

ARTICLES

Exploring Reminder Calls Intended to Increase Interviewer Compliance with Data Collection Protocols

Amanda Nagle¹, Gina Walejko²

¹ Center for Adaptive Design, US Census Bureau, ² Decennial Statistical Studies Division, U.S Census Bureau

Keywords: interviewer compliance, reminder calls, autocalles, autodialer, capi

<https://doi.org/10.29115/SP-2018-0022>

Survey Practice

Vol. 11, Issue 2, 2018

Survey managers sometimes require field interviewers to adhere to novel data collection protocols, which they may not consistently implement for a variety of reasons. This research examines whether automated reminder telephone calls encourage interviewers to perform such new instructions. Our findings suggest that calls may be related to interviewer compliance. The calls are inexpensive compared to other forms of training and supervision. We suggest that investigators explore the use of reminder calls when asking interviewers to implement unfamiliar data collection techniques.

Introduction

Implementing innovative procedures in face-to-face surveys may call for interviewers to follow instructions on case management systems, and such instructions may require interviewers to transmit data from these case management systems to a central server on set schedules. For example, case management systems may communicate the prioritization of certain addresses to personal interviewers, which requires case status information be transmitted on a daily basis (Walejko and Wagner 2015). Unfortunately, these interviewers may not always adhere to such transmission instructions, causing information such as case prioritization to not be updated (Walejko and Wagner 2015). Thus, survey managers who want to implement new or modified techniques may be interested in cost-effective ways to get interviewers to follow transmission instructions or other novel protocols. This study examines how automated telephone reminder calls relate to interviewers' performance in transmitting data on a set schedule, a task essential to implementing an adaptive experimental survey design in the Survey of Income and Program Participation (SIPP). It is a retrospective observational study, so we cannot determine a causal relationship between reminder telephone calls to interviewers and transmission compliance, but the relationship between the two can be informative.

Background

Researchers have asked field interviewers to perform modified protocols in many data collections, including the 2013 and 2014 Census Tests (Walejko and Miller 2015; Walejko and Wagner 2015), the National Survey of Family Growth (Wagner et al. 2012), and SIPP (Walejko, Zotti, and Tolliver 2016). Research on the 2013 and 2014 Census Tests found that interviewers were non-compliant with

some experimental data collection protocols, including visiting specified households and transmitting data at certain times (Walejko and Wagner 2015). In a series of experiments in the National Survey of Family Growth, case management systems instructed interviewers to prioritize certain cases by making more frequent visits, and researchers found that interviewers made significantly more attempts to prioritized cases in only seven of sixteen experimental groups (Wagner et al. 2012).

Such issues illustrate the need for methods that increase interviewer compliance. Sending reminders such as autodialed telephone calls to interviewers may increase compliance with protocols. The National Agricultural Statistical Service tested sending autodialed calls to respondents during data collection instead of reminder postcards. They found that reminder calls increased response rates, but the effect was smaller than sending a reminder postcard (McCarthy 2007). Although research has not documented extending reminder calls to interviewers, it may be a cost-effective way to increase compliance with novel procedures.

The Survey of Income and Program Participation collects data and measures change for topics including economic well-being, family dynamics, assets, health insurance, food security, and — as its name suggests — participation in different programs, such as income-based food subsistence programs. During the third wave of SIPP in 2016, survey managers instructed interviewers to transmit performance data manually from their laptops to U.S. Census Bureau servers on a set schedule that was different from previous SIPP waves. An automated system sent all interviewers a call containing a recorded message reminding him or her to transmit after completing work on Wednesdays and before beginning work on Fridays. Prior to 2016, SIPP interviewers were asked to transmit each time they worked, and they were not reminded to transmit in any systematic way beyond their initial training. This previous transmission procedure was much simpler than the 2016 procedure described in detail below, and data on interviewer transmissions from 2015 were not available, so no analysis of compliance in 2015 could be performed.

This paper aims to answer four research questions: (1) how compliant were interviewers at following this new data transmission protocol, (2) were reminder calls related to higher compliance, (3) how does compliance vary between new and experienced interviewers, and (4) are reminder calls a cost effective way to encourage compliant behavior in interviewers?

Methods

SIPP is a longitudinal, face-to-face household survey conducted by the Census Bureau. During Wave III data collection, April to July 2016, researchers conducted an adaptive design experiment that instructed interviewers to work prioritized cases with the goal of increasing sample representativeness. In the adaptive design experiment, control interviewers' case management system showed all cases as having the same "priority" to interview, while treatment

interviewers' case management systems indicated that some cases were "high priority." Cases were prioritized using case status information, which required interviewers to transmit contact attempt data to a central server. Interviewers also needed to perform data transmissions to receive updated case prioritizations. Although the adaptive design was implemented as an experiment, the transmission protocol was not. Both control and treatment interviewers received the same transmission instructions and reminder telephone calls, so the automated reminder call effort was not an experimental manipulation.

Transmissions

Case priorities were altered throughout data collection in response to accruing data on completed and outstanding cases. To update the priorities displayed on their laptops, interviewers were required to transmit case status information to central servers on Wednesday evenings after their last contact attempt, and trainings and supervisors instructed each interviewer to perform these manual transmissions on this schedule. Using transmitted data, algorithms were employed to update case priorities on Thursdays, and the new priorities were delivered to every interviewer's laptop case management system when he or she transmitted before their first contact attempt on Fridays. In previous SIPP waves, interviewers were asked to transmit each time they worked, but they were not instructed to transmit at a specified time of day.

In order for an interviewer to transmit data to the central Census Bureau servers or receive an update from the servers, interviewers had to connect their laptop to the Internet, navigate to the correct screen, and select "transmit." This process could take several minutes and must have been done at home or at work. Hereafter, this process will be referred to as a transmission. Although not a particularly difficult task, a transmission required planning and intentional action.

Automated Reminder Calls

A computer system called PhoneTree¹ placed automated reminder calls to interviewers. When the telephone call was answered by a person or an answering machine, a prerecorded audio message played. Automated calls went to all interviewers with available telephone numbers each Wednesday and Friday with two exceptions, April 1st and May 6th, resulting in a total of 20 days of calls. Throughout this paper, Wednesdays and Fridays are called *transmission days*.

Interviewers

¹ PhoneTree is a system used by the Census Bureau to send automated telephone calls. PhoneTree attempted to call all interviewer telephone numbers between 4:00 and 4:30 pm Eastern Standard Time on both Wednesdays and Fridays. The caller ID showed a Census Bureau number.

Each week the automated call system updated interviewer telephone numbers from databases located at each Census Bureau regional office. Every newly assigned SIPP interviewer was called on the next transmission day, and the telephone numbers of interviewers who stopped working on the survey were removed. The total number of calls to SIPP interviewers per transmission day was approximately 1,600. Interviewers were considered “experienced” if they were hired before October 1, 2015, and “new” if they were hired on or after that date. October 1, 2015, is the day that hiring started for the 2016 wave of SIPP, so interviewers hired before October 1 most likely worked on a study prior to SIPP 2016. Throughout data collection, about 1,400 experienced interviewers and 300 new interviewers worked on cases. Experienced interviewers were aware that the transmission rules for this study were different from previous waves of SIPP and other surveys. Regardless of whether or not they were in the control or treatment, all SIPP interviewers were reminded to transmit on the new schedule.

Data

The research team used interviewer recorded contact history data to determine whether interviewers made compliant transmissions. These data included a timestamp for each recorded contact attempt and transmission. Using this information, the research team determined the first and last action — contact attempt or transmission — performed by every interviewer each day. An interviewer was marked compliant for a Wednesday date if their last action before Wednesday at 11:59 pm was a transmission, even if the action did not occur on Wednesday. An interviewer was marked compliant on a Friday date if their first action after that Friday at 12:00 am was a transmission, even if that action did not occur on that Friday. Accordingly, a binary compliance indicator was created for each interviewer for each transmission day.

Reminder telephone data included about 32,000 records, one for each interviewer for each of the 20 days called. Records included date, time, and outcome of the call. The research team grouped call outcomes into three categories: Message Received, Message Delivered, and No Message. The Message Received category indicated that a person who answered the telephone received at least some of the recorded message. The Message Delivered category contained messages that were delivered to a telephone number associated with an interviewer, but it was not known whether he or she listened to the message. The No Message category was for calls during which PhoneTree was unable to reach a person or leave a message. A complete list of call outcomes and categories can be found in Table 1.

Results and Discussion

First, the research team analyzed compliance and call outcomes separately. Descriptive statistics and estimates over time were calculated to understand interviewer behavior and patterns. Next, we combined and analyzed data about compliance and automated reminder call messages to

Table 1. Automated reminder telephone call outcomes.

Outcome	Description
Message received	Message received by a person
Answered by person	Message was delivered to a person
Hung up early	Call answered; person hung up before message finished
Message delivered	Message delivered to a telephone
OGM ^a too long	Outgoing message too long to be delivered in whole
Answered by machine	Message was delivered to an answering machine
No message	No message delivered
Call failed	Error detected during call
No connect	No signal detected after dialing
Busy after voice	Telephone company busy or off-hook service message
Max no answers	Line was busy or not answered for max attempts

^a outgoing message

determine if there was a relationship between the two. Finally, we examined cost data to understand the feasibility of automated reminder calls for future data collections.

Analysis of Automated Reminder Call Outcomes

Averaged across all transmit days, 39.9% of automated reminder calls had a Message Received outcome, 36.2% had a Message Delivered outcome, and 23.9% had an No Message outcome. The percent of message outcomes by day can be seen in Figure 1. Messages received decrease and messages delivered increase over time, perhaps because more interviewers let the reminder calls go to voicemail as data collection progressed.

Analysis of Transmission Compliance

Overall transmission compliance, or the total number of times interviewers were compliant divided by the total number of transmission days, was 87.7%. To put this statistic in context, we compared it to interviewer transmission compliance for the 2013 Census Test. The test used the same case management system and mostly experienced interviewers but asked interviewers to transmit twice daily, once before they started working and once after they stopped working. Calculated by the same method as overall transmission compliance in this study, overall transmission compliance in the 2013 Census Test was 71% (Walejko and Wagner, forthcoming).

A compliance percent was calculated for each day by dividing the number of compliant transmissions by the total number of interviewers who were given a compliance indicator for that date. Considering these daily averages, the average compliance across the eleven Wednesdays was 87.8%, and the average of the nine Fridays was 84.5%. Using daily averages to test for a difference

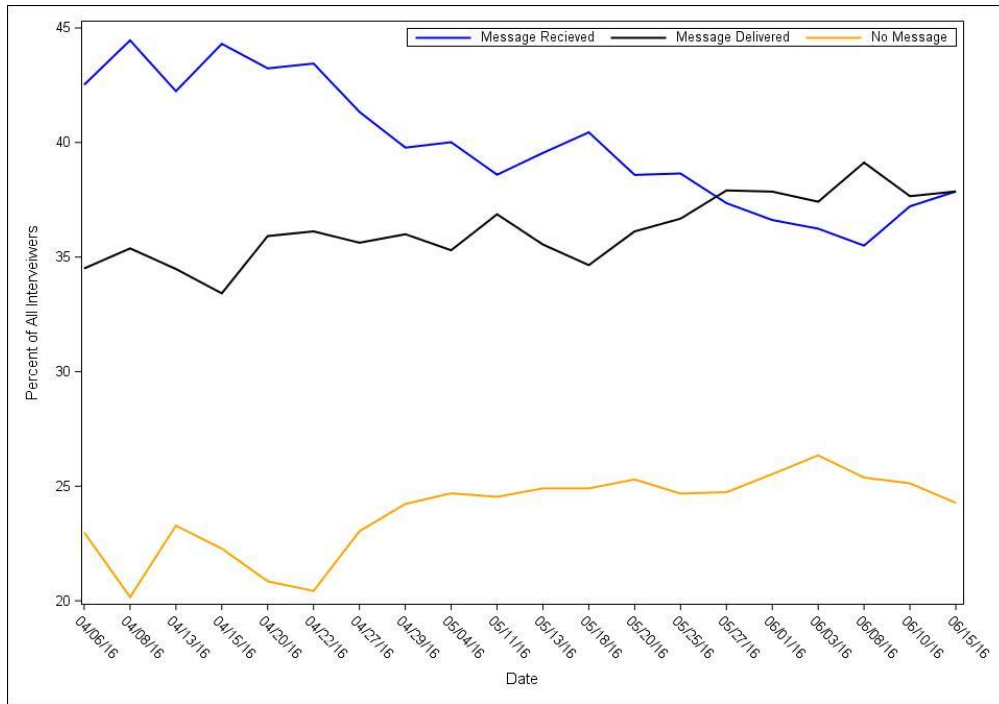


Figure 1.

Percent automated reminder calls by outcome across days.

between means in Wednesday and Friday compliance, the two days did not have significantly different mean compliance at the $\alpha=0.05$ level (t-statistic=-1.8 p-value=0.08).

The research team computed an overall transmission compliance percentage score for each interviewer. The score was equal to the ratio of the number of transmission dates in which an interviewer transmitted at the correct time to the total number of transmission dates in which he or she was assigned cases²:

$$\text{Transmission Compliance Percentage Score} = \frac{\text{Transmission Days Compliant}}{\text{Transmission Days Worked}}$$

Summary statistics of the distributions of compliance scores for experienced and new interviewers are shown in Table 2. Interviewers working fewer than three transmit days ($n \approx 300$) were dropped from the analyses. First, a Kolmogorov-Smirnov test was used to evaluate the null hypothesis of normality in the distributions of compliance scores for both new and experienced interviewers. In both cases, the null hypothesis was rejected at the $\alpha = 0.05$ level ($D=.116$ and $D=.151$, respectively). Given the nonnormal distributions, we examined characteristics of the distributions beyond the

² This includes times when an interviewer performed a compliant transmission on a day other than Wednesday or Friday and did not work on Wednesday or Friday. It does not include dates in which the interviewer was not assigned cases on the survey, such as before he or she was hired.

Table 2. Selected summary statistics, interviewers who worked >2 days.

	Mean	Median	Mode	Lower confidence limit for the 25th percentile score	Upper confidence limit for the 25th percentile score
Experienced interviewers	89.0 ^a	91.7	1.0	83.3	83.3
New interviewers	84.0	87.5	1.0	75.0	76.5

^a The mean compliance of new interviewers is significantly lower than the mean compliance of experienced interviewers ($t=4.79$, $p\text{-value}<.05$).

mean. The first two columns of Table 2 show that the median scores are higher than the mean scores, and the mode scores are higher than the median scores, which typically indicate that both distributions are left-skewed (Hippel 2005).

The median compliance scores for new and experienced interviewers were compared using a Mood's median test, with a null hypothesis of equality of medians. The null hypothesis was rejected ($z=-4.85$, $p\text{-value}<.05$), meaning the median compliance of new interviewers is significantly lower than the median compliance of experienced interviewers. Finally, in order to further explore the lower tail of the compliance score distributions, confidence intervals for the 25th percentile of each distribution were produced using the Woodruff method (Sitter and Changbao 2001). The last two columns in Table 2 provide the lower and upper bounds of the 95% confidence interval of the first quartile compliance score for each interviewer experience group. Those intervals do not overlap, meaning the 25th percentile score for new interviewers is significantly lower than the 25th percentile score for experienced interviewers. These analyses indicate that, not only is the median compliance score of the new interviewers lower, but the least compliant quarter of new interviewers is also less compliant than the least compliant quarter of experienced interviewers. These differences in compliance could be due to the selection bias of experienced interviewers. When a new hire is consistently noncompliant or underperforming, they are less likely to be asked back to work on future surveys and become experienced interviewers.

Relationship between Compliance and Automated Reminder Call Message

Two identical hierarchical logistic regressions were run, one with only new interviewers and one with only experienced interviewers. The regressions included random intercepts for each interviewer. The datasets used in the models contained one record for each interviewer for each day that he or she worked. In order for the hierarchical models to converge and compute an intercept for each interviewer, only interviewers who worked five transmission days or greater were included in each model. This brought the total observations for experienced and new interviewers to approximately 1,100 and 250, respectively. The model used was:

$$\text{Logit}(p_{ij}) = \beta_i + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij}$$

Table 3. Odds ratio estimates.

Effect	Odds ratio estimate	95% Wald confidence limits	
Message received vs. no message	1.23	1.126	1.344
Message delivered vs. no message	1.152	1.139	1.375
Message received vs. message delivered	-0.0176	-0.117	0.0817
Friday vs. Wednesday	0.691	0.644	0.743

where $\beta_i = \beta_o + \varepsilon_i$

and

$$\text{Logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right)$$

where p is a binary compliance variable, x_1 and x_2 are binary indicator variables for Message Received and Message Delivered, respectively, x_3 is a binary variable for Friday, and β_i is a random intercept for each interviewer, i . No Message was the omitted category. Other variables were not included in the model because additional data on interviewer characteristics, competencies, and workload were not available for this analysis. Parameter estimates from the two models were compared by performing a t-test on each parameter. No significant differences were found, so a model combining new and experienced interviewers was run. Table 3 contains the odds ratios from the combined model. Compared to interviewers whose telephone number received no message, interviewers whose telephone number received the message or had a message delivered were significantly more likely to comply with transmission protocols. No significant difference in compliance was found between interviewers in the message received category or the message delivered category. On Fridays, interviewers were significantly less likely to comply than on Wednesdays.

Cost

The total cost of the automated reminder telephone calls in this study was \$2,641.23, which is approximately eight cents per call, or \$1.60 per interviewer. Since SIPP interviewers were not reminded to transmit in earlier waves, cost data on other reminder methods in SIPP is not available. However, evidence from a 2015 study found that the costs of automated calls were lower than postcards when at least 8,000 reminders are sent (Shoup et al. 2015). Other Census Bureau surveys have instituted automated email reminders to supervisors when interviewer performance such as cost per case becomes statistically different from the expected value. Although the variable costs of such automated reminder systems would be close to zero, there are fixed costs to setting up these systems, and research should be done to test whether such reminders to supervisors rather than interviewers result in increased interviewer compliance.

Organizations can purchase calls from vendors on per call or per minute rates without programming special systems. Some current prices found from an online search for “automated phone calls” are five cents per minute and 200 calls at eight cents per call. These prices make auto calls available to a wide range of surveys.

Conclusion

This study suggests that interviewer compliance with new transmission protocols was high compared to previous studies of interviewer transmission compliance (Walejko and Wagner, forthcoming), and interviewers who listened to the message were more compliant than those who received no message. This is consistent with the National Agricultural Statistical Service study on reminder calls to respondents that found respondents who picked up the automated call responded at higher rates than those who received a voicemail (McCarthy 2007). Thus, reminder calls to interviewers produced similar compliant behavior as reminder calls to respondents when the target audience received the message. The cost of automated reminder telephone calls was small in relation to other costs such as interviewer retraining.

McCarthy (2007) posited the potential utility of automated calls for messages to office staff and interviewers as an area for future research. Our results provide evidence for this utility and suggest automated reminder calls may be a useful tool for researchers hoping to obtain interviewer compliance with new or difficult procedures on a constrained budget. The low cost of automated calls allows researchers to try automated reminder call programs without great expense. If reminder calls are found to be impactful on interviewer compliance, data collections and experimental designs reliant on changes to interviewer behavior may have less implementation error, and experimental effects will be easier to identify.

This study is a retrospective observational study, so we cannot determine a causal relationship between reminder telephone calls to interviewers and compliant data transmission behavior. We recommend incorporating automated reminder calls as an experiment in future data collections to verify causal effects of automated calls on interviewer behavior. For example, future data collections could use a randomized experiment to test the difference in compliance measures between interviewers who receive daily telephone calls and interviewers who do not receive the automated reminders. Furthermore, analyses of such an experiment could include measures of workload and interviewer characteristics, which were not available for this analysis.

Disclaimer

Any views expressed are those of the authors and are not necessarily those of the U.S. Census Bureau.

Acknowledgments

The authors thank Allison Zotti for her help with programming and sharing her knowledge about the data used in this study, Stephanie Coffey for her guidance and help with data analysis, and Rachel Davis for her editing and input. We also thank the U.S. Census Bureau for its support and for conducting the data collection on this study.

REFERENCES

- Hippel, P. 2005. "Mean, Median, and Skew: Correcting a Textbook Rule." *Journal of Statistics Education* 13 (2). www.amstat.org/publications/jse/v13n2/vonhippel.html.
- McCarthy, J. 2007. "Messages as an Alternative to Follow-up Survey Reminder Postcards: The Effectiveness of Autodialer Software Used during the Pennsylvania County Estimates Survey." United States Department of Agriculture, National Agricultural Statistics Service. https://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/reports/RDD0704_mccarthy_autodialer.pdf.
- Shoup, J.A., C. Madrid, C. Koehler, C. Lamb, J. Ellis, D. Ritzwoller, and M. Daley. 2015. "Effectiveness and Cost of Influenza Vaccine Reminders for Adults with Asthma or Chronic Obstructive Pulmonary Disease." *American Journal of Managed Care* 7 (21): 405–13. <http://www.ajmc.com/journals/issue/2015/2015-vol21-n7/effectiveness-and-cost-of-influenza-vaccine-reminders-for-adults-with-asthma-or-copd>.
- Wagner, J., B. West, N. Kirgis, J. Lepkowski, W. Axinn, and S. Ndiaye. 2012. "Use of Paradata in a Responsive Design Framework to Manage a Field Data Collection." *Journal of Official Statistics* 28 (4): 477–99.
- Walejko, G., and P. Miller. 2015. "The 2013 Census Test: Piloting Methods to Reduce 2020 Census Costs." *Survey Practice* 8 (6). <http://www.surveypractice.org/article/2828-the-2013-census-test-piloting-methods-to-reduce-2020-census-costs>.
- Walejko, G., and J. Wagner. 2015. "Challenges to Innovation in Face-to-Face Surveys Posed by Interviewer Noncompliance." In *The 70th Annual Conference of the American Association of Public Opinion Research, Hollywood, FL, May 14–17, 2015*.
- . Forthcoming. "A Study of Interviewer Compliance in 2013 and 2014 Census Test Adaptive Designs." *Journal of Official Statistics*.
- Walejko, G., A. Zotti, and K. Tolliver. 2016. "Survey of Income and Program Participation Case Prioritization Experiment Results." In *The Joint Statistical Meetings 2016 Survey Research Methods Section, Chicago, IL, July 30–August 4, 2016*.