A Study of Interviewer Compliance in 2013 and 2014 Census Test Adaptive Designs

Gina Walejko[1] and James Wagner[2]

Researchers are interested in the effectiveness of adaptive and responsive survey designs that monitor and respond to data using tailored or targeted interventions. These designs often require adherence to protocols, which can be difficult when surveys allow in-person interviewers flexibility in managing cases. This article describes examples of interviewer noncompliance and compliance in adaptive design experiments that occurred in two United States decennial census tests. The two studies tested adaptive procedures including having interviewers work prioritized cases and substitute face-to-face attempts with telephone calls. When to perform such procedures was communicated to interviewers via case management systems that necessitated twice-daily transmissions of data. We discuss reasons when noncompliance may occur and ways to improve compliance.

Key words: Computer-assisted personal interviewing; decennial census.

1. Introduction

Researchers are interested in measuring the effectiveness of adaptive and responsive survey designs that monitor frame data, paradata, and survey response data and react to this information using tailored or targeted interventions (Groves and Heeringa 2006; Kirgis and Lepkowski 2013). While several studies have successfully evaluated adaptive design experiments that call cases at specific times or stop effort on unproductive cases in computer-assisted telephone interviewing (CATI) systems (e.g., Coffey 2013; Luiten and Schouten 2013; Wagner 2013a), those that measure the effectiveness of adaptive designs in computer-assisted personal interviewing (CAPI) environments are scarce.

This article suggests that few in-person adaptive design studies have been executed and reported because interviewer noncompliance can limit the effectiveness of these interventions, making them difficult or impossible to evaluate. In contrast, when interviewers follow intervention protocols, researchers can evaluate their effectiveness. In-person adaptive design experiments often rely on computerized case management systems that allow interviewers much flexibility in managing their workload, including the number of calls made to each case and the timing of those calls (Morton-Williams 1993). Overhauling these case management systems completely to test an adaptive design

experiment may be unfeasible. Thus – in CAPI adaptive design experiments – randomization often relies on in-person interviewers working cases precisely as communicated by their case management systems, which interviewers may choose not to follow. This decision may or may not be for legitimate reasons, such as a change in their schedule making it inefficient to drive to an address across town or having personal knowledge about a housing unit that increases its likelihood to be interviewed. Whether for valid reasons or not, noncompliance limits the ability of survey organizations to implement centralized decision rules typical of adaptive designs.

This article describes examples of interviewer noncompliance and compliance in adaptive design experiments that occurred in two decennial census tests – tests that required in-person interviewers to follow field procedures specified in their case management systems to implement adaptive designs. Conducted by the United States Census Bureau, the two tests occurred at different sites and used different interviewers, supervisors, trainings, monitoring infrastructure, and case management systems. We briefly present the test results from the first experiment and then describe how interviewers did not implement the treatment. We next discuss results from the second experiment and then show that interviewers largely were compliant and applied the treatment. Where possible in both experiments, we attempt to explain why interviewers did or did not comply. Our analysis illustrates the challenges associated with controlling field procedures when implementing adaptive designs in CAPI environments and may be of value to survey staff interested in experimenting with or implementing field procedures that rely on interviewers following instructions.

2.   Background

Responsive or adaptive designs attempt to alter survey protocols either by targeting particular cases or subgroups to receive differential treatment or by changing protocols over time. Often the goal of these interventions is to optimize the allocation of resources such that total survey error is minimized for a fixed cost.

Several web and telephone surveys have used computerized case management systems to implement responsive or adaptive interventions that achieve improvement in cost or data quality by prioritizing cases, calling cases at specific times, or assigning groups of cases to specific interviewers. Statistics Canada prioritized cases in several CATI surveys – including the Households and the Environment Survey and the Survey of Labour and Income Dynamics – and found this adaptive approach led to lower total system time (i.e., cost savings) and similar response rates in both surveys when compared to control methods (Laflamme and St-Jean 2011). The Survey of Consumer Attitudes (SCA) experimentally altered CATI calling algorithms to call during time windows when cases had the highest estimated probability of contact and found calls made during such time windows had a higher contact rate in the experimental group (Wagner 2013a). Statistics Netherlands assessed a pilot test designed to increase representativeness and reduce cost in the Survey of Consumer Sentiment. During the first survey wave, they grouped cases into high, low, or medium cooperation based on predicted scores. Depending on group, cases were sent one of three invitations to participate: web, mail, or choice (i.e., web or mail). In CATI follow-up to nonrespondents, the same survey assigned different call schedules to

groups with high, medium, or low contact propensities, and the highest-performing CATI interviewers were assigned to the lowest cooperation group and vice versa. They found tailored strategies increased representativeness at comparable – although slightly higher – cost (2.6 percent) and obtained similar response rates (Luiten and Schouten 2013).

CAPI surveys attempting adaptive survey interventions have found mixed results due, in part, to a lack of compliance with requests from central office staff. For example, the National Survey of Family Growth (NSFG) experimented with interviewers working prioritized cases. In only two of 16 separate experimental interventions were response rates significantly higher in the experimental group, which may have been due to lack of compliance. Interviewers made more calls on prioritized cases in all 16 experiments but in only seven were call attempts significantly higher (Wagner et al. 2012).

Other adaptive survey designs that relied on in-person interviewers to implement experimental manipulations could not be evaluated since control and treatment interviewers behaved the same. Similar to the SCA, each day the NSFG estimated time windows during which cases had the highest probability of contact. CAPI case management systems stored and showed the recommended call time to in-person interviewers in the treatment condition. Interviewers in the control who were not shown such call times happened to coincide attempts with recommended windows 23.0 percent of calls while treatment interviewers who were shown recommended call times made attempts during suggested windows only 23.6 percent of the time. In debriefings, interviewers said they did not follow recommendations because geographically clustered cases did not always have the same suggested time windows. The authors note that – rather than attempt to balance the efficiencies of clustered cases and predicted time windows – interviewers stuck with typical behavior, calling cases at time windows of their convenience (Wagner 2013a).

No evidence suggests that interviewer compliance is worse in adaptive survey designs than other types of field surveys. In fact, other survey experiments have experienced issues analyzing results because interviewers did not follow procedures. NSFG interviewers were asked to leave a "Sorry I Missed You" card at households where such a notice was estimated to increase the probability of contact. Interviewers ignored these instructions (Wagner 2013b) leaving researchers unable to evaluate the effects of the card. Biemer et al. (2013) report interviewers admitted they did not record every call attempt as required to avoid having cases hit a specified cap on the number of allowable call attempts or because interviewers were unclear about what constituted a call attempt (e.g., a "drive-by" sighting that no one is home).

To advance our understanding of interviewer compliance and its effects on evaluating adaptive interventions, this article examines the results of interviewer behavior associated with the 2013 and 2014 Census Tests. These decennial census field studies tested adaptive procedures including having interviewers (1) work prioritized cases and (2) supplement face-to-face attempts with telephone calls to specified sample units. The first intervention also depended upon interviewer compliance with a requested twice-daily transmission of data made from laptop computers to databases maintained in the central office.

The focus of the article is on interviewer compliance, an important issue for adaptive designs in CAPI settings. The experimental results themselves are less interesting since they are difficult to interpret in the presence of noncompliance, and the methods may not

be useful for surveys other than the decennial census. First, we describe a case prioritization intervention. Other surveys have successfully implemented this type of intervention (Wagner et al. 2012; Peytchev et al. 2010). In this article, interviewers did not implement the intervention as designed. Further, it did not increase contact and completion rates. We then compare this experiment to an intervention that – largely – interviewers correctly implemented. This intervention led to a reduction in personal visit attempts per case. Our discussion concludes with reasons for noncompliance and how requests to CAPI interviewers in experiments might be improved.

3.   Data and Methods

3.1.   2013 Census Test

The 2013 Census Test piloted subsequent decennial census test procedures between October and December 2013. Census Bureau staff selected 2,077 sample addresses from six block group pairs in Philadelphia. One block group in each matched pair was assigned randomly to "No Priority Condition" interviewers and the other to "Adaptive Condition" interviewers. No Priority interviewers served as the control group for the test.

3.1.1.   Interviewers

Eighteen interviewers who had recently finished working on another survey were selected to work on this pilot because supervisors recommended them, and they had better histories of recording contact attempt information in a previous survey. Eight interviewers were assigned randomly to work Adaptive Condition cases and ten to work No Priority Condition cases. Two supervisors from the field office managed each condition separately. More detail on the 2013 Census Test can be found in Walejko et al. (2014).

3.1.2.   Intervention Goal

One goal of the 2013 Census Test was to measure the effect of case prioritization on efficiency. Up to seven cases with the highest predicted propensity to respond on the next contact attempt were prioritized on each Adaptive interviewer's case list. (Adaptive interviewers may have received more or fewer than seven "high priority" cases due to reassignments between interviewers, interviewers not transmitting, or other anomalies.) Cases were rescored, and priority cases were updated daily. Geography was not used in creating this prioritization. Prioritized cases could fall anywhere within the six block groups assigned to the Adaptive condition. For this intervention to be implemented, interviewers needed to attempt all seven high priority cases every day they worked. Success metrics for this intervention included higher contact and completion rates on prioritized cases.

3.1.3.   Training, Supervising, and Monitoring

Supervisors instructed Adaptive and No Priority interviewers separately over the course of a two-day training. Supervisors instructed all interviewers that the test was about following instructions provided to them through their case management systems. Trainings, training manuals, and job aids highlighted the importance of Adaptive

interviewers attempting all priority cases every day they worked. Adaptive interviewers were only to attempt a "regular" case (i.e., not high priority) if it was nearby or had an appointment. No Priority interviewers were instructed to work cases using Census Bureau survey guidelines that allow flexibility in which cases they visited and when they made contact attempts. After monitoring in-person interviewers' performances and observing poor compliance, supervisors performed a half-day refresher training on the 17th day of data collection for Adaptive interviewers to increase understanding of 2013 Census Test procedures including working prioritized cases.

An interviewer performance report monitored all interviewers daily on specific field procedures. The report tracked data transmission compliance as well as daily counts of attempted high priority cases for each of the Adaptive interviewers. Headquarters and field staff conducted a daily meeting during which they discussed this report and other interviewer performance topics. Supervisors were instructed to address noncompliance observed in the report by talking to interviewers.

3.1.4. Case Management System

The 2013 Census Test used many existing Census Bureau information technology resources including a computerized case management system located on interviewer laptops. (See Figure 1 for a screenshot of the 2013 Census Test Adaptive Condition case management system.) Each high priority case, designated with a unique control number, was preceded by an exclamation point, and high priority cases were sorted to the top of the case list. Cases did not have a priority indicator on No Priority interviewers' case lists.

3.1.5. Data Transmissions

Interviewers in the 2013 Census Test needed to transmit data from their case management systems to the operation control system twice daily, once before they started work and once after they completed work for the day. Before-work transmissions pulled any updated interviewer instructions from the central control system to interviewers' case lists. Daily instructions updated which cases were prioritized. After-work transmissions pushed contact history information and outcome codes from interviewer laptops to the control system so that instructions for the next day could be calculated by business rules. Transmissions needed to occur after and before set times – not too late at night or early in the morning. Due to the six-time zone span of the U.S. (three in the continental U.S.), a decennial census would need transmissions to occur so work transmitted late at night in the

Fig. 1. Screenshot of 2013 Census test adaptive condition case management system.

west could be read in by business rules that assigned instructions and would be available early in the morning for interviewers who transmitted in the east.

In the 2013 Census Test, all transmissions required interviewers to log into their laptop case management system, connect to the internet, and click a "transmission" button (i.e., transmissions were manual). Interviewers were instructed to transmit once before working, no earlier than 8:00 a.m., and once in the evening after they were done working, no later than 10:00 p.m.

3.2. 2014 Census Test

In 2014, the Census Bureau carried out a larger field test of adaptive procedures. The CAPI portion of this test included 46,247 sample addresses located in Washington, DC and Montgomery Country, Maryland and ran during August and September 2014. More detail on the 2014 Census Test can be found in Poehler and colleagues (2016). These sample addresses were assigned in geographic clusters to one of three interviewing conditions: a "Control Condition," an "Experimental Contact Strategy Condition," and an "Adaptive Condition." (We do not analyze the Adaptive Condition in this article because it employed CATI interviewers to call sample addresses rather than CAPI interviewers.) Control Condition interviewers had much flexibility as to how they contacted cases, similar to 2010 Census enumerators. They were instructed to perform a personal visit first and then up to two more personal visits and three telephone attempts at their discretion. Experimental Contact Strategy interviewers were instructed to make no more than three total contact attempts (although the case management system allowed more than three) with the first being a personal visit, the next being a telephone contact, and the final a personal visit. Although not an example of "dynamic" adaptive design, the Experimental Contact Strategy can, in our opinion, be considered a "static" adaptive design (see Schouten et al. 2013) because the protocol instructing interviewers to make the second attempt by telephone was applied to only addresses that had been successfully matched to telephone numbers, 81.2 percent of cases.

3.2.1. Interviewers

The Census Bureau recruited interviewers from the area specifically to work on the 2014 Census Test. Interviewers were new hires who were not required to have past interviewing experience, although many had worked on previous decennial censuses. Crew leaders, also new hires, supervised these interviewers, and were, in turn, supervised by higher-level managers. The Control and Experimental Contact Strategy had 304 and 389 employees who recorded making at least one call attempt or transmitting, respectively. (Employees were not spread evenly across conditions because conditions did not have the same number of sample cases. Two additional interviewers transmitted but did not make any contact attempts.)

3.2.2. Intervention Goal

One goal of the 2014 Census Test was to measure the cost effect of substituting costlier in-person visits in place of telephone call attempts made by CAPI interviewers on

second contact attempts for addresses with matched telephone numbers. The goal of this approach was to decrease nonresponse follow up costs by reducing the total number of attempts and, specifically, the number of personal visit contact attempts. For this intervention to be implemented correctly, Experimental Contact Strategy interviewers needed to attempt a phone call to all open cases with matched telephone numbers on the second contact attempt. (Using the Census Bureau's Master Address File ID, 81.2 percent of cases were matched to phone numbers available from several commercial data files.) In contrast with the 2013 intervention, this intervention did not require transmission since it could be implemented algorithmically via the case management system (i.e., using a programmed rule such as "if the case has a telephone number and one attempt, the next attempt should be made via the telephone"). Success metrics for this intervention included lower costs, measured by fewer contact attempts per case, fewer personal visits per case, and fewer average attempts – notably personal visit attempts – per completion.

3.2.3. Training, Supervising, and Monitoring

Over the course of three days, crew leaders instructed interviewers how to plan their day, follow field procedures, record contact history information, transmit their data, and perform interviews. Trainings, training manuals, and job aids instructed Experimental Contact Strategy interviewers to call all cases with matched telephone numbers after first attempting a personal visit. One half-day of training was devoted to interviewers performing production interviews and supervisors reviewing this work. Supervisors were instructed to use reports that monitored their interviewers' activities including interviewers' transmissions. (Reports did not monitor whether second contact attempts were done by phone or in-person.)

3.2.4. Case Management System

For the 2014 Census Test, the Census Bureau developed a new computerized case management system available to interviewers as an iPhone application. This system functioned similarly to that used in 2013, providing interviewers with a list of their cases and instructions on how to work each case as well as allowing interviewers to collect interview data and record contact attempt information. The Experimental Contact Strategy interviewer's case management system indicated when to do a telephone attempt and provided these interviewers with the matched telephone numbers. (See Figure 2 for a screenshot of this case management system. The box with a "T" inside it indicates that the interviewer should make a telephone attempt on the indicated case.)

3.2.5. Data Transmissions

In contrast with the technical systems used for the 2013 Census Test, the 2014 Census Test case management system was designed to manage data transmissions automatically, and transmissions were not necessary for interviewers to be displayed the correct mode. The system attempted automatic transmissions when two hours had passed since the last successful transmission, when an interviewer logged into the app or completed a contact attempt, and when an interviewer completed a case or logged out of the app. The case
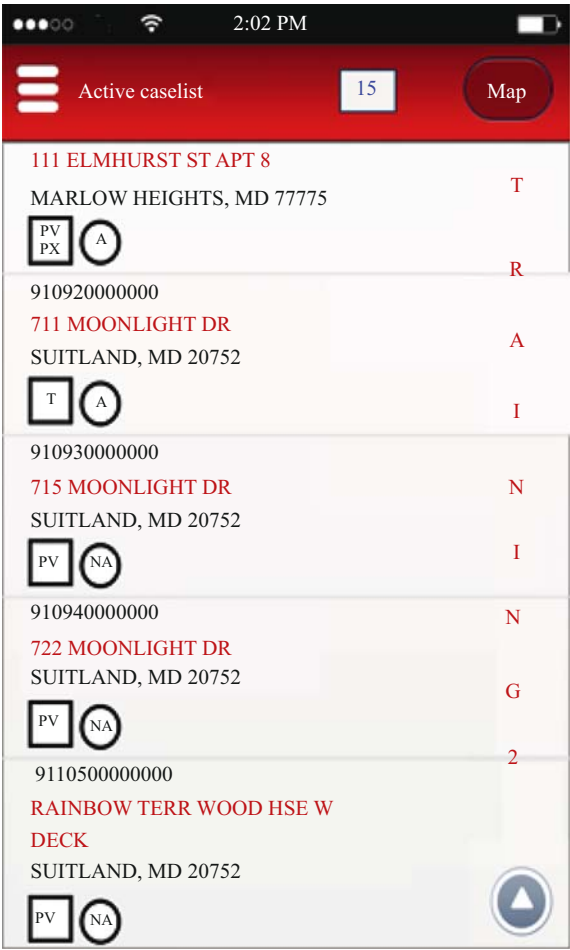
Fig. 2.   Screenshot of 2014 Census test experimental contact strategy condition case management system.

management system itself kept track of whether a case had a matched phone number as well as the number, mode, and outcome of each contact, allowing mode to be displayed correctly without a transmission.

   Some protocols not analyzed in this article necessitated twice-daily transmissions, and automated transmissions would not work if interviewers became disconnected from the Census Bureau's network, for example, by driving or walking through an area without cell coverage. For this reason, 2014 Census Test interviewers were instructed to transmit manually twice each day that they worked, once before working no earlier than 7:00 a.m. and again after working but no later than midnight. (After 2013 Census Test results and debriefings uncovered interviewers had difficulty transmitting between 8:00 a.m. and 10:00 p.m., the time period within which interviewers were instructed to transmit on days they worked was expanded to between 7:00 a.m. and midnight for the 2014 Census Test.) Interviewers were able to view when their last successful transmission occurred using the application.

4. 2013 Census Test Results

Table 1 summarizes the design of both 2013 and 2014 Census Tests. Results of the 2013 Census Test did not support the hypothesis that the Adaptive Condition would have higher contact and completion rates than the No Priority Condition in the 2013 Census Test. Instead, contact rates on personal visits were significantly higher in the No Priority Condition than the Adaptive Condition. Furthermore, completion rates on personal visits were the same (Table 2). To help understand these results, we examine interviewer compliance with implementing two necessary actions, transmitting data twice daily and attempting prioritized cases daily. These results showed interviewers did not comply in either transmitting data or working prioritized cases. The fact that the Adaptive Condition interviewers did worse in terms of contact and completion rates seems to indicate that this intervention would not achieve the stated aims. However, given the selective nature with which it is applied, higher or lower rates may have been achieved if the intervention had been applied to the full sample. In any event, poor compliance with the requested actions discussed in the next two sections limits the ability of the central office to implement case prioritization schemes aimed at controlling which cases respond.

4.1. 2013 Test: Interviewers Transmit Data Twice Daily

In the 2013 Census Test, interviewers transmitted as instructed (i.e., once before working and once after working between 8:00 a.m. and 10 p.m.) over 71 percent

Table 1. 2013 and 2014 Census test designs.

	2013		2014	
	No priority	Adaptive	Control	Experimental contact strategy
Interviewers	10	8	304	389
Location	Philadelphia, PA		Washington, DC and Montgomery county, MD	
Case management system	Modified existing survey system using laptops		New system using cell phone application	
Training	Separate for each condition; 2-day training on procedures; half-day refresher training; training manual, job aid		Separate for each condition; 3-day training on procedures with 1/2 of day for supervisor review of work; training manual, job aid	
Monitoring and supervision	Performance monitoring report; daily meetings; feedback to interviewers via supervisors		Performance monitoring reports; feedback to interviewers via supervisors	

Table 2. 2013 Census test contact and completion rates on personal visits between adaptive and no priority condition interviewers.

Condition	Number	Contact percent	Standard error of percent	p-value	Completion percent	Standard error of percent	p-value
Adaptive	1,283	24.50	3.20		18.97	2.77	
				0.03			0.86
No Priority	1,354	31.73	2.35		19.69	3.13	

Note: Standard errors and significance take into account clustering by interviewer.
Note: Includes both compliant and non-compliant transmissions.
Note: Excludes personal visit attempts where an appointment was set.

(standard error, 7.6 percent) of days worked (i.e., all days each interviewer worked summed over all interviewers). Figure 3 shows the number of interviewers grouped by five categories of percent compliant daily transmissions. Compliant transmission days ranged between 14 and 100 percent by interviewer, with nine of 18 having over 80 percent compliant transmission days. Five interviewers had 40 percent or fewer compliant transmission days.

Transmission compliance varied across time ranging between 0 and 100 percent over the 2013 Census Test field period. Figure 4 shows the percent of interviewers who worked and transmitted correctly each day, where compliance is measured as transmitting as instructed – once before working no earlier than 8:00 a.m. and once in the evening after they were done working, no later than 10:00 p.m. Small numbers of working interviewers explain peaks in low compliance and high compliance. On December 2, a day with no compliant transmissions, one interviewer worked, and on November 28, only five interviewers worked. On November 24 and December 3, days with 100 percent compliance, fewer than three interviewers worked. On other days with poor compliance, interviewers often transmitted earlier or later than instructed. For example, on November 7, fifteen of sixteen working interviewers transmitted before 8:00 a.m., but only three
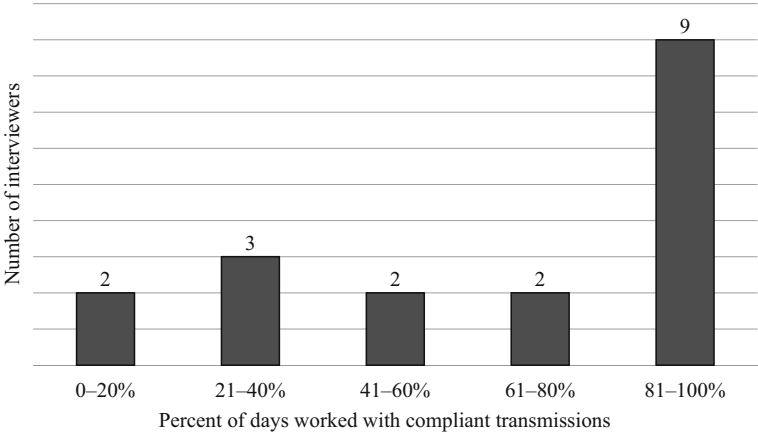

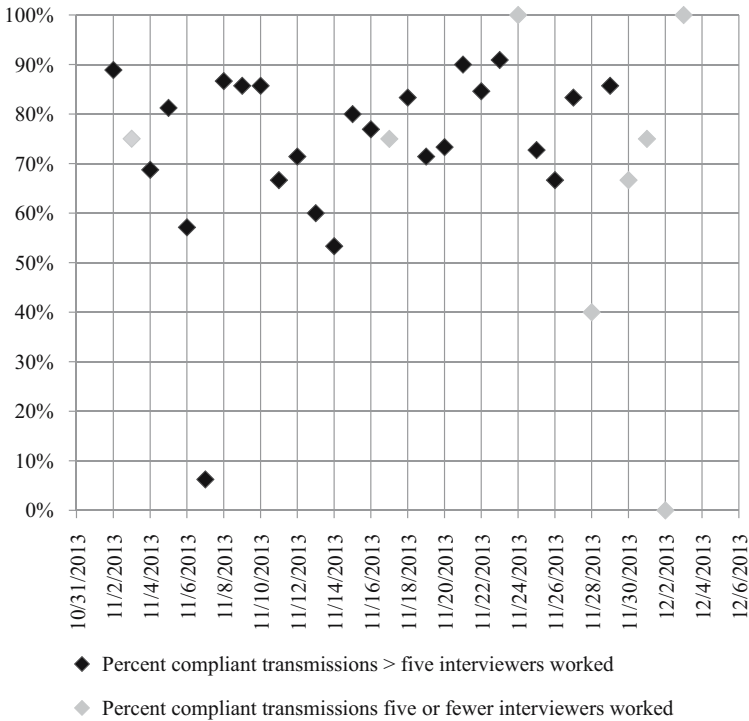*Fig. 3. 2013 Census test compliant daily transmissions.*

Fig. 4.   *2013 Census test percent compliant transmissions by day.*

transmitted after 8:00 a.m. as instructed. On the five days with the smallest percent compliant transmissions, 73 percent of interviewers transmitted between 6:00 a.m. and midnight, either earlier or later than instructed (i.e., between 8:00 a.m. and 10:00 p.m.). These results and debriefing led the 2014 Census Test to expand the time period within which interviewers were instructed to transmit on days worked.

4.2. 2013 Test: Interviewers Attempt Prioritized Cases Daily

In order to provide a basis for assessing compliance, 2013 Census Test interviewers needed to receive high priority cases each day they worked. Because 2013 Census Test interviewers did not transmit correctly on 29 percent of days they worked, we broaden the definition of successful transmissions to include those that occurred between 6:00 a.m. and midnight, which includes more days for analysis. To avoid confusion, we will call these "reliable transmissions."

In 2013, the eight adaptive interviewers worked all high priority cases on 45 percent of days with reliable transmissions. These interviewer days are compliant. On seven percent of days with reliable transmissions, interviewers did not attempt all high priority cases but also did not attempt other, "regular" cases, which may indicate they ran out of time before attempting all high priority cases. These interviewer days are potentially compliant. Interviewers did not attempt all high priority cases and worked regular cases on nearly 48 percent of days they transmitted reliably. These days are not compliant.
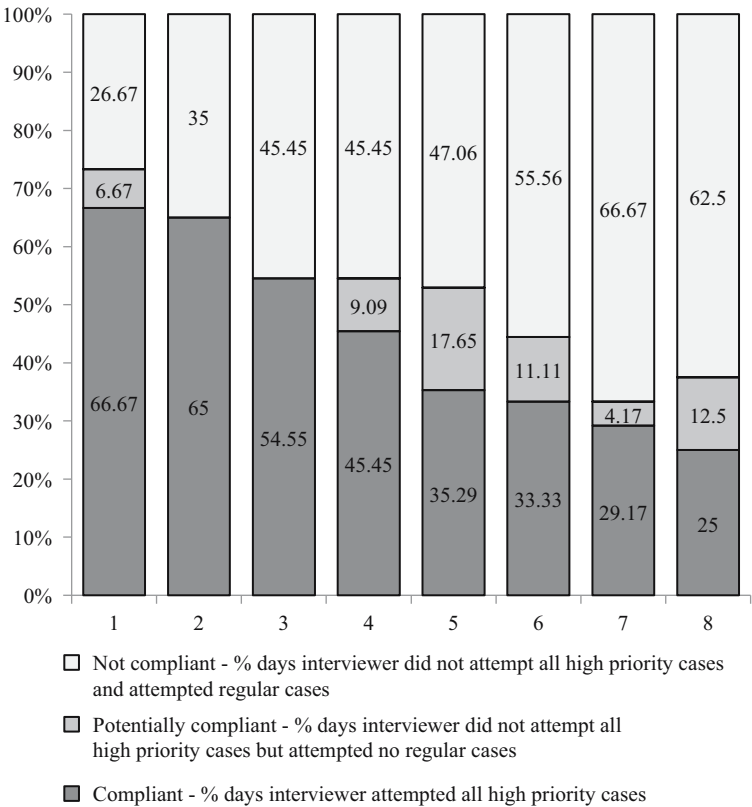
Fig. 5. Percent compliant or potentially compliant days for CAPI interviewers with reliable transmissions in 2013 Census test (Adaptive condition).

Across all eight interviewers, the days they worked all high priority cases ranged between 25 and 67 percent (Figure 5). Two interviewers worked all their high priority cases over 65 percent of the days on which they worked and made reliable transmissions, while two attempted regular cases even though they did not attempt all their high priority cases over 60 percent of the days they worked. This lack of compliance is an interesting result in its own right, as it hampers the ability of data collection operations to implement centrally directed interventions. In this case, it appears that the intervention would not have met its goals, but the lack of compliance is an important finding for other field surveys attempting to prioritize cases.

5. 2014 Census Test Results

In contrast to the 2013 Census Test, results from the 2014 Census Test support the hypothesis that Experimental Contact Strategy interviewers performed actions leading to a reduction in cost; they made fewer average contacts and personal visits per case than Control interviewers. They also had a lower average attempts per complete than Control interviewers and a notably lower average number of in-person attempts per complete than Control interviewers (Table 3). However, the mean number of attempts for the

Table 3. 2014 Census test mean attempts and average attempts per complete between control and experimental contact strategy conditions.

Condition	Number of cases	Mean attempts	Mean in-person attempts	Percent complete	Average attempts per complete	Average in-person attempts per complete
Control	7,394	4.29	3.14	0.62	6.95	5.07
Exp. Contact Strategy	8,873	3.25	2.41	0.57	5.75	4.26

Note: Includes only cases with matched telephone numbers.

Experimental Contact Strategy (3.25) is still greater than 3, indicating some noncompliance. This reduction in effort also appears to have reduced the completion rate relative to the Control Condition (0.57 for the Experimental Contact Strategy Condition and 0.62 for the Control Condition). The Experimental Contact Strategy Condition had about 76 percent of the effort measured as attempts relative to the Control Condition and produced 92 percent of the completion rate relative to the Control Condition. However, the poor compliance observed in the 2013 Census Test led researchers to investigate in more detail the extent to which interviewer compliance may have affected the 2014 data.

5.1. 2014 Test: Interviewers Perform Telephone Calls when Instructed

As shown in Table 4, Experimental Contact Strategy interviewers for the 2014 Census Test followed mode instructions on over 88 percent of contact attempts to cases with matched numbers. (See Table 5 for a summary of compliant procedures.) They performed personal visits as instructed over 99 percent of the time on the first contact attempt. Compliance in attempting contact by telephone on the second attempt was 82 percent, which differs starkly from the control interviewers who performed personal visits 72 percent of the time on the second attempt. In this experiment, case management systems directed interviewers to perform a contact attempt in a particular mode, and it appears this directive changed interviewer behavior when comparing the second contact attempt of the Experimental Contact Strategy and Control interviewers. (The interpretation of results does not change when analysis includes cases without matched numbers.)

5.2. 2014 Test: Interviewers Transmit Data Twice Daily

Unlike the 2013 Census Test, 2014 Experimental Contact Strategy interviewers did not need to transmit twice daily to receive updated mode instructions. However, other protocols not analyzed in this article did rely on data transmissions, so researchers analyzed whether automated transmissions helped interviewers to transmit their data twice daily, once before and once after work.

In the 2014 Census Test, transmissions – either automated or manual – occurred once before working no earlier than 7:00 a.m. and once after working before midnight on 43

Table 4. *Number and percent of compliant and non-compliant mode contact attempts by contact attempt number for the 2014 Census test (Experimental contact strategy and control conditions).*

| Contact attempt | Experimental contact strategy | | | | | Control | | | | |
| | Compliant mode | | Non compliant mode | | Standard error of percent | Personal visit mode | | Telephone mode | | Standard error of percent |
	n	Percent	n	Percent		n	Percent	n	Percent	
1st	8,787	99.03	86	0.97	0.28	7,335	99.20	59	0.80	0.29
2nd	5,722	81.67	1,284	18.33	1.40	4,284	72.15	1,654	27.85	2.23
3rd	4,942	84.94	876	15.06	0.90	3,121	65.73	1,627	34.27	2.15
4th+	6,010	83.83	1,159	16.17	1.05	8,440	61.75	5,228	38.25	1.86
Total	**25,461**	**88.20**	**3,405**	**11.80**	**0.60**	**23,180**	**73.01**	**8,568**	**26.99**	**1.24**

Note: Includes only cases with matched telephone numbers.

Note: The Experimental Contact Strategy called for the first and third attempts to be in-person and the second by telephone. Although interviewers were instructed to perform three attempts total, the case management system allowed for more.

Note: Excludes records of inbound telephone call attempts.

Note: Standard errors take into account clustering by interviewer.

Table 5. 2013 and 2014 Census test procedures.

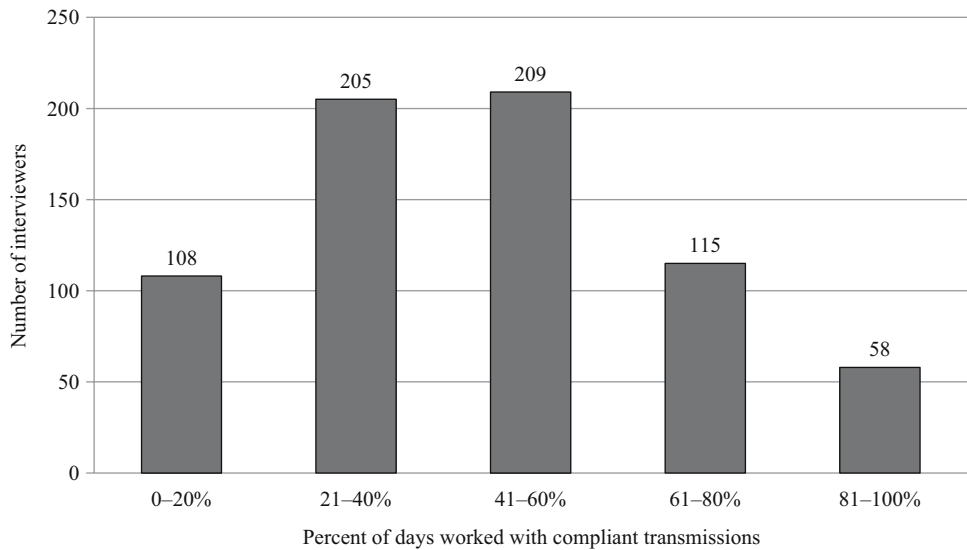|	2013		2014	
	No priority	Adaptive	Control	Experimental contact strategy
Data transmissions	Manual; 2 times on work days before and after attempts; no earlier than 8:00 a.m.; no later than 10 p.m.		Automatic and manual; 2 times on work days before and after attempts; no earlier than 7:00 a.m.; no later than midnight	
High priority cases	Cases attempted at interviewer discretion; no high priority	7 highest propensity cases prioritized daily; interviewers attempt all daily	*Not analyzed here*	
Telephone calls	*Not analyzed here*	Ring cases with numbers at interviewer discretion		Ring cases with numbers on 2nd attempt
Total contact attempts allowed	3 in-person before proxy attempts; interviewers ring cases with numbers up to 2 times	3 in-person before proxy attempts; CATI before field	6 total before proxy attempts with 3 in-person and 3 phone if number	3 total before proxy with 2 in-person and 1 phone if number

Fig. 6.   *2014 Census test compliant daily transmissions.*

percent (standard error, 0.9 percent) of days worked (i.e., all days each interviewer worked summed over all interviewers). Compliant transmission days ranged between 0 and 100 percent by interviewer, with 58 of 695 interviewers (8.3 percent) having over 80 percent compliant transmission days and 108 interviewers (15.5 percent) having less than 21 percent compliant transmission days (Figure 6).

6.   Discussion

6.1.   Limitations

Results should be considered in conjunction with several study limitations. First, the 2013 and 2014 Census Tests recruited interviewers from just two geographic sites. As a result, the tests are not generalizable to the broader United States. Second, the 2013 test contained a sample size of only 18 interviewers. A larger interviewer workforce could have led to different results.

6.2.   Summary

Interviewers were somewhat compliant in transmitting data and receiving updates twice daily during the 2013 Census Test. Overall, interviewers transmitted correctly on 71 percent (standard error, 7.6 percent) of days worked. For this test, transmissions were manual. We found lower compliance during the 2014 Census Test where case management systems transmitted automatically, but – as a backup – trainings and training materials instructed interviewers to transmit manually twice daily. Interviewers did not perform this backup transmission, and compliant manual or automated transmissions occurred only 43 percent (standard error, 0.9 percent) of the days interviewers worked.

Lack of compliance and improper functioning of automated transmissions meant instructions (i.e., prioritized cases) were not updated every day interviewers worked during the 2013 Census Test.

In the 2013 Census Test, prioritized cases did not have higher contact and completion rates than nonprioritized cases running counter to our hypothesized result. However, the request to attempt contact on prioritized cases met with low compliance. During the 2013 Census Test, Adaptive interviewers worked all high priority cases on fewer than half (45 percent) of the days they worked, and this percentage is limited to days with reliable transmissions only. While acknowledging 100 percent compliance is unrealistic, noncompliance observed in the 2013 Census Test affected our ability to analyze an adaptive intervention by limiting the number of days we could evaluate interviewers working *all* prioritized cases on their case lists to only a nonrandom 32 percent of days. Regardless of the potential benefit, this noncompliance limits the ability of the central office to intervene by prioritizing cases.

The 2014 Census Test results showed Experimental Contact Strategy interviewers performed actions that led to cost reductions including having lower average attempts per complete than Control interviewers. In contrast with the 2013 Test, we observed quite high compliance with the request that in-person interviewers attempt telephone calls rather than personal visits at certain points in 2014 data collection.

6.3. Reasons for Noncompliance

There are several reasons interviewers may have been noncompliant in transmitting and working prioritized cases. First, as with any kind of job, interviewers may have life circumstances such as a sudden change in their planned schedule due to a sick family member or a safety concern with approaching a sampled housing unit that prevents them from carrying out their assigned tasks. For example, one Adaptive interviewer's high priority case was a house where illegal drug trade occurred, so they did not visit it. Under such circumstances during the 2013 Census Test, it was unrealistic for the interviewer to carry out contact attempts following algorithmic rules.

Second, instructions relayed via the case management system allowed interviewers the flexibility to be noncompliant. Years ago, case management for CAPI interviewers constituted a pen-and-paper system that communicated which addresses to work – usually located nearest to where an interviewer lived – with space to fill out contact information. A historical artifact of paper, most digital CAPI case management systems today supply interviewers with a list of cases to work and leave much to their discretion, including when to work, which cases to attempt, and how frequently to make attempts. The 2013 Census Test interviewers could and did choose to work regular, nonprioritized cases. For example, in debriefings some interviewers did not like having to return to the same block the next day. Although CAPI sample management systems designed to constrain interviewers to follow instructions would be preferred for testing experimental manipulations, reprogramming such systems for a test is cost prohibitive for most organizations that perform CAPI surveys. This finding led to the development of new systems that constrained interviewers to attempt contact on sets of cases selected daily by the central office, as these were the only cases displayed to interviewers.

Third, technical issues could explain at least some of the discussed noncompliance in transmitting data to and from laptop or smartphone case management systems. For example, during the 2013 Census Test, a few interviewers whose data showed continuous transmission issues claimed they were transmitting twice daily as instructed or that they could not transmit as instructed. In debriefings, several interviewers said they needed to transmit more than once before the server would connect to their laptop. Because the 2014 Census Test used smartphones rather than laptops, it is possible that interviewers attempted but were unable to transmit because they lost reception causing them to no longer be connected to the network. During 2014 debriefings, a majority of interviewers and supervisors brought up issues with cellphone reception.

It is also possible that interviewer noncompliance happened when following new procedures competed with other interviewer activities. For example, interviewers in the 2013 Census Test mentioned it was difficult to transmit before 10 p.m. on days when interviewing lasted late into the evening. Supervisors in the 2013 test also reported it was difficult to balance managing interviewer noncompliance with other supervising responsibilities and that a rolled-up report of potential problems to discuss with interviewers could help to alleviate this time pressure. The 2013 and 2014 test interviewers remarked that, when they saw a respondent near their address, they attempted an in-person interview with that respondent, even if their address was not a high priority case or was supposed to be attempted via telephone.

Fifth, the nature of the intervention itself may have led to interviewers to be more accepting of 2014 Census Test procedures than 2013 procedures. In debriefings, all 2014 Census Test Experimental Contact Strategy interviewers reported understanding the test procedures for conducting telephone calls – attempt a contact in the mode that the case management system instructed. However, at least one Adaptive interviewer in the 2013 Census Test admitted it was unclear why case management systems deemed cases as high priority, indicating confusion regarding the nature of the intervention itself.

Finally, it may be that Adaptive interviewers' experience with previous surveys, where they had wider discretion, may have made it more difficult to train them to work under a new centrally directed approach. A common theme in debriefings with experienced interviewers were differences between previous data collection procedures including the 2010 Census. For example, a few Adaptive interviewers did not like planning their route in the morning after an early data transmission provided them new instructions, as they were accustomed to doing it the night before. The interviewers used in the 2013 Census Test were experienced. While we do not have data on 2014 Census Test interviewers' past interviewing experience, being less seasoned and – thus – less inclined to recall past protocols, may have played a role in interviewers following the suggested mode.

Current practice allows interviewers wide latitude for deciding how to conduct their work, but the experimental adaptive design interventions described in this article restrict this range. A tension exists between the centralized, data-driven control of interviewers and decentralized decision making by interviewers who rely upon their expertise and local knowledge. While experienced, expert interviewers with local knowledge may perform at a higher level than if centrally directed, interviewers vary in their ability to plan efficient trips and recruit respondents in practice (Wagner and Olson 2011; O'Muircheartaigh and

Campanelli 1997; Purdon et al. 1999; Pickery and Loosveldt 2002; Durrant and Steele 2009). Further, local interviewers are unable to make decisions that balance response across more cases than their own sample. While centralized, data-driven interventions like those described in this article may go against current interviewer practices, they improve the ability of data collection organizations to control important aspects of the response process, including balanced respondent pools and overall data collection costs. Finding a balance between centralized and decentralized procedures remains a complex function involving the available interviewing staff, the capabilities of the data collection organization, the particulars of the survey design, and the overall goals of the survey. Finding the correct equilibrium may also require considering training, explaining the intervention purpose to interviewers, and other actions considered in the next section.

6.4. Ways to Improve Compliance

Although not possible in the experiments described here, researchers could construct some protocols to constrain CAPI interviewers into compliance. Kreuter and colleagues (2014) found setting prespecified appointments based on the prior wave interview date in the Medical Expenditure Panel Survey Household Component significantly decreased the number of attempts (e.g., phone, in-person, letter) to get an interview. To increase the likelihood that interviewers kept appointments, they mailed sample addresses a postcard with the appointment date and time so "interviewers could not simply ignore the treatment without the risk of upsetting respondents who expected the appointment to be kept" (page 212).

Survey organizations may also develop computerized case management systems to constrain interviewers to follow protocols. For the 2015 and 2016 Census Tests, partially in response to the findings from 2013 and 2014, the Census Bureau redesigned decennial test case management systems to give interviewers cases to be worked on a daily basis. This new system also asked interviewers what their schedule would be and took into account how many and at what times interviewers would be working. Such a design allowed interviewers no flexibility in whom they visited and attempted to constrain when they worked (Blumerman et al. 2015).

It may be that better case management system designs could further improve compliance. The field of decision support systems examines how to construct systems that enable informed decision making, including following requested actions. Much of this work aims at enabling medical professionals to implement treatments following evidence-based best practices. Kawamoto and colleagues (2005) summarize the lessons learned from this literature regarding approaches that ensure compliance.

Improved interviewer training may also increase compliance. Fowler and Mangione (1988) demonstrated that extended interviewer training could improve compliance with standardized interviewing practices. In the realm of nonresponse, Groves and McGonagle (2001) showed that training interviewers with methods for tailoring survey introductions could improve response rates. Indeed, in 2013 Census Test debriefings supervisors recommended self-assessments that would test interviewers' understanding of procedures while allowing supervisors to gauge interviewer knowledge. Interviewers also called for more training with additional role-playing situations.

Other approaches – such as incentives – may improve compliance. Tourangeau and colleagues (2012) found offering incentives to interviewers for every identified eligible person led to higher eligibility rates. However, Rosen et al. (2011) offered incentives to interviewers who completed cases with a low estimated propensity of response. They found incentives did not change interviewer behavior, and low propensity cases in an experimental group did not receive more effort than low propensity cases in a control group. Peytchev and colleagues (2010) offered interviewers incentives for converting cases with low response propensities but found completion rates between low propensity control and treatment cases to be the same, possibly due to high response rates for low propensity cases (i.e., 90.3 percent). Evidence that incentives prompt interviewers to follow field procedures is mixed, and we need more research to determine if and when such approaches can increase interviewer compliance, thereby improving our ability to test adaptive designs in CAPI environments.

Lastly, aligning performance standards with adaptive protocols could increase the likelihood that interviewers follow procedures. In widely cited research on what motivates individuals at their jobs, Hackman and Oldham (1976) argue workers need "knowledge of results" in the form of feedback that clearly aligns with the effectiveness of their job performance. Interviewers, too, may benefit from not only feedback on how well they are doing at following protocols, such as working prioritized cases, but also explanation as to how following such instructions ties to their overall job performance.

7.   Conclusion

In sum, we view a lack of in-person interviewer compliance as an obstacle to the implementation of adaptive designs, which hinders our ability to evaluate their successes in CAPI settings. When interviewers do not comply with data transmissions or working all prioritized cases each day, analysis is limited to nonrandom subsets of days or cases. Thus, we cannot say whether an observed difference between treatment and control is due to the adaptive design or the interviewer choosing when to comply – or not comply – with protocols.

Fortunately, lack of CAPI interviewer compliance is a problem with solutions and – as illustrated here – not a barrier to all in-person adaptive designs. Future adaptive design research needs to strengthen both the actions requested from interviewers and the ways in which these requests are delivered. Furthermore, the field would benefit from a study designed specifically to understand reasons for in-person interviewer compliance and noncompliance with a variety of protocols.

Adaptive designs that rely on in-person interviewers to implement protocols must consider the balance between flexibility and prescription. Survey methodologists and systems programmers should deliberately acknowledge when interviewer's local, accumulated knowledge outweighs the prescriptiveness that can be built into case management systems. These decisions likely depend on the survey. For data collections like the decennial census that, in 2010, hired over 500,000 employees across the United States, many of whom had limited interviewing experience, the balance might best tip towards prescription.

8. References

Biemer, P., P. Chen, and K. Wang. 2013. "Using Level-of-Effort Paradata in Non-Response Adjustments with Application to Field Surveys." *Journal of the Royal Statistical Society Series A* 176(1): 147–168. Doi: http://dx.doi.org/10.1111/j.1467-985X.2012.01058.x.

Blumerman, L., E. Moffett, M. Bentley, T. Boone, and M. Chapin. 2015. "2020 Census Operational Plan Overview and Operational Areas." Presentation to the Census Bureau's National Advisory Committee. October 8, 2015.

Campanelli, P., P. Sturgis, and S. Purdon. 1997. "*Can You Hear Me Knocking? An Investigation into the Impact of Interviewers on Survey Response Rates*. London, GB, National Centre for Social Research.

Coffey, S. 2013. "Implementing Adaptive Design for the National Survey of College Graduates." Article presented to FedCASIC, Suitland, Maryland, March 20, 2013.

Durrant, G. and F. Steele. 2009. "Multilevel Modelling of Refusal and Non-Contact in Household Surveys: Evidence from Six UK Government Surveys." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172(2): 361–381. Doi: http://dx.doi.org/10.1111/j.1467-985X.2008.00565.x.

Fowler, F. and T. Mangione. 1988. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Sage Publications: Newbury Park, CA.

Groves, R. and S. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169: 439–457. Doi: http://dx.doi.org/10.1111/j.1467-985X.2006.00423.x.

Groves, R. and K. McGonagle. 2001. "A Theory-Guided Interviewer Training Protocol Regarding Survey Participation." *Journal of Official Statistics* 17(2): 249–265.

Hackman, J. and G. Oldham. 1976. "Motivation through the Design of Work: Test of a Theory." *Organizational Behavior and Human Performance* 16: 250–279. Doi: http://dx.doi.org/10.1016/0030-5073(76)90016-7.

Kawamoto, K., C. Houlihan, E. Balas, and D. Lobach. 2005. "Improving Clinical Practice Using Clinical Decision Support Systems: A Systematic Review of Trials to Identify Features Critical to Success." *British Medical Journal* 330(7494): 765. Doi: http://dx.doi.org/10.1136/bmj.38398.500764.8F.

Kirgis, N. and J. Lepkowski. 2013. "Design and Management Strategies for Paradata-Driven Responsive design: Illustrations from the 2006–2010 National Survey of Family Growth." In *Improving Surveys with Paradata: Analytic Uses of Process Information*, edited by F. Kreuter, 121–144. Hoboken, New Jersey: John Wiley and Sons.

Kreuter, F., A. Mercer, and W. Hicks. 2014. "Increasing Fieldwork Efficiency through Prespecified Appointments." *Journal of Survey Statistics and Methodology* 2(2): 210–223. Doi: http://dx.doi.org/10.1093/jssam/smu005.

Laflamme, F. and H. St-Jean. 2011. "Highlights from the First Two Pilots of Responsive Collection Design for CATI Surveys." In Proceedings of the Joint Statistical Meetings, American Statistical Association. Available at: https://www.amstat.org/sections/srms/proceedings/y2011/Files/301087_66138.pdf (accessed March 2016).

Luiten, A. and B. Schouten. 2013. "Tailored Fieldwork Design to Increase Representative Household Survey Response: An Experiment in the Survey of Consumer Satisfaction." *Journal of the Royal Statistical Society, Series A* 176(1): 169–189. Doi: http://dx.doi.org/10.1111/j.1467-985X.2012.01080.x.

Morton-Williams, J. 1993. *Interviewer Approaches*. England: Dartmouth Publishing Company Limited.

O'Muircheartaigh, C. and P. Campanelli. 1999. "A Multilevel Exploration of the Role of Interviewers in Survey Non-Response." *Journal of the Royal Statistical Society, Series A* 162(3): 437–446. Doi: http://dx.doi.org/10.1111/1467-985X.00147.

Peytchev, A., S. Riley, J. Rosen, J. Murphy, and M. Lindblad. 2010. "Reduction of Nonrespose Bias in Surveys through Case Prioritization." *Survey Research Methods* 4(1): 21–29. Doi: http://dx.doi.org/10.18148/srm/2010.v4i1.3037.

Pickery, J. and G. Loosveldt. 2002. "A Multilevel Multinomial Analysis of Interviewer Effects on Various Components of Unit Nonresponse." *Quality and Quantity* 36(4): 427–437. Doi: http://dx.doi.org/10.1023/A:1020905911108.

Poehler, E., D. Cronkite, P. Sanchez, A. Wakim, G. Dusch, H. Walrath, R. King, and J. Jones. 2016. "2020 Research and Testing: 2014 Census Test Nonresponse Followup Panel Comparisons and Instrument Analysis." Washington, DC: U.S. Census Bureau.

Purdon, S., P. Campanelli, and P. Sturgis. 1999. "Interviewers Calling Strategies on Face-to-Face Interview Surveys." *Journal of Official Statistics* 15(2): 199–216.

Rosen, J., J. Murphy, A. Peytchev, S. Riley, and M. Lindblad. 2011. "The Effects of Differential Interviewer Incentives on a Field Data Collection Effort." *Field Methods* 23(1): 24–36. Doi: http://dx.doi.org/10.1177/1525822X10383390.

Schouten, B., M. Calinescu, and A. Luiten. 2013. "Optimizing Quality of Response through Adaptive Survey Designs." The Hague: Statistics Netherlands.

Tourangeau, R., F. Kreuter, and S. Eckman. 2012. "Motivated Underreporting in Screening Interviews." *Public Opinion Quarterly* 76(3): 453–469. Doi: http://dx.doi.org/10.2307/41684579.

Wagner, J. and K. Olson. 2011. "Where Do Interviewers Go When They Do What They Do? An Analysis of Interviewer Travel in Two Field Surveys." In Proceedings of the Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Miami, July 30-August 4, 2011.

Wagner, J. 2013a. "Adaptive Contact Strategies in Telephone and Face-to-Face Surveys." *Survey Research Methods* 7(1): 45–55. Doi: http://dx.doi.org/10.18148/srm/2013.v7i1.5037.

Wagner, J. 2013b. "Using Paradata-Driven Models to Improve Contact Rates." In *Improving Surveys with Paradata: Analytic Uses of Process Information*, edited by F. Kreuter, 145–170. Hoboken, New Jersey: John Wiley and Sons.

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): 477–499.

Walejko, G., A. Keller, G. Dusch, and P. Miller. 2014. "2020 Research and Testing: 2013 Census Test Assessment." Washington, DC: U.S. Census Bureau.