as possible. Non-sampling related survey errors may arise from numerous sources but usually fall into two main categories: authenticity problems and errors introduced by the interviewer.

Authenticity errors occur when FRs do not collect data directly from a respondent. This may occur quite innocuously, such as when an FR assumes the answer and does not ask the respondent the question. In these instances, FRs who interview for SIPP are trained to never assume an answer and instead verify the information with the respondent, such as "you said earlier that you had health insurance through Medicaid, is that correct?" In extreme cases, authenticity problems occur when FRs skip a section of the questionnaire or even fabricate the

entire interview. The Census Bureau has an established method of investigating these cases and handling FRs who have been suspected of falsifying data.

Interviewer errors occur because of differences in the presentation of the interview questions. Specifically, not all respondents are receiving the survey in the same fashion. This can occur for a host of reasons, for example emotional loading (such as through tone of voice), leading the respondent, or a failure to follow administrative protocols. For questions of a sensitive nature, even the mere presence of an interviewer may bias responses.

During Wave 2 of the 2014 SIPP Panel, a CARI-QA pilot test was initiated using professional survey coders from the Census Bureau's National Processing Center (NPC) in Jeffersonville, IN. Originally, the Census Bureau planned to initiate a coding program sampling 10% of all cases (i.e., the proportion of cases selected for traditional reinterview). This was the sample size deemed necessary to establish CARI as a viable alternative to reinterview as a QA method. However, due to budgetary constraints in Wave 2 of the 2014 SIPP Panel, CARI QA sample size was restricted to a number only sufficient for a pilot test instead of 10% of the sample. The goal of the pilot test was threefold: (1) test the ability of NPC to manage the flow of cases into the coding process, (2) code at least one case for each FR, and (3) test the reliability of coding results.

PRESENT STUDY

The primary aims of the present study are to (1) assess how the CARI consent rate differed along a variety of dimensions, (2) examine how much of the variance in the odds of the CARI cooperation rate is attributable to respondents, Field Representatives (FRs), Field Supervisors (FSs), Supervisory Statisticians Field (SSFs), and regional offices (ROs), and (3) review and discuss findings from our QA CARI Component. Because of the introduction of CARI in the 2014 SIPP, we have a unique opportunity to address these aims, which were previously impossible to examine with the old SIPP survey design.

DATA AND METHODS

CARI

Our analysis draws on data from Wave 1 of the 2014 SIPP Panel. The Wave 1 sample included approximately 53,000 households, nearly 30,000 of which we successfully interviewed. Approximately, 66% of sampled households had at least one member who agreed to be recorded (Fee, Welton, and Marlay, 2015).

We begin by estimating the means (or proportions) of all variables used in the analyses by whether a household consented to CARI or not, and testing for significant differences between households that consented to CARI and those that did not. Next, we estimate a five-level organizational model using multilevel modeling techniques. Our outcome of interest is measured at the household level (Level 1). Households are nested within FRs (Level 2). FRs are nested within Field Supervisors (Level 3). FSs are nested within SSFs (Level 4). SSFs are nested within ROs (Level 5). For Wave 1 of the 2014 SIPP panel, there were 28,227 interviewed households

(who reported valid level measures, such as FS and SSF) nested within 1,334 FRs, nested within 508 FSs, nested within 48 SSFs, nested in six regional offices. Figure 1 depicts the data structure of our models.

[Figure 1 about here]

Multilevel modeling is appropriate because it accounts for the clustering of respondents by FR, FRs within FSs, FSs within SSFs, and SSFs within ROs. Ignoring the more detailed levels of analysis can lead to incorrect conclusions by affecting estimated variances and the power to detect covariate effects (Ene, Leighton, Blue, and Bell, 2014). We expect that each level of analysis will significantly contribute to CARI consent. We hypothesize that households and FRs will account for the largest proportion of the explained variance in CARI consent, given households provide the answers to questions and FRs have the responsibility of keying the answers.

Dependent Variables

The dependent variable is *whether the household consented to CARI or not*, dummy coded (1 = Yes, 0 = No). Although CARI is asked of each adult household member (i.e., age ≥ 15), SIPP samples households in Wave 1; therefore, we use the household respondent's CARI response.[2] The CARI question reads, "This interview may be recorded for quality control purposes. Is this OK?"

Control Variables

Respondent-level characteristics. Our analysis controls for a number of characteristics of the respondent. Again, note that because our analysis focuses on whether the household consented to be recorded, the respondent-level characteristics are those of the household reference person. These variables include:

- *Race/ethnicity*: White (reference group), Black, Asian, other race, Mexican (American), and other Hispanic.
- *Sex*: 1 = Female, 0 = Male.
- *Age*: a continuous variable coded in years.
- *Marital status*: married (reference group), widowed, divorced/separated, and never married.
- *Household income*: 1 = Below 200% of poverty threshold, 0 = Above 200% of poverty threshold.
- *Employment status*: 1 = Currently employed, 0 = Not currently employed.
- *Education*: less than high school degree, high school degree (reference group), some college, Bachelor's degree, and post-graduate degree.
- *Household size*: a continuous variable coded in number of persons in the household.
- I*nterview was conducted in a language other than English*:1 = Yes, 0 = No.
- *Foreign-born status*: 1 = Yes, 0 = No.
- *Incentive amount*: $0 (reference group), $20, and $40.
- *Housing tenure*: owns home (reference group), rents home, and lives rent-free in home.

[2] The household respondent is the first eligible adult household member. This person provides the household roster and demographic information during the interview.

- Whether any household member *ever served in the Armed Forces*: 1 = Yes, 0 = No.
- *Case difficulty*:
    - *Number of contact attempts*: a continuous variable that accounts for the total number of telephone and personal visit contact attempts made to each household.
    - *Whether another FR was ever assigned the same household*: 1= Yes, 0 = No.

FR-level characteristics. In addition to characteristics of the respondent, our analysis also contains independent variables related to the field representative, including:

- *U.S. Census Bureau experience*: a continuous variable coded in years
- *Certification test score*: a continuous variable coded as a percent (0% - 100%). The certification test was administered to the FRs at the end of their weeklong training session. It quizzes the FRs along a number of dimensions designed to test their knowledge of three areas – SIPP content, instrument navigation, and field procedures. Higher scores reflect greater knowledge of SIPP and the instrument.

Regional office. All states are included in one of six regional offices. The six regional offices include: New York, Philadelphia, Chicago, Atlanta, Denver, and Los Angeles.

RESULTS

CARI

Table 1 shows the means or proportions for all variables used in the analyses. Our results are consistent with prior research (Fee et al., 2015) with roughly 66% of households consenting to be recorded. Among households that did not consent to CARI, roughly 19% were headed by someone who is foreign-born and nearly 16% were headed by someone of Hispanic descent, compared to roughly 10% and nearly 10% for households that did consent to CARI. Households that did and did not consent to CARI also differed on a variety of other demographic characteristics – those that consented were significantly more likely than those that did not to own their home and to have served in the Armed Forces. However, households that did not consent to CARI were more likely to conduct the interview in a language other than English and be currently employed compared to households that consented. In addition, households that did consent to CARI were less difficult compared to households that did not consent. Households that consented were less likely to switch FRs and were contacted less frequently. There are also significant differences for FR-level characteristics between households that did not consent to CARI and those that did. Households that consented were more likely to have FRs who scored higher on the certification test and have FRs who had nearly a year less U.S. Census Bureau experience. Households in all regional offices, except for Philadelphia and Atlanta, significantly differed in their CARI consent rate. For instance, 10% of households did not consent to CARI and 7% did consent to CARI were located in the New York regional office. Contrary to the pattern seen in the New York regional office, a greater proportion of households in the Chicago regional office consented to CARI. Approximately 23% of households that consented to CARI and 15% that did not consent to CARI were in the Chicago regional office.

[Table 1 about here]

Table 2 shows the results from multilevel models. Two components largely contribute to the explained variance in the CARI consent rate. Households and other unknown factors explain 45.69% and FRs explain 44.71% of the variance in the CARI consent rate. However, both FSs and SSFs significantly contribute to the explained variance in the CARI consent rate; FSs explain 4.82% and SSFs explain 4.16% of the variance. Regional office explains very little (0.589%) of the explained variance in the CARI consent rate. The variation in the CARI consent rate between FS areas decreases when FR-level traits are controlled (Model 2). More precisely, nearly 18% of the explainable variation in FS mean CARI consent rates is explained by FR-level traits. Model 3 controls for household-level traits and explains a smaller percentage (5%) of the explainable variation in FS mean CARI consent rates. However, household-level traits explain a sizable percentage (25%) of the variation in SSF mean CARI consent rates. On the contrary, controlling for FR- and household-level traits, the variation among FR and Regional office mean CARI consent rates still remains to be explained.

[Table 2 about here]

QA Component

All data compiled as part of the SIPP CARI-QA pilot test were recorded during Wave 2 interviews. Related to the pilot test's first goal, testing the ability of NPC to manage the flow of cases into the CARI coding system, NPC was successfully able to code 2,034 cases in a 2-month time span. Regarding the pilot test's second goal, coding at least one case for each interviewer, of the 2,034 coded cases there were a total of 1,096 interviewers rated. In Wave 2 of 2014 SIPP panel, a total of 1,209 interviewers conducted a partial or complete interview. Thus, the coding pilot test fell somewhat short of its second goal. However, the interviewers without coding data completed significantly less cases on average. The 1,209 interviewers had an average interview count of 18.05. The 113 interviewers without coding information had an average interview count of 7.01.

Each question selected for the QA program was rated across seven categories with a unique series of predefined codes for each category. The seven categories were (1) Authenticity, (2) Recording Problems, (3) Reading, (4) Interview Administration, (5) Probing, (6) Conduct, (7) Other (miscellaneous).

In the CARI-QA program, the 1,976 coded cases with CARI data had a total of 3,099 coded CARI questions. As for the pilot test's third goal, to test the reliability of coding results, an IRR analysis found an overall agreement rate of 43.04% between CARI-QA coders (results not shown). Meaning in 43.04% of all questions double-coded, the coders applied the exact same scores. This indicates only a fair degree of agreement among coders and does little to support the reliability of the CARI-QA results. However, due to the complex coding options, it is likely two coders could assign different results to an audio clip despite a similar overall view of question delivery. For example, within the Recording Problems category, coders had nine answer choices and could choose up to three. Of particular note for this study, on all CARI consent questions, NPC coders had an agreement rate of 44.81% (results not shown). Considering that there was only a fair amount of agreement among coders was not surprising as it is entirely reasonable for one coder to code an audio clip as "audible, but question not heard" and another as "unintelligible."

CONCLUSION

CARI

Consistent with prior research (Fee et al., 2015), 66% of households agreed to be recorded during his/her SIPP interview. Households that did and did not consent to CARI differed along a variety of dimensions. For instance, among interviews that were conducted in a language other than English, 3% of households consented to CARI, whereas 8% of households did not consent to CARI. Households that consented to CARI also significantly differed by foreign-born status; nearly 19% of household respondents who were foreign-born did not consent to CARI and roughly 10% did consent to CARI. Additionally, CARI consent rates differed by FR-level traits. Households that did consent to CARI were less difficult to interview; they were less likely to change FRs and completed an interview with fewer contact attempts.

Multilevel modeling results revealed that all levels, except for regional office, significantly contributed to the explained variance in CARI consent rate. Consistent with our hypothesis, both respondents and FRs explained more of the variance in the CARI consent rate than other levels. These two levels account for roughly 90% of the explainable variance in the CARI consent rate. These findings are in line with expectations, given households provide the answer to the CARI consent question and FRs ask and key the answer to CARI. Moreover, controlling for respondent- and FR-level characteristics helps to explain a sizable proportion of the explainable variation in FS and SSF mean CARI consent rates.

A benefit of CARI is the help improve data quality. Multilevel results show that regional offices do not significantly explain the variance in CARI consent. This indicates that all six regional offices are more similar than dissimilar in their CARI consent rates. However, CARI consent rates do significantly differ among SSFs, FSs, FRs, and respondents. The differences among SSFs, FSs, and FRs may be explained by the differences in the adherence to training and/or fielding procedures with some areas reporting higher CARI consent rates while others report low CARI consent rates.

QA Component

Although the IRR test did not confer robust reliability for the CARI coding pilot test, it is reasonable to conclude that within the coding data particular insight does exist. The low Kappa statistic in many coding categories is likely caused by an overabundance of answer options. For example, the Recording Problems category has nine answer options and an agreement rate of 19.35% (results not shown). Other categories utilized only five answer categories, but focused predominately on subjective issues open to interpretation (i.e., speed of diction or professional behavior). In future QA assessment of CARI coders, answer categories should be fewer or more standardized across categories.

Conventional wisdom dictates that people adhere more to established standards when being monitored. Likewise, it is an equally common belief that those who prefer not be observed must have an ulterior motive to evade monitoring. However, most people prefer to not have someone watching over their shoulder. Nonetheless, a successfully implemented CARI-QA program

requires the observance of *all* sampled cases. Any non-random assignment of CARI consent lowers the reliability of the coding results.

Similarly, successful survey administration requires that all respondents receive the questionnaire in a similar fashion. Otherwise, there is no guarantee that response variations are due to genuine respondent differences, and not the manner in which the question was delivered. Thus, CARI has the potential to serve as a revolutionary tool for ensuring data quality. During the SIPP Wave 2 QA pilot test, the QA processing system and its coding workforce showed an ability to skillfully manage the workload of an incredibly complex survey. With a streamlined coding interface and more established training system, the CARI- QA program could achieve meaningful improvements in the data collection process.

REFERENCES

Ene, M., Leighton, E. A., Blue, G. L., & Bell, B. A. (2014). Multilevel Models for Categorical Data Using SAS® PROC GLIMMIX: The Basics.

Fee, H., Welton, T.A., & Marlay, M.C. (2015). CARI That Weight: Obtaining Consent to Record SIPP Interviews. Paper presented at the Annual Federal Computer-Assisted Survey Information Collection (FedCASIC) Workshop, U.S. Census Bureau, Washington DC.

Hicks, W., Edwards, B., Tourangeau, K., McBride, B., Jarris-Kojetin, L., & Moss, A. (2010). Using CARI Tools to Understand Measurement Error. Public Opinion Quarterly, 985-1003.

Thissen, R., Fisher, C., Barber, L., & Sattaluri, S. (2008). Computer Audio-Recorded Interviewing (CARI): A Tool for Monitoring Field Interviewers and Improving Field Data Collection. Statistics Canada Symposium 2008 (pp. 1-9). Ottawa: Statistics Canada Symposium.

Thissen, R., Park, H., & Nguyen, M. (2013). Computer Audio Recording: A Practical Technology for Improving Survey Quality. Survey Practice, 1-7.