Interviewer Effects in a Small-Scale Experimental Research Survey of Demographic Questions

Elizabeth Nichols(1), Jennifer Hunter Childs(1) and Rolando Rodríguez(2)[1]
(1)Center for Survey Measurement, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233
(2) Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract
In November 2006, the U.S. Census Bureau fielded a RDD CATI survey called the Questionnaire Design and Experimental Research Survey, which had two experimental panels. The primary objective of this experiment was to test different questions that capture time spent at residences in order to measure within household coverage. Additionally, researchers embedded an experiment focusing on how to confirm age given a valid date of birth. A total of 1870 interviews were completed, approximately evenly split between the two panels. Ten different interviewers were assigned to each of the two panels. Because of a lack of resources, interviewers did not rotate through the panels. In an attempt to balance the panels, staff used amount of interviewing experience to make the two interviewing groups approximately equivalent.

We analyzed respondent burden, as measured by interview length, as one of the comparisons between the two experimental panels. This paper shows how results of regression models change when accounting for the small number of interviewers through fixed effects as compared with adding a random interviewer effect. This paper demonstrates the importance of controlling for interviewer effects, even in a relatively simple, small-scale experiment.

Key Words: Date of birth question; Age confirmation; Interviewer Random Effect

1. Introduction

In preparation for 2010 Census operations, in November 2006 the U.S. Census Bureau fielded a split-panel Random-Digit-Dial (RDD) Computer Assisted Telephone Interview (CATI) survey called the Questionnaire Design and Experimental Research Survey (QDERS). The purpose of fielding this survey was to compare two different questionnaire designs that were under consideration for the 2010 Census Coverage Measurement (CCM) operations. The within-household CCM operations measure how accurately the Census counted everyone in the U.S. That is, did the Census count everyone once and only once, and in the right place according to the Census Residence Rules (National Research Council, 2006)? The primary experiment in the 2006 QDERS was to discern which of two approaches to ascertain where each person should be counted in the census worked best. The secondary experiment involved how to best confirm age given a valid date of birth.

[1] *This report is released to inform interested parties of research and to encourage discussion. Any views expressed on methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau.*

A total of 1870 interviews were completed, approximately evenly split between the two QDERS panels. Ten different interviewers were assigned to the each of the two panels. Because of a lack of resources, interviewers did not rotate through the panels. In an attempt to balance the panels, staff used amount of interviewing experience, gender, and skill level to make the two interviewing groups approximately equivalent.

In this paper, we analyze respondent burden, as measured by interview duration, as one of the comparisons between the two experimental panels. This paper shows how results of regression models predicting interview length change when accounting for the small number of interviewers through fixed effects as compared with adding a random interviewer effect. This paper demonstrates the importance of controlling for interviewer effects, even in a relatively simple, small-scale experiment.

2. Methodology for Survey

The data for these experiments were collected in the 2006 QDERS, which is a split-panel controlled experiment developed by the Census Bureau's Statistical Research Division[2] for conducting methodological experiments offline from the agency's ongoing production surveys. Motivation for the 2006 QDERS came from the CCM operation for the 2010 Census. QDERS was used to test different CCM questionnaires prior to the 2010 Census. QDERS 2006 followed the CCM format of collecting an independent roster of current occupants of the address and then the demographics and residence information for each occupant.

QDERS 2006 was conducted between November 3 and November 21, 2006 using a RDD sample via CATI from one of the Census Bureau's centralized calling centers. The sample was nationally representative (excluding Alaska and Hawaii), with independent samples for each of the two panels.

Twenty interviewers were provided classroom training on only one of the two panels. All interviewers had previously been trained on how to conduct an RDD survey. Therefore, the classroom training focused only on the specific content of the QDERS instrument. We attempted to balance the two groups of ten interviewers in terms of interviewer characteristics such as tenure, experience with a similar instrument, skill level and gender.

There were two experiments in the 2006 QDERS for which we show how controlling for interviewer effects within the regression models shapes our conclusions.

Experiment 1: Residence Determination

The main purpose of QDERS 2006 was to compare two different approaches for collecting information to aid in residence determination: the "Cycle" questions and the "Dates" questions (Childs et al., 2007). Because respondents' own determination of usual residence differs in key ways from the Census Bureau's definition (Gerber, 1994), the Census Bureau implemented a series of questions in the CCM to assign the correct residence status according to the official rule. For people who lived or stayed only one

[2] The Statistical Research Division has since split into three centers. The authors now work in the Center for Survey Measurement and the Center for Statistical Research and Methodology.

place, the rules are fairly straightforward. However, complexity arises when assigning Census Day residence for people with more than one address.

- If a person has two addresses, the Cycle questions ask about how often the person goes back-and-forth between the places. These questions offer predefined patterns, for example cycling between places every week or every month, and ask the respondent to determine where the person spent most of the time during a specified time period (e.g., March and April). Notes are obtained for people with three or more addresses. Cycle questions were used to evaluate coverage in Census 2000. Cycle questions were used in a CAPI CCM instrument which was fielded approximately five months after the 2006 Census Test.
- The Dates questions involved collecting dates of stays for each address mentioned by respondents during the interview, instead of asking about patterns of going back-and-forth. The motivation behind this approach comes from the fact that the Cycle questions presume set patterns of living situations that may or may not reflect the realities of peoples' lives (Martin, 2004). The Dates approach does not make this assumption about regular patterns. The Dates questions were not used in 2000, but they were used in a PAPI CCM instrument which was fielded approximately 10 months after the 2006 Census Test.

The 2006 QDERS allowed us the opportunity to compare the "cycle" and "dates" approaches directly and measure interview duration without the confounds of mode and elapsed time between interviews that we had with the 2006 Census Test. We measure respondent burden by looking at the time taken in the interview for the two panels. Our hypothesis is that the panel that collects dates of stays for each address is no more burdensome and therefore will take no more time than the panel which asks the cycle questions.

Experiment 2: Confirming Age

The second experiment in QDERS investigates how to ask date of birth and age for a CAPI instrument. For decennial census operations, age data has typically been gathered as of Census Day (April 1st), so the U.S. Census Bureau can produce statistics on the age of the U.S. population at a very specific point in time. In past censuses, with few or no automated data collection instruments, respondents were asked to provide both age as of April 1st as well as date of birth. This increased the likelihood of gathering valid age data (Spencer and Perkins, 1998). As the Census Bureau has tested automated data collection instruments in the years leading up to the 2010 Census, we have acknowledged the technological advantages offered by automation. When a respondent provides a date of birth for either him or herself, or for another member of the household, the interviewer's computer automatically computes an age for the interviewer to verify with the respondent (see Martin, et al., 2007). Spencer and Perkins (1998) recommended accepting age calculated from birth date when possible. This question sequence also serves as an edit to pick up an error in the date of birth given by the respondent or a data entry error by the interviewer. For the CCM, date of birth will be gathered from the respondent, and then age will be confirmed as of Census Day, April 1, 2010.

The problem is that the CCM operation begins almost five months after Census Day and continues for up to seven months; thus, the age confirmed in this operation is an age on a specific day in the past. We suspect that confirming an age in the past may be difficult for respondents, and thus it might take the respondent longer to answer that survey question.

Our hypothesis is supported by past research using paper census questionnaires. In Census 2000, Carter and Brady (2002) found two situations where respondents had problems reporting age correctly on the census form – both the self-administered form and the interviewer-administered paper Nonresponse Followup (NRFU) form – which asked for age as of April 1st.[3] The most notable problem identified was that respondents misreported age when the person's birthday was after April 1st and the form was checked-in after the birthday (meaning the respondent likely completed the form after the person's birthday, which made the person's current age older than his or her age on April 1st). Forty percent of the people in this category over-reported their age (suggesting they reported their current age, and not their age as of April 1). The other problem occurred when the person's birthday was before April 1st, and their form was checked in before the birthday (meaning that the respondent completed the form before their birthday and before April 1st). In this situation, the person should report their age as of a date in the future. In 10.3 percent of cases, the person underreported his or her age, indicating that they were reporting their current, not future, age. This suggests that respondents did, indeed, have difficulty reporting age as of a date other than the current one.

Norris (2005) found indications of a similar problem in the 2004 Census Test. In that test, the NRFU questionnaire was automated using a hand-held computer. NRFU was fielded approximately one to three months after Census Day. Age was asked as of Census Day and then date of birth was asked. An edit on age was included in that instrument such that if the calculated Census Day age (based on the date of birth given) did not match the reported Census Day age, a confirmation question appeared:

> "For the Census, we need to record age as of April 1, 2004. Based on the date I just entered, your age was [fill: calculated age as of Census Day]. Earlier I recorded your age as [fill: reported Census Day age]. Which age is correct as of April 1, 2004?"

Norris (2005) found that for 17 percent of the people in one of the test sites and 11 percent of the people in the other test site, the edit check appeared and the respondent reported that the Census Day age that they previously reported was wrong. Although Norris (2005) does not speculate or investigate the reason so many people fell into this situation, we pose these possible reasons: (1) the interviewer asked the age question incorrectly (most likely excluding the Census Day reference date, thus eliciting the current age), (2) the age question was asked as worded on the screen but the respondent was confused and reported the current age instead, (3) the interviewer made a typing error while entering the age reported, (4) the respondent guessed at the Census Day age and when presented with the edit, realized the mistake, or (5) some combination of the above and the respondent was confused by the wording of the edit check which attempts to verify age as of a date in the past. Norris (2005) also found that for 5 percent of the people in each site, the edit check appeared and the respondent reported that the Census Day age that they provided was correct and the calculated age was wrong. In this situation, either the respondent initially gave the wrong date of birth, or the respondent was confused by the wording of the edit check, which, again, attempted to verify age as of a date in the past. Additionally, behavior coding results showed that respondents had to request clarification in seven percent of all administrations of the initial age question (Hunter and Landreth, 2006), also suggesting the age question was difficult for

[3] Because Census 2000 was predominantly a paper-only census, there was not the capacity to calculate and verify age as of Census Day given a date of birth.

respondents to answer. All of these data support the possibility that asking or confirming age in the past may be a difficult cognitive task for respondents.

In the 2006 QDERS, a split-panel experiment was conducted testing two different age-confirmation questions. If the respondent provided a date of birth, one panel calculated and confirmed current age, and the other panel calculated and confirmed age as of Census Day (April 1). After date of birth was collected, the interviewer asked one of the two following age confirmation questions:

> Panel A: "That would make NAME (fill: age as of today) years old. Is that correct?"
>
> Panel B: "For the Census Bureau, we need to record age as of April 1, 2006. So, just to confirm, NAME was (fill: age on 4/1/2006) years old on April 1, 2006?"

If the respondent said "no" to this question in either panel, meaning the calculated age was not correct, we then went to a screen to confirm the date of birth given. That screen asked:

> "I have recorded NAME's date of birth as (fill: MONTH, DAY, YEAR). Is that correct?"

If the respondent reports that the calculated age was not correct, but the date of birth was correct, we assume this implies the respondent was confused by the age-confirmation question.

We measure respondent burden by looking at the time taken in the demographic module of the two panels, where the only difference is how age was confirmed. Our hypothesis is that the panel that confirms current age is less burdensome and therefore will take less time than the panel which verifies Census Day age.

Summary of QDERS 2006 Experimental Design

There were two panels in the 2006 QDERS.

- Panel A contained the Dates questions and confirmed current age in the demographic section.
- Panel B contained the Cycle questions and confirmed age as of Census Day (April 1, 2006).

Both panels followed the same order of questions, first collecting a roster of current occupants, the demographics of each person, collection of addresses for each person, and then the residence determination questions (Cycle or Dates). Questions in the demographic section include date of birth, confirmation of age or collection of current age if date of birth was not given, sex, relationship, Hispanic origin and race. Interview duration in each section of the interview was recorded. So, although the two experiments were not crossed, it is possible to independently measure interview duration in those sections of the interview.

3. Methodology for the Analysis

When we discuss the "effect of interviewer" we mean, in a broad sense, a change in the answers/outcomes of a respondent due (directly or indirectly) to the interviewer. See Belli (2010) and Groves, et al., (2004) for discussion of interviewer effects.

In this paper, our outcome of interest is interview burden (measured in interview duration or time), so we ask ourselves the question "How do interviewers change the interview

duration or time required of the respondents?" We must then determine how best to modify our statistical models to account for these effects.

We can view respondents as being "nested" within interviewers: each respondent is interviewed by only one interviewer, but each interviewer has many respondents. Given a statistical analysis of interview duration, we can ask how that effect differs when applied to each interviewer "nesting."

In the 2006 QDERS, there were 20 interviewers total, 10 per panel. For budget reasons the interviewers did not rotate panels. Each interviewer conducted between 49 and 131 interviews, with the median being 92. Fourteen of the 20 interviewers conducted 80 or more interviews.

We can look at the effect of these 20 interviewers on the QDERS data by looking at a simple regression of number of addresses collected on interview duration. Across both panels, the number of addresses collected in each interview ranged from one to five for this sample. Households with only one address were households where everyone lived or stayed only at the sampled address for the year.

Figure 1 shows the results of a regression of interview time or duration (on the log scale) on the number of addresses by panel, where we run a separate regression for each interviewer as well as on the data as a whole (the black line). The logical trend of increasing time for more addresses is seen within each nesting. What we also see is that the expected time taken for single-address respondents differs for each nesting, and that this time difference in maintained across many of the nestings. Given random assignment of respondents into nestings, this suggests the possibility that the interviewers themselves are affecting the baseline respondent time.

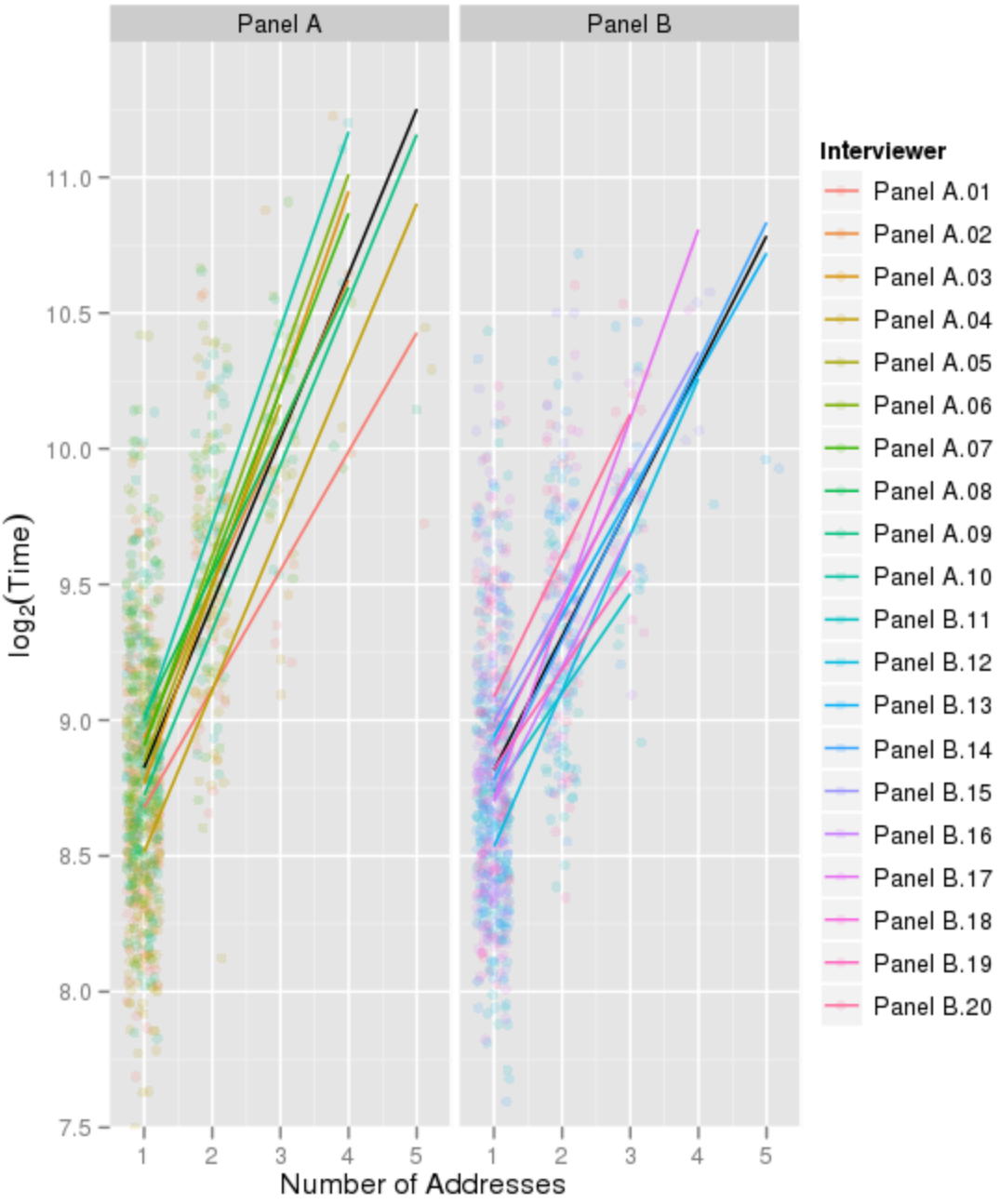

Figure 1: Regression of log of interview time by number of addresses for each interviewer by panel

Since we are trying to ascertain the difference in timing between the panels, we would be wise to include interviewer characteristics in our analyses. For regressions, we can accomplish this via fixed and/or random interviewer effects.

Definitions of these concepts differ across the literature, but in general a fixed effects model assumes the observations are uncorrelated while a random effects model can accommodate correlation in the observations. In these data, the observations occur in clusters (by interviewer) and it is likely that observations in a cluster are correlated. In

other words, probably there is correlation in the interview duration conducted by an interviewer. The random effects model accounts for this correlation while the fixed effects model does not.

Secondly, a random effect allows us to generalize results to a population, while fixed effects leave us with conclusions about only the sample at hand. (For a thorough overview of fixed and random effects and models that use them, see Raudenbush and Bryk, 2002).

Another reason for our interest in interviewer random effects is that the inclusion of a random effect will tend to increase standard errors associated with interviewer-level fixed effects. This makes a random-effects model less prone to Type I errors; that is, we guard against declaring the significance of non-significant fixed effects. This is especially important when we look at the effect of panel type on interview duration, as we can view panel type as a property of the interviewer, since interviewers were assigned to only one panel.

We used SAS' Proc Mixed for our modeling as it can account for both fixed and random effects.See http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mixed_sect006.htm for additional discussion of the modeling.

4. Limitations

The two experiments contained in QDERS 2006 (residence determination and age confirmation) were not crossed. At the point in the questionnaire when demographic questions were asked, the two panels were still identical; thus, we do not believe the residence determination experiment confounded interview time during the demographic section of the survey containing the age-confirmation questions. However, since the demographic questions came before the residence determination questions (dates or cycle), the interview time could have been affected by the age confirmation experiment. When we examine the interview duration for the age confirmation experiment, we use only the time spent in the demographic module. When we examine the interview time for the dates/cycle experiment, we use total interview time excluding the demographic module time.

The 2006 QDERS was fielded in November, which is later than most census operations would be conducted. This could affect the conclusions of our age confirmation experiment. There was more than a seven-month lag time between Census Day (April 1) and the interview day in 2006 QDERS. If birthdays were equally distributed throughout the year, more people would fall into the situation of being one year older than they were on April 1st in QDERS than in most census operations (which typically end by October). If the demographic module containing a verification of age in the past takes longer than the module with a verification of current age with a seven-month time lag, we cannot conclude that the same finding would exist with a shorter time lag.

This test did not investigate whether questions were accurately read. Confounding factors (such as the interviewer not reading the question clearly or a respondent who knew English as a second language) would influence any conclusions we might draw.

5. Response Rate Results

The total sample size for each QDERS panel was 2,996. Using the response rate calculation standards established by the American Association for Public Opinion Research (AAPOR, 2006), excluding cases of ineligibility and unknown eligibility, the response rates for Panels A and B were 60.77 percent and 55.92 percent, respectively.[4] These response rates were significantly different from one another ($p < .01$). The overall response rate was 58.4 percent as shown in Table 1.

For the response rate calculation, we considered interviews where the interviewer got to the end of the instrument for at least one person in the responding household. For the timing analysis presented in this paper we further limit the sample to those interviews which were fully complete for everyone in the household. During data analysis, we discovered that the two supervisors had conducted 11 interviews and that three interviewers conducted a total of four interviews from the wrong panel. We removed those cases in our analysis data set.[5] Finally, we removed one outlier interview in Panel B where eight addresses were collected. For the analysis presented in this paper, there are 972 households in Panel A which contained the Dates residence questions and confirmed current age in the demographic section and 878 households in Panel B which contained the Cycle residence questions and confirmed Census Day age in the demographic section.

Table 1: Sample Size and Response Rates

	2006 QDERS	Panel A Dates	Panel B Cycle
	Total	Current age	Census age
Total Sample Size	5,992	2,996	2,996
Response Rate	58.4%	60.77	55.92
Cases for Response Rate	1,870	982	888
Cases Eligible for Timing Analysis		972	878

Although we attempted to balance the two groups of ten interviewers between the panels, there may have been unanticipated, systematic, uncontrolled differences between the two groups. The significant difference in response rates between panels noted above suggest that Panel A might have included interviewers who were more skilled at gaining participation by respondents than those who worked on Panel B. Whether this skill of convincing respondents to participate in the survey translates into other differences between panels is unclear, but it is further reason for examining the interviewer effects in our modeling.

[4] Rates reflect the AAPOR RR6 definition (AAPOR, 2006).
[5] For incoming QDERS calls, the case was assigned to the next available interviewer without regard to the interviewer's assigned panel.

6. Interview Duration Results

Experiment 1: Residence Determination Timing Results

In Column A of Table 2, the regression model to predict the $\log_2$ of interview time (minus the time in the demographic section) contains the following predictors: number of addresses reported at a household level (No. of addresses), number of people in the household (Size of HH), two fixed effects for the interviewer (Interviewer experience with a similar survey and Interviewer skill level), respondent race, and panel (which can be considered an interviewer fixed effect as well since each interviewer only conducted one panel).[6] The model in Column A also contains a random effect of the interviewer on the intercept. Column B uses the exact same model except it does not include the random effect of the interviewer. The only difference between the two panels is which residence determination questions were asked – Cycle or Dates.

Table 2: Linear Regression Model of Log_2(Time in Seconds) to complete the Interview (minus time in the demographic module)

	Column A With Random Interviewer Effect		Column B Without Random Interviewer Effect	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Intercept	7.8870**	0.1074	7.8993**	0.03854
No. of addresses per HH	0.5491**	0.01557	0.5462**	0.01642
Size of HH	0.09989**	0.007089	0.09775**	0.007470
Interviewer experience with similar survey				
No	-0.05578	0.09427	-0.08403**	0.02461
Yes (control)	0		0	
Interviewer skill level				
Average	0.009014	0.1110	0.04240	0.03147
Good	-0.08431	0.09006	-0.06562**	0.02347
Excellent (control)	0		0	
Race of Respondent				
Black	0.1122**	0.03639	0.09941**	0.03831
Don't know/Refused	0.1561*	0.06308	0.1269	0.06634
Multiple races	0.1034	0.06619	0.1195	0.06969
Other	0.09631	0.06054	0.08635	0.06387
White (control)	0		0	
Panel				
B: Cycle	-0.1449	0.08197	-0.1398**	0.02202
A: Dates (Control)	0		0	
N=1850 * $p \leq .05$ ** $p < .01$				

In both models in Table 2, we see that the interview duration increases as more addresses are collected. We also see that the interview duration increases as more people are rostered. For both models, respondents who were Black or African American had a

[6] These covariates were included either because the two experimental panels differed in these characteristics or because the variable was expected to impact the amount of interview time.

longer interview than respondents who were White. However, our conclusions differ when we look at the two models for effect of the interviewer. In Column A of Table 2, the time it took to complete the interview did not differ between panels when we account for the interviewers by adding a random interviewer effect. Likewise, the fixed effects for the interviewers (experience with a similar survey and skill level) are not significant. However, if we do not add this random interviewer effect and assume that the interviewer effect is accounted for by controlling for the two fixed effects of experience with a similar survey and skill level, we find that the panel is significant as shown in Column B of Table 2.

If we had used the model in Column B and drawn our conclusions from that model, we would have rejected our hypothesis that the Dates approach took no more interview time than the Cycle approach. And it is true, for these 20 interviewers, the Dates approach interviews did take longer than the Cycle interviews, but if we consider our 20 interviewers as only a subset of all possible interviewers and add the random interviewer effect and consider the model in Column A, we would not reject our hypothesis. We would conclude that, in general, the Dates approach takes no longer than the Cycle approach.

Experiment 2: Demographic Module Timing Results

Now let's turn to the demographic module. The same model is used to predict the $\log_2$ of interview time in the demographic section, where the only difference in the two panels is how age is confirmed. Again, the model contains the following predictors: number of addresses reported at a household level (No. of addresses), number of people in the household (Size of HH), two fixed effects for the interviewer (Interviewer experience with a similar survey and Interviewer skill level), respondent race, and panel. The model in Column A also contains a random effect of the interviewer. Column B uses the same model except it does not include the random effect of the interviewer.

Table 3: Linear Regression Model of Log_2(Time in Seconds) to complete the Demographic Module

	Column A With Random Interviewer Effect		Column B Without Random Interviewer Effect	
	Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Intercept	4.8501**	0.1481	4.8608**	0.04594
No. of addresses per HH	0.03437	0.01797	0.03006	0.01957
Size of HH	0.4070**	0.008181	0.4001**	0.008905
Interviewer experience with similar survey				
No	-0.05816	0.1312	-0.06378*	0.02933
Yes (control)	0		0	
Interviewer skill level				
Average	0.1932	0.1542	0.2111**	0.03751
Good	-0.07077	0.1254	-0.05869*	0.02798
Excellent (control)	0		0	
Race of Respondent				
Black	0.09282*	0.04199	0.1177*	0.04567
Don't know/Refused	0.3182**	0.07280	0.2262**	0.07908
Multiple races	0.6325**	0.07638	0.6886**	0.08308
Other	0.1674*	0.06986	0.1336	0.07614
White (control)	0		0	
Panel				
B: Census Day	0.2864*	0.1140	0.2966**	0.02625
A: Current Age (control)	0		0	
N=1850				
* $p \leq .05$				
** $p < .01$				

The time it took to complete the demographic section differed between panels. For any given household size in this study, the demographic section in Panel B, which confirms age as of Census Day, took significantly longer to complete than did the demographic section in Panel A, which confirms current age (see Table 3). This finding holds true with or without adding the random interviewer effect. This means that for these 20 interviewers in particular, the interview with confirming an age as of a specific day in the past took longer and it means that this result is more legitimately generalized to other interviewers.

Other covariates were also significant predictors of time spent in the demographic section. In both models, households with more people in them led to longer demographic sections, as expected, since demographic data are collected for each person in the household. The race of the respondent was also a significant predictor of interview length given the other variables in the model. Respondents who did not identify a race or self-identified as a race other than White (except for "Other race" in the model with no random interviewer effect) took longer to complete the interview than did respondents who were White. This could have been due to taking additional time answering the race question itself, which was also in this section, and may not have been influenced by the age confirmation question.

The fixed effects for the interviewer (experience and skill) were significant in the model with no random interviewer effect, but were not significant in the model containing the random interviewer effect. Like Table 2, for these 20 interviewers these characteristics help predict interview time, but once we generalize to all interviewers; these covariates are no longer significant predictors of interview time for this survey.

The only difference between panels in the demographic section is how age is confirmed. The average time taken in the demographic section where the age was confirmed as of Census Day (i.e., age was confirmed as of a date in the past) was 1 minute and 32 seconds. The average time taken in the section that confirmed current age was 1 minute and 14 seconds. There was, on average, an 18-second difference between the two panels.[7] Although 18 seconds does not seem like much respondent burden, over thousands of interviews as in a census, the hours saved could be significant.

We do not believe the difference in time in this section was due to one panel having more instances of an incorrect date of birth or age (see Nichols et al., 2008, for full analysis of age confirmation experiment). For rostered people with a complete date of birth, respondents reported only five instances of the age being incorrect in the Dates panel. Only seven instances were reported in the Cycle panel. Beyond all other factors, the increase in the amount of time needed for the panel that confirms age as of Census Day suggests that interviewers and respondents were going more slowly through this section, perhaps due to the more cognitively difficult nature of verifying an age in the past. Verifying an age in the past could be a four-step process: 1) think about how old you are now, 2) think about when your birthday is, 3) decide whether your birthday was before or after April 1, and 4) decide how old you were on April 1. This process is repeated for everyone in the household. It makes sense that this four-step cognitive process would take longer than a single step of only having to remember one's current age. Additionally, if the interviewer has to assist the respondent by probing or offering clarification, this would also increase the time to administer the question.

7. Conclusions

When modeling results from an experimental survey collected by 20 interviewers, each assigned to one of two panels, results differed when we included a random interviewer effect and when we did not. With the random interviewer effect, we find that amount of respondent burden as measured by interview duration is no different if we use either the "Cycle" or the "Dates" residence determination questions. We also conclude that confirming an age that someone was on a specific day in the past takes more time than

[7] This is using pure means, not factoring in the covariates.

confirming current age. If we had not included the random interviewer effect, we would have come to the second conclusion, but our first conclusion would have been faulty because it was not generalizable to a different cast of interviewers.

The larger issue for other survey researchers to consider is when and how to know if you need to include fixed or random interviewer effects in your modeling. The answer depends upon the variability of your interviewers and your desire to generalize results to other potential interviewers. In a simple single-predictor regression at the respondent level, we know that if our data vary wildly about the regression line, the standard error of our estimate of the mean response at each value of the predictor will be large. Similarly in a multi-level model, if our regression parameters vary wildly across nestings (interviewers), our estimates of those parameters would also have a large standard error. Figure 1 illustrates this: it is clear that intercept values vary substantially across interviewers. Fixed effects of interviewer can help mitigate variability due to interviewers (mirroring the role of respondent-level predictors). Random effects of interviewer allow us to account for additional variability due to additional interviewer characteristics that we did not include in the model or that were not measured (mirroring the role of the error term). This allows for a better degree of generalization, as it forces more conservative statements concerning significance.

Acknowledgements

We thank Bob Fay who first encouraged us to investigate interviewer effects with these data. We thank Jennifer Rothgeb and Aref Dajani who helped in the preparation of the 2006 QDERS data. Elizabeth Murphy assisted with editing an earlier paper concerning age confirmation results. Nancy Bates, Mary Mulry, Kathleen Ashenfelter and Theresa DeMaio assisted with editing this paper for AAPOR.

References

American Association for Public Opinion Research (2006). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 4th edition. Lenexa, Kansas: AAPOR.

Belli, R. F. (2010). Interviewer Variance Between Event History Calendar and Conventional Questionnaire Interviews. *Public Opinion*, *74*(1), 140-153. doi: 10.1093/poq/nfp089.

Carter, N. and Brady, S. (2002). "Date of Reference for Age and Birth Date used by Respondents of Census 2000." Census 2000 Evaluation H.10, November 14, 2002.

Childs, J.H., Nichols, E., Dajani, A., and Rothgeb, J. (2007). A new approach to measuring residence status, *2007 Proceedings of the American Statistical Association, Survey Research Methods Section [CD-ROM]*, Alexandria, VA: American Statistical Association.

Gerber, E. (1994). "The Language of Residence: Respondent Understandings and Census Rules." Final Report of the Cognitive Study of Living Situations. U.S. Bureau of the Census, Center for Survey Methods Research.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., and Tourangeua, R. (2004). *Survey Methodology*. Hoboken, NJ: John Wiley & Sons, Inc.

Hunter, J. and Landreth, A. (2006). "Behavior Coding Analysis Report: Evaluating Bilingual Versions of the Non-Response Follow-Up (NRFU) for the 2004 Census Test." Study Series (Survey Methodology #2006-7). Available online at: http://www.census.gov/srd/papers/pdf/ssm2006-07.pdf.

Martin, E. (2004) "Some Evidence on Patterns of Reporting Multiple Residences and Cycles of Staying in Them in EFU," Internal Census Bureau Memorandum prepared June 1, 2004.

Martin, E., Childs, J.H., DeMaio, T., Hill, J., Reiser, C., Gerber, G., Styles, K., and Dillman, D. (2007). *Guidelines for Designing Questionnaires for Administration in Different Modes.* U.S. Census Bureau, Washington, DC 20233.

National Research Council (2006). *Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census.* Panel on Residence Rules in the Decennial Census. Daniel L. Cork and Paul R. Voss (Eds.) Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nichols, E., Childs, J. H., and Rodriguez, R. (2008). 2006 Questionnaire Design and Experimental Research Survey: Demographic Questions Analysis. *Statistical Research Division Research Report Series (Survey Methodology # 2008-1).* U.S. Census Bureau. Available online at http://www.census.gov/srd/www/byyear.html

Norris, S. (2005). "2004 Census Test Evaluation #4: Data Quality in the Use of the Hand Held Computer for Nonresponse Followup Enumeration During the 2004 Census Test." 2004 Census Test Memo #35. U.S. Census Bureau. September 14, 2005.

Raudenbush, S., and Bryk, A. (2002). *Hierarchical Linear Models –Applications and Data Analysis Methods.* Second edition. Thousand Oaks, CA: Sage Publications, Inc.

Spencer, G., and Perkins, Jr., R. C. (1998). "The Effect of Different Question Formats on Age and Birthdate Statistics From the 1996 National Content Survey." Internal Census Bureau report.