Adaptive Intervention Methodology for Reduction of Respondent Contact Burden in the American Community Survey

Robert Ashmead[1], Eric Slud[1,2], and Todd Hughes[3]

The notion of respondent contact burden in sample surveys is defined, and a multi-stage process to develop policies for curtailing nonresponse follow-up is described with the goal of reducing this burden on prospective survey respondents. The method depends on contact history paradata containing information about contact attempts both for respondents and for sampled nonrespondents. By analysis of past data, policies to stop case follow-up based on control variables measured in paradata can be developed by calculating propensities to respond for paradata-defined subgroups of sampled cases. Competing policies can be assessed by comparing outcomes (lost interviews, numbers of contacts, patterns of reluctant participation, or refusal to participate) as if these stopping policies had been followed in past data. Finally, embedded survey experiments may be used to assess contact-burden reduction policies when these are implemented in the field. The multi-stage method described here abstracts the stages followed in a series of research studies aimed at reducing contact burden in the Computer Assisted Telephone Interview (CATI) and Computer Assisted Personal Interview (CAPI) modes of the American Community Survey (ACS), which culminated in implementation of policy changes in the ACS.

Key words: Contact History Instrument; nonresponse follow-up; respondent burden; paradata.

1. Introduction

1.1. General Background on Respondent Contact Burden

It is in the nature of household sample surveys that by responding, a person gives up a certain amount of time, effort, and possibly privacy. Survey researchers refer to this cost as respondent burden or response burden, and they associate increased respondent burden with diminished data quality and response rates (Bradburn 1978; Sharp and Frankel 1983)

[1] U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC, 20233 U.S.A. Emails: eric.v.slud@census.gov, robert.douglas.ashmead@census.gov
[2] Mathematics Department, Kirwan Hall Room 2314, University of Maryland, College Park 20742
[3] UCLA Center for Health Policy Research, 10960 Wilshire Blvd, Suite 1550, Los Angeles, CA 90024, U.S.A. Email: toddhughes@ucla.edu

both within the given survey and in future surveys. In particular, a respondent who feels burdened by answering one survey might be less likely to accept the next request to participate in a survey. Many authors argue that behavior is primarily affected by the *perceived* respondent burden (Bradburn 1978; Fricker et al. 2014), which filters burden through the respondent's experience. For a recent summary of this literature, see the Eurostat document Hedlin et al. (2005).

In most literature on household surveys, 'respondent burden' refers to the survey instrument or interview process itself (length, difficulty or sensitivity of questions, etc.). In his classic paper on respondent burden, Bradburn (1978) restricted himself to the four general headings, 1) length of interview; 2) amount of effort required by respondent; 3) amount of stress on respondent; and 4) frequency with which respondent is interviewed. Fricker et al. (2014) develop a structural model of respondent burden in terms of three inputs: motivation, task difficulty, and recruitment effort. In this model, difficulty and recruitment are treated as 'related to survey characteristics', with recruitment effort a latent construct with the indicators of *frequent contact*, *frequent personal visit*, and *converted refusal*. The second of these indicators is defined by whether the interviewer visited the respondent more than the median number of visits, so that this and the converted-refusal indicator include survey recruitment preceding the respondent's exposure to the survey instrument. However, in the other elements of the Fricker et al. (2014) model and in almost all other literature not associated with the American Community Survey (ACS), burden was viewed as an aspect of the respondent answering the survey itself, not of the contacts leading to response. Many authors dating back to Bradburn (1978) mentioned the relevance of cumulative burden from all other survey solicitations to a respondent's decision whether or not to participate in a given survey.

In this article, we focus attention on measurement and reduction of what we term the *respondent contact burden* imposed on a potential respondent by the cumulative efforts of survey personnel to make contact with that person or household in any mode, prior to the actual completion of the survey. This can take the form of repeated mailings, telephone calls, personal visits including quick assessments from a car of whether anyone was home or leaving materials at the doorstep, and so on. Especially in surveys where persistent attempts are made to secure interviews in the field, contact burden may be viewed as a social cost that should be mitigated as far as possible without compromising the validity of the survey results. Reduction of contact burden is especially important for large national surveys like ACS which require continuing public acceptance and trust.

An essential part of measuring respondent contact burden and designing nonresponse follow-up strategies is the collection of survey paradata. Paradata might simply be the tallied numbers of visits or calls on each case, but systematic reporting of contact histories by field interviewers has recently become common practice, as formalized in systems such as the Contact History Instrument (Dyer 2004). In lieu of a formal instrument, interviewer observations might be used to summarize or replace paradata (Groves et al. 2007), and characteristics of respondents from historical data (Luiten and Schouten 2013) could serve a related purpose. Groves and Heeringa (2006) are frequently cited as early advocates for the use of paradata in 'responsive' survey design. Many papers (Bates et al. 2010; Maitland et al. 2009; Bates et al. 2008; Olson and Groves 2012) propose to use paradata from contact histories for predicting survey-response propensities. A few authors focus on

level-of-effort paradata (Slud 1998 and 1999; Biemer et al. 2013) and the search for proxy measures (Kreuter et al. 2010) that might predict response rates but not outcomes. Although some of these papers, particularly Bates et al. (2010), emphasize the measurement aspects and data quality of contact histories, there has been little work specifically on the measurement of contact burden with a view to designing interventions to mitigate it.

So-called adaptive interventions or adaptive study designs arise in various branches of statistics, from randomized controlled trials in biomedical statistics (Bothwell and Podolsky 2016) to sample surveys. The adaptive design generally takes the form of some scheduled change in the study protocol as a contingent result of interim outcome data. Because these "interventions" enter *after* the randomization stage, which in the survey context consists in choosing the probability sample, great care must be taken to ensure that the interim outcomes which influence the conduct of the study affect the desired study data as little as possible. In conducting sample surveys, the concern is that altering the data-collection (usually the handling of sampled but not yet responding units at the time of intervention) not affect the quality or values of the data yet to be collected. Data quality, which might be harmed by overly persistent contact strategies, is sometimes assessed within ACS research by the level of item nonresponse, but is generally difficult to measure without an independent audit of the correctness of survey responses.

The most common type of intervention in adaptive study designs is the use of interim outcomes to determine whether follow-up of a not-yet-responding study participant, or the study as a whole, should be terminated. Such interventions are generally rule-based, depending on the values of certain *control variables* up to the interim times when follow-up might be terminated, and the rules are specified as part of the study design before randomization. The control variables governing curtailed follow-up might be summary statistics from the responses collected so far in the entire study, e.g., the total number of sampled survey units who have responded within a geographical area, or might be contact-history variables or model-based predictions of the probability of response, based on case-history data specific to each observational unit whose follow-up might be curtailed. Examples of stopping rules at the whole-survey level include the idea to stop based on levels of the 'R-indicator' as a measure of 'representativeness' (Schouten et al. 2009) or of an estimated probability that a key survey outcome would change more than a specified amount (Wagner and Raghunathan 2010). By contrast, our discussion in this article of adaptive designs supported by contact history paradata is restricted solely to nonresponse-follow-up policies in sample surveys, applied at the level of the individual case.

Our discussion of contact burden concentrates heavily on research done for the ACS. We recognize that ACS is far from typical in the contact burden it imposes. Because of the mandatory nature of the ACS, and the Census Bureau's determination to maintain survey quality through exceptionally high overall response rates (approximately 97% survey-weighted response rate among eligible housing units, over the years 2005–2014), telephone and personal interviewers in the ACS are persistent in their follow-up with sampled housing units. This persistence is extremely successful at securing interviews, even after many prior contact attempts, but may cause persons in sampled units to feel unduly burdened. The research methods developed at the Census Bureau to mitigate this type of burden with minimal adverse impacts, such as lost interviews, are part of a broad effort to reduce the burden experienced by persons sampled in the ACS.

1.2. *Multi-Stage Process for Developing and Testing Contact Policies*

The theme of this article is that adaptive intervention policies for curtailing burdensome contact follow-up can be developed and tested in a multi-stage process, based on the regular collection of contact history data. Table 1 gives a summary of this multi-stage process (A)–(E) that can be followed closely for repeatedly administered surveys.

The first stage is to identify control parameters (such as maximum number and duration and type of admissible contact attempts) and devise intervention policies in terms of them, subject to administrative or logistical constraints. In a truly adaptive setting, these control parameters would include measures of propensity to respond, developed through models of response and time-dependent covariates fitted to earlier contact history data. Next, relevant outcome measures from existing data sources are specified, such as interview case completion rate and number of contacts and perhaps indicators of reluctance such as refusals by potential respondents, but also some measures of data quality such as item nonresponse or the magnitude of change in key survey estimates from those under the default intervention policy. The third stage consists of a retrospective analysis of contact history data to generate descriptive hypothetical data concerning outcomes that would have been realized had each of the intervention policies been followed. Based on this analysis, a policy is chosen for implementation based on a cost-benefit comparison or trade-off of expected outcome metrics under the policies and on the identification of constraints on allowed values for those outcome measurements. Such constraints would almost always include bounds on the changes in key survey estimates under the changed policies versus the previous policy. The new policies that meet administrative constraints and have admissible cost-benefit profiles are the ones that can actually be chosen. Lastly, the chosen policy is implemented under field conditions that are as close as possible to controlled experimental conditions for comparison with the regime under the earlier policy. Data are collected on contact histories and related outcome measures and assessed for acceptability and possible policy improvements.

The definition of 'admissible' policies requires some clarification. The general meaning of admissibility is that an allowable policy must not be dominated with respect to all costs and benefits by any other feasible policy or probability-weighted combination of policies. In game theory, economics, or statistical decision theory (Ferguson 1967), *multiple* outcome measures (all measured on a scale in which larger signed values are better) together constitute an outcome vector for each feasible policy, and all probability-weighted (i.e., convex) combinations of the feasible outcomes sweep out a convex multidimensional region. The admissible policies are those for which the outcome vector lies on the lower boundary or envelope of this feasible outcome region.

The objective of this article is to show these stages of development of adaptive designs in practice, along with the remaining methodological developments needed to make the design work well, with some focus on limitations of this process as applied so far.

2. ACS Experiments and Analyses on Contact Burden

As part of a broader ACS research effort on the reduction of burden perceived by respondents and potential respondents, the multi-stage process (A)–(E) has been followed

Table 1. Multi-stage development of adaptive policies to reduce burden of nonresponse follow-up.

Stage	Description	Examples
A. Identify control parameters and scope of policies for intervention	Data-collection control parameters that can be manipulated to reduce contact burden	Maximum number, type and duration of contact attempts
B. Specify policies and outcome metrics for evaluating them	Policies controlled and outcomes measured from existing data sources	Interview completion rate, number of contacts, cumulative sum of contacts weighted by burden
C. Analyze retrospective data under intervention policies	Using existing data, calculate outcome metrics as if new policies had been followed	A series of complete survey cycles of contact histories and outcomes
D. Choose a policy for implementation	Compare metric changes under competing policies, identify constraints on outcomes, and choose an admissible policy	Trade-off between number of contacts versus number of lost interviews or other quality measures
E. Assess new policy for acceptable outcomes	Under controlled field conditions, confirm outcome measures fall in desired range, compare operational alternatives	Randomized design elements, in controlled experiment comparing new policy outcomes versus outcomes from earlier period

in research on two separate modes of attempted contact, the Computer Assisted Telephone Interview (CATI) mode and the Computer Assisted Personal Interview (CAPI).

ACS data collection consists of three months of effort for each monthly panel of sampled households (U.S. Census Bureau 2014). In the first month, up to five pieces of mail are sent to sampled households in an effort to obtain an internet or mail self-response. Nonresponse follow-up proceeds to a second month for almost all 'mailable' ACS-sampled housing unit addresses with at least one known telephone number. The CATI operation schedules repeated call attempts terminating either after an interview; after a certain maximum number of call attempts; after reaching control thresholds of numbers of refusals, hang-ups, or unproductive calls; or at the end of the second month. A fraction of about 1/3 of the sampled housing units that do not provide a sufficiently complete CATI interview are sub-sampled into the personal-interviewer CAPI phase of data-collection. In addition, a higher proportion (closer to 2/3) of all national non-mailable sampled addresses are directly sampled into CAPI. The CAPI phase consists of trained Field Representatives (FRs) attempting to secure interviews by personal visits to household or, where possible, by further telephone contact attempts.

Contact history data for CATI are generated through the central calling operations under a system called WebCATI. CAPI field operations are assigned and managed centrally through a Case Management System and tracked and reported through a Unified Tracking System (UTS) based primarily on FR self-reports of case contact attempt histories through the Contact History Instrument (CHI), a centralized online system into which FRs enter details of their work on each case. CHI is described in Dyer (2004) and its design is further discussed by Bates et al. (2010). Although not designed for the purpose of measuring contact burden (Dyer 2004), CHI has been re-designed at least once with that in mind (Virgile 2015). These systems, WebCATI and CHI, may be the only ongoing real-time paradata systems in the U.S. containing longitudinal information on contacts and respondent reluctance for sampled households who choose *not* to respond to a survey.

CATI and CAPI modes – both in ACS and more generally – have unique and distinctly different characteristics. As a result, the control parameters, constraints, potential policies, and metrics that might be used as part of the multi-stage process (A)−(E) may be quite different depending on mode. Still, we argue that the same process can be used in both cases. The ACS CATI and CAPI research on the reduction of respondent burden is documented in a series of reports (Table 2) which collectively represent the multi-stage process (A)−(E). Some of the research documents span multiple stages. In the following sections we summarize the findings from ACS research on CATI and CAPI contact burden. Table 2 is designed to help the reader match the stage of the multi-stage process (A)−(E) to the references.

2.1. Research on CATI Burden

In 2012, at the beginning of this line of research, CATI follow-up on a case would stop after the earliest call at which the cumulative total of refusals was at least 2, the total of hang-ups reached 3, the number of 'unproductive' (i.e., unanswered) calls reached 20, or the total number of calls was 25. The Census Bureau hoped to identify strategies that would reduce respondent burden from CATI follow-up with small effects on response

Table 2.   Research references on CATI and CAPI burden, by stage.

Stage	CATI	CAPI
A. Identify control parameters and scope of policies	(Zelenak and Davis 2013)	(Zelenak 2014); (Griffin and Nelson 2014)
B. Specify policies and outcome metrics	(Griffin and Hughes 2013); (Slud and Erdman 2013)	(Griffin 2014); (Griffin and Nelson 2014)
C. Analyze retrospective data under interventions	(Griffin and Hughes 2013); (Slud and Erdman 2013)	(Griffin et al. 2015)
D. Choose a policy for implementation	(Griffin and Hughes 2013)	(Griffin et al. 2015)
E. Assess new policy for acceptable outcomes	(Griffin 2013); (Griffin 2014)	(Hughes et al. 2016)

rates. The data analyzed in this research consisted of preprocessed CATI status and history (WebCATI) files and CAPI history (CHI) files covering the national ACS monthly samples for which contacts began during the period June 2011 through Feb. 2012. The data set consisted of 1,097,985 housing-unit records (cases), of which 307,054 were only in CAPI, 600,203 only in CATI, and 190,728 in both.

Zelenak and Davis (2013) summarized ACS CATI case history data, through cross-tabulations of case outcomes (under both CATI and CAPI) with configurations of *control parameters*, by which we mean the numbers of call attempts, of hang-ups, of contacts with members of the sampled household or some other person, and cumulative counts of several categories of expressed reluctance or outright refusal to respond. The main results were tabulations of CATI cases according to the counts of calls, hang-ups, and refusals and the interview status at CATI termination, and the case distributions of final outcome types by total number of calls and of total number of calls by final outcome.

The CATI contact history data were analyzed further in Griffin and Hughes (2013) to suggest specific changes to the rules for termination of CATI case follow-up in terms of the control parameters. It was known to be technologically feasible to manage CATI operations so as to terminate case follow-up when different parameter thresholds were reached. More complicated or contingent interventions in CATI case follow-up were deemed infeasible because they would have required developing and testing new functionality for CATI software. As a result, Griffin and Hughes (2013) proposed 14 (centralized, automatic) termination policies involving reduction of the refusal maximum to 1, the hang-up maximum to 2, the maximum number of unproductive calls to 15 or 12, or the maximum number of calls to 20 or 15. They chose outcome variables from the WebCATI contact histories: total CATI numbers of contact attempts and CAPI workload (the 1/3 of eligible housing-units in CATI not providing CATI interviews that would be subsampled and passed on to CAPI nonresponse follow-up); CATI calls resulting in contacts; a weighted score for calls that combined the cumulative counts of previous refusals, hang-ups, requests to call back; and CATI and CAPI unweighted response rates.

The report then reanalyzed the CATI data according to the outcome metrics that would have been realized under the specified alternative policies.

As a more methodologically oriented complement to the Griffin and Hughes (2013) report, Slud and Erdman (2013) analyzed the same CATI data by calculating the discrete hazard functions associated with interview completion under a time-dependent state measurement defined by Calls, Hang-ups and Refusals. This *discrete hazard function* applied to the case group G at call number k is defined (Klein and Moeschberger 2003) as the proportion of cases in group G followed for k or more calls that terminate in an interview at the k'th call. This function of k may be viewed as the interview yield rate within group G, specific to the number k of calls. As Slud and Erdman (2013) emphasized, this is a conditional response propensity that can be used to suggest adaptive case-follow-up policies for different case groups defined from paradata observable up to each specified number of calls. While Griffin and Hughes (2013, Table 8, p. 10) compared proposed new policies different from the earlier one by calculating ratios of cost saving, calls eliminated, or post-resistance calls eliminated per lost interview, trading off favorable cost and contact outcomes against the unfavorable lost-interview outcomes, Slud and Erdman proposed more complicated and less feasible policies which would extend follow-up only for those groups where the discrete hazard (conditional call-specific interview-yield function) would exceed some threshold. The value of policies based on a decomposition into groups G by contact history depends on verifying that a relatively small set of groups succeeds in strongly separating the group-and-call-specific interview yield. Slud and Erdman (2013, Fig. 1, p. 6) showed that this was indeed possible. The most striking finding was that households that at any stage requested a call back had a strikingly higher propensity to complete an interview, almost regardless of other aspects of their contact history.

As part of the research leading up to the Griffin and Hughes (2013) report, it was confirmed that the ACS centralized call management system could easily be programmed to accommodate changed maximum hang-up and productive- and total- call parameters but not to curtail follow-up based on more complicated combinations of these or other case-specific contact-history variables such as those proposed by Slud and Erdman (2013). Based on the Griffin and Hughes (2013) analysis, the ACS Office determined to change the parameters ending CATI case follow-up beginning with the March 2013 ACS panel (i.e., beginning with CATI operations in April 2013) by reducing the maximum numbers of nonproductive and total calls. Griffin (2013) assessed the changes in experienced outcomes for CATI operations in the months April–May 2013 by comparison with outcomes for the January–February 2013 panels. The field implementation of the CATI call parameters provided at best a very imperfect observational setting, partly because the initial mailings in the January panel collide with the Christmas and New Year holiday period and because January and February tend to have seasonally high mail response rates, but also because new recording technology and telephone scripts were initiated in March 2013 CATI operations. The outcome metrics were observed as part of April and May CATI operations and can be compared either with those from February-March CATI (i.e., those from the January-February panels) or with the ACS June 2011-Feb. 2012 data analyzed in Griffin and Hughes (2013). However, any observed changes could not be ascribed solely to the new CATI call parameter changes. In fact, Griffin (2013) found all of the outcomes (productive and unproductive calls, contacts, interviews, subsampled CAPI

workload) to be in the expected range based on the earlier data analysis, but these comparisons could not be presented with statistical precision due to the uncontrolled (because universal) nature of the implementation.

Griffin and Hughes (2013) broke new ground in quantifying cumulative contact burden by defining a simple *contact score* to summarize a case contact history in terms of severity of burden. Their objective was to provide a single descriptive statistic of burden in terms of which to measure differences between the outcomes of their 14 proposed call-stopping policies versus the previous (up to 2012) call-stopping rules. They displayed these changes in their Table 5, p. 8. This burden score was defined as a case-specific sum over all call attempts, of a score defined at each call attempt as follows. If the case had never resulted in a call-back request at previous calls, the score was 0 if there had been no previous hang-ups or refusals, 1 if there had been one immediate hang-up (but no refusals), 2 if there had been in previous calls either 2 hang-ups or 1 refusal (but not both), and 5 if there had been previously at least one hangup and one refusal. If there had been a previous call-back request by the potential respondent, then the scores under these four different prior-history conditions were defined respectively as 0.5, 0.5, 1, and 3. While these scores were based on introspection by Griffin and Hughes (2013) rather than any research on actual perceived burden by potential respondents, they represent the first example we know of where contact history data are codified into a burden outcome at case level, viewed as a quantity to be minimized.

2.2. *Research on CAPI Burden*

In the past, FRs in ACS relied on their judgment and feedback from their supervisors to determine when to terminate attempted contacts with a sampled case under the CAPI mode of data collection. The ACS Office approved a series of research projects to explore whether contact burden might be lessened by systematic policies for terminating CAPI case follow-up without large losses in response rates. The resulting sequence of research reports, culminating in an August 2015 Pilot study on modified CAPI follow-up control processes, followed the same multi-stage process (A)–(E). First, Zelenak (2014) provided baseline descriptions of CAPI workloads and outcomes from the June 2011–Feb. 2012 dataset previously explored for CATI histories. She began by cross-classifying CATI-nonresponder cases subsampled into CAPI according to their CATI outcomes, and gave descriptive summaries of CAPI workloads and of CAPI cases classified by outcomes and by numbers of contact attempts. From the vantage point of contact burden reduction, the most interesting finding (see Table 8 on page 16 of the report) was the relatively small number of mean and median contact attempts (1 to 3) for cases ending as Interviews or Ineligibles and the much larger (3 to 8) mean and median numbers for cases with ultimate outcomes of Refusals or other Noninterviews.

Next, Griffin and Nelson (2014) created baseline summaries of many different outcomes from the CAPI contact history dataset on January through December 2012 operations. They distinguished response rates from contact rates, cooperation rates and refusal rates, and tabulated outcomes according to whether sampled housing units turned out to be occupied. (This last issue is particular to ACS, where 'vacant' unit interviews related to housing type are collected when FRs can find corroboration of vacancy.) They also measured differences in completeness among interviews based on contact attempts

and levels of cooperation, and delved deeper into strategic patterns followed by FRs, in classifying CAPI workloads according to sequential patterns of telephone and personal visits. Finally, they drew attention to known errors in CHI self-reporting by FRs (Bates et al. 2008) with respect to reluctance displayed by potential respondents.

This research laid the groundwork for the definition of outcome metrics and policy alternatives for curtailing nonresponse follow-up. With the goal of reducing contact burden, Griffin (2014) identified possible changes to CAPI data collection rules, ruling out large proposed system or instrument changes. In addition, it was thought advisable to minimize the impact of proposed changes on the policies and incentives governing FRs. A team of ACS researchers familiar with the CAPI baseline research collaborated in proposing rule changes that might lessen burden without dramatically worsening response rates. Griffin (2014) collected the ideas and distilled them into policy proposals to stop CAPI case follow-up after reaching thresholds of numbers and types of contacts, of estimates of perceived contact burden, or of estimated response propensity. Refinements of these ideas allowing thresholds to depend on CATI outcomes or geography were also considered. This CAPI research generated a list of 27 policies grouped according to which control variables measuring CHI states would trigger termination of CAPI follow-up attempts on a case.

Griffin et al. (2015) analyzed data on the 2012 ACS case histories to tabulate the outcome metrics that would have resulted from implementing each of the 27 CAPI stopping rules. The rules were compared with respect to a large set of different outcomes: contacts, broken down as personal visits or telephone contacts; interviews and non-interviews according to whether or not contact was made with the sampled household; contacts at which either expressed reluctance or 'firm reluctance' was recorded in CHI; contacts after either reluctance or firm reluctance was expressed; and a cumulative burden score for CAPI analogous to the summary contact score previously devised for CATI by Griffin and Hughes (2013). The burden score combined an initial score for the total burden incurred in mail and CATI modes, with an incremental numeric value from 1 to 15 for each CAPI contact attempt, based on an intuitive scoring of the relative burden of various types of contact coded in CHI. In this scoring, personal visits were assigned more burden points than telephone calls, contacts more than non-contacts, and additional points were added if a potential respondent indicated reluctance. The cumulative burden score served as the primary proxy for respondent perceived contact burden. All of the different outcomes analyzed were alternative measurements for workload/cost or nonresponse or perceived contact burden, and the approach was to balance increases in nonresponse due to implementing the new stopping policies versus reductions in contacts or contact burden. This balancing was done in the same sequence as for CATI policy changes, first through cost-benefit ratios (Table 15 of Griffin et al. 2015) measuring the reduction in cost or contacts per lost interview, or reduction in burden-score or highly-burdened cases per lost interview, and then (Figures 1−2 of the same report) through the nearness of the vector of three policy outcomes for each policy – mean contacts, mean burden score, and case non-interview rate – to the lower boundary of the convex combinations of all such policy outcome vectors. This lower boundary is understood as the result of cost-benefit trade-offs, providing the set of best policy outcomes that can be achieved by applying the different policies with various probabilities. A small set of competing policies were thus identified as most interesting, further reduced to a single recommended policy after clarifying the administrative

constraint that an overall CAPI nonresponse rate greater than about 0.085 could not be borne. The chosen policy, number 18 among those displayed, was a stopping rule to terminate CAPI contacts on cases that exceed the threshold of 40, a level roughly corresponding to the 90'th percentile of cumulative burden among all CAPI cases in 2012 ACS data.

With respect to lost interviews and reduced burden, the stopping policy chosen for implementation was the most aggressive of those near the lower envelope of the nonresponse, contact and burden score outcome combinations that met the upper-bound constraint of 0.085 on CAPI nonresponse (Griffin et al. 2015). Interestingly, the policy chosen was not one of the three competing policies defined in terms of thresholded response propensities, but rather was designed to achieve maximum reductions in the burden score. It resembled the CATI stopping-policy chosen for implementation in 2013 through its explicit definition in terms of parameters controlling contact burden. The avowed purpose of both research initiatives had been to find a stopping policy to achieve maximum reduction of contact burden subject to small and acceptable losses in interview completion rates. The CAPI policy of stopping according to a burden-score threshold was chosen for actual (pilot) implementation only after additional research (not published) verifying on the ACS 2012 data that this policy did not lead to unacceptably large CAPI nonresponse or small contact and burden reductions when tabulated by state and Survey Statistician Field Area (a set of 48 geographic organizational units for CAPI field work).

In order to test the effectiveness of the cumulative burden score stopping rule under field conditions, the Census Bureau conducted a field pilot in the CAPI operation of the ACS during August 2015 (Hughes et al. 2016). The pilot required that FRs transmit data from their laptops to a central management system (UTS) twice daily on days worked, at the beginning of the work-day before attempting contacts and after all of the day's attempts, so that burden scores could be updated correctly. The transmission prior to contact attempts updated the case burden scores with information from the day before and removed any cases with burden scores above the burden threshold from FRs' laptops. A secondary goal of the pilot was to learn whether survey outcomes would be affected by displaying current burden scores on FR laptops. The previous ACS protocol required FRs to transmit only once per day, at the end of their workday. The UTS system updated the cumulative case burden scores using entries from the FR in CHI. Therefore, the cumulative burden score calculation depended on accurate and timely FR entry of case-specific data. The CHI includes information on each contact attempt made to sample units as well as additional levels of effort not classified as contact attempts (e.g., locating the unit or geocoding).

The pilot was conducted in one-quarter of the field geographies for ACS interviewing. The remaining field geographies followed the standard field protocol. Within the Survey Statistician Field Areas (SSFAs) selected for inclusion in the pilot, individual Field Supervisor (FS) areas and all the FRs within that FS area (FSA) were assigned randomly to one of three experimental treatments:

- Treatment 1 (Control): burden score not displayed, cases not removed;
- Treatment 2: burden score displayed, cases removed; and
- Treatment 3: burden score not displayed, cases removed.

Overall, 4.5 percent and 4.1 percent of cases were pulled from Treatments 2 and 3 respectively as a consequence of reaching the cumulative burden score threshold. Only

Table 3.   2016 CAPI pilot estimated differences in outcomes per case.

	Trt 1	Trt 2−3	Difference	Percent difference	*p*-value
Contact attempts	3.90	3.67	0.234	6%	0.074
Contacts	1.00	0.94	0.062	6%	0.067
Contacts with firm reluctance	0.08	0.07	0.017	21%	0.032
Response rate	93.1%	91.8%	1.3%	1.4%	0.104

Source: American Community Survey paradata, 2015.

minimal differences between Treatment 2 and Treatment 3 were observed, and therefore Treatment 2 and 3 observations were pooled in order to increase the precision of estimates of differences with Treatment 1. Table 3 shows the 6% decrease in both contact attempts and contacts per case observed for Treatments 2−3 compared with Treatment 1, along with a 21% reduction in contacts with firm reluctance per case. The CAPI response rate for Treatment 1 was 93.1% compared to 91.8% for Treatments 2−3. The estimated difference was 1.3 percentage points [*p*-value = 0.104]. The estimated difference 1.3% was relatively small compared to the approximately 4% of cases that were stopped due to high burden score. This suggests that a large proportion of the stopped cases would not have resulted in complete interviews had FRs continued to make attempts.

In the pilot, the cumulative burden score stopping rule succeeded in reducing the percentage of cases with extremely high burden score. Treatments 2−3 had only 0.3 percent of cases with burden scores over 60 (and less than 0.1 percent with scores over 80), while control and Treatment 1 had over 2.0 percent of cases with burden scores over 60 (and 0.5 percent of cases with burden scores over 80). Treatments 2−3 had cases with burden scores as high as 60 or 80 only because of failures to comply with the beginning- and end-of-workday transmission protocol updating the burden score, together with multiple intra-day contact attempts by some FRs.

As a result of the Hughes et al. (2016) report, the Census Bureau decided to implement the cumulative burden score stopping rule nationwide beginning in June 2016. As adopted, the protocol displayed a burden category to FRs, but not the exact score.

2.3.   Before-and-After Assessment

At the end of the process of Table 1 for contact burden reduction, what can be said about the overall actual impact of the changes? In the ACS CATI and CAPI examples discussed at length above, the final step E was an assessment based on newly collected ACS data, and this can be enriched in each example with the retrospective data-analysis results done in step C to inform the choice of a contact-burden reduction policy.

As mentioned in the ACS reports cited in the paragraphs above, statistical inferences for the effects of the policy changes are generally not available because of the need to correct for period and regime differences, which could only be done through a model. For example, Table 4 shows the per-case number of calls (contact attempts) and CATI interview completion rate for CATI cases, based on the June 2011 – Feb. 2012 data analyzed in Zelenak and Davis (2013), Griffin and Hughes (2013), and Slud and Erdman (2013), and on the comparison in Griffin (2013) of CATI outcomes just before and after

Table 4.   CATI call and interview rates per case, before and after rule change.

Stopping rule	Time period	Attempts/case	Interviews/case
Old	June 2011-Feb. 2012	7.25	0.242
Old	Jan.-Feb. 2013	6.70	0.168
New*	June 2011-Feb. 2012	5.72	0.215
New	Mar.-Apr. 2013	5.16	0.150

Source: Zelenak and Davis (2013, Table 4); Griffin and Hughes (2013, Tables 1,4)
*Outcomes calculated for New rule as though followed in the older data set.

the implementation of the new stopping rule in March 2013 with reduced maximum numbers of nonproductive and total calls. Because of a decreasing trend over the past several years in CATI interview-completion rates, the attempt and interview rates are very different across the two time periods. However, the reductions in outcome measures are similar for both periods, in the range 21–23% for attempts per case and roughly 11% for reductions in interview completion rates.

In the progression to a new policy for terminating nonresponse follow-up by FRs in CAPI, Table 5 similarly shows performance through the August 2015 ACS pilot study described above and in Hughes et al. (2016). As in Table 4, the comparison is partly observational, with results displayed for the month before the pilot, July 2015, and for the "Treatment Group 0" which we use to denote the control group of ACS data collections in SSFAs not chosen for inclusion in the pilot. The SSFA choice was not random, and contrasts with that control group and with the ACS outcomes in the month before the pilot do not yield statistical inferences, only descriptive comparisons. The designation of treatment group in data collected in July 2015 arises because data from the Field Supervisory Units (FSAs) that were assigned to the separate treatments in the August pilots are tallied separately in July. The numbers of attempts per case are systematically different in 2015 from their level in 2012, a change that can largely be ascribed to a redesign of the CHI system implemented in January 2014 that explicitly aimed to reduce the potential for underreporting contact attempts. The overall pattern of Table 4 shows a

Table 5.   CAPI attempt, interview and burden per case, before and after rule change.

Rule/regime	Time period	Attempts/case	Interviews/case	VHBurd rate
Old	all 2012	2.94	0.950	0.050
Trtgp 0	July 2015	3.84	0.931	0.062
Trtgp 1	July 2015	3.89	0.931	0.060
Trtgp 2−3	July 2015	3.92	0.936	0.062
Trtgp 0	Aug. 2015	3.87	0.934	0.067
Trtgp 1	Aug. 2015	3.90	0.935	0.063
New*	all-2012	2.79	0.920	0.019
Trtgp 2−3	Aug. 2015	3.67	0.918	0.041

Source: Hughes et al. (2016, Tables 5.7, 5.15) and ACS paradata, 2015.
VHBurd is the case indicator of having a burden score above 45.
*Outcomes calculated for New rule as though followed in the older data set.

clear and similar decrease in attempts per case for the new versus old stopping regime, for each separate year of data collection. There was a clear decrease of roughly 2% in CAPI response rate, and a dramatic decrease in the incidence of cases with 'very high' burden score (i.e., above 45, which was the 95'th percentile level for burden score in the ACS 2012 paradata). The decrease in VHBurd from implementation of the new policy was very different in the 2012 data and the 2015 pilot because cases exceeding the threshold 40 defining the contact-attempt stopping rule could not in practice be put into effect instantly but waited for a further transmission from the FR, allowing further contact with the same case during the same day (or the next, in case of transmission delay). Thus, the VHBurd reduction achieved was much less than the possible reduction if case follow-up had been stopped immediately.

The contrast between treatment groups within the August 2015 pilot was based on randomization and therefore did yield meaningful variances and confidence intervals for contrasts between the Treatment group and the combination of Treatment groups 2–3. As described in Hughes et al. (2015), the reduction in attempts per case and CAPI response rate between Treatment Group 1 (which did not remove cases from the FR's case portfolio) and Groups 2–3 (which did) were found to be nearly significant with p-values .074 and .104. A similar permutational calculation for VHBurd reduction in the August pilot yields a highly significant result (p-value < 0.001).

3.   Limitations of Multistage Approach to Burden Reduction

The contact history databases we are aware of require FRs to enter paradata accurately. However, FRs may have competing incentives not to do so. FRs are evaluated (in part) on their completion rate, and in the context of the ACS CAPI pilot study, when a case is pulled in the future from a FR's workload for exceeding the burden score, the case may count against that FR's response rate. This may influence FRs not to record some contact attempts that result in high amounts of respondent contact burden. In the pilot study, there was little evidence that such behavior occurred; however, the removed cases did not count against the FRs during that month, and FRs were found generally not to have a good understanding of the burden score increments, knowledge that they might acquire over time. This raises the importance of heightened monitoring of field reporting behaviors after the full implementation of the new stopping rule. Additionally, the chosen measures of contact burden in studies conducted up to the present, such as the burden score increments used in the ACS CAPI stopping rules based on cumulative burden score, have never been validated against actual perceptions of contact burden by potential respondents. Cognitive research along this line is needed, specifically to validate the current score used in ACS or to suggest modifications.

Even when contact history data have been properly recorded and maintained, retrospective analysis of those data may not be a reliable guide to factors influencing response rates and other measures of survey quality when new procedures are implemented. For this reason, the effects of implementation should always be tested in the field, ideally in a randomized study permitting assessment of the relative effects of the new policy versus other less predictable changes in the survey environment. That is the motivation for including stage E in Table 1

Conducting randomized controlled experiments in a survey field setting can be difficult in part because of compliance failures and administrative procedures which either allow participants to know how their treatment group differs from others or which change the treatment for some participants after randomization. Analogous issues arise in randomized clinical studies where patients do not fully comply with treatment regimens, for example by failing to take pills on schedule, or where they or their doctors can infer their treatment assignment, or where there is some unanticipated treatment switching. All of these failures to follow the formal randomization protocol did occur in the 2015 ACS CAPI pilot of the cumulative burden score stopping rule. Among the limitations in the pilot were the reassignment of cases to FRs in different treatment groups, and transmission compliance was much lower than was needed to ensure that burden scores were daily entered accurately and cases were pulled in a timely manner. In addition, because cases were rarely supposed to reach the burden score threshold, the implementation of the stopping rule was not expected to cause drastic changes in metrics of contact or response. For these reasons, and also because the geographies chosen for the pilot were not a random sample of all ACS geographies, the results were not representative of the whole ACS. Thus the pilot study was partly observational rather than controlled.

4. Conclusions and Directions for Research

In this article, we proposed a multi-stage process for the development of adaptive policies to reduce the burden of nonresponse follow-up and described how it has been implemented in two response modes in the ACS. We argued that in a repeatedly administered survey like the ACS, which collects contact history paradata, this multi-stage process can be replicated to create successful burden reduction policies. We also described the cumulative burden scores that have been used by the ACS to study the contact-burden reduction of newly implemented CATI and CAPI contact-attempt stopping rules. In CAPI, a stopping rule for contact attempts based on reaching a certain threshold in the cumulative burden score was chosen for a pilot study and national implementation. The idea of burden scoring as a sum of increments with severity determined by contact history paradata is one that could be adopted and modified for other surveys as a method to measure contact burden. For survey organizations interested in monitoring and mitigating contact burden on potential respondents, it is essential first to develop a system to collect paradata concerning contact attempts. Then, after a period of quantifying baseline measures of contact burden and other process outcomes, the multi-stage process in Table 1 can be undertaken.

Additional iterations of the multi-stage process can be beneficial as conditions and resources change. As a sequel to the research described in Section 2.1, Mills (2016) analyzed CATI contact histories during the six-month period following a change in 2016 of telephone number sources from third-party vendors to the Census Bureau's Center for Administrative Records Research and Applications. Motivated by the dual goals of increasing the efficiency of CATI and decreasing respondent burden, Mills used CATI data to simulate policies altering the control parameters of total-workload, maximum-follow-up, and maximum-follow-up after contact. This work – analyzing new CATI data, choosing outcome measures reflecting cost and efficiency of producing completed

interviews in CATI, and proposing policies reducing case follow-up based on model-based scores for reliability of telephone numbers or on time to first contact – cuts across the stages (A)–(C) of Table 1.

In the experience of Hughes et al. (2016), the assessment under field conditions of implemented adaptive policies to curtail nonresponse follow-up required detailed attention to the impact on morale and incentives of field staff. Since the paradata on which stopping policies are based depend until now on FR self-reporting, the lack of research on the interaction between FR incentives and contact burden is a limitation on the effectiveness of field experiments. More broadly, either because of FR behavioral changes in adapting to changes in their compensation and incentives, or because policy changes induce changes in respondent propensities over time, conclusions from analysis of contact histories must be confirmed or re-evaluated after adaptive policies are implemented.

If survey organizations are seriously concerned about respondent contact burden, then the instruments used to measure it, such as CHI, should be redesigned. Additionally, policies should be evaluated from the perspective of the potential respondent. Though policies such as the cumulative burden score stopping rule are clearly successful at reducing burden according to its own metric, it is not yet clear what the effect is on the potential respondent's perception of burden.

5.   References

Bates, N., J. Dahlhamer, P. Phipps, A. Safir, and L. Tan. 2010. "Assessing Contact History Paradata Quality Across Several Federal Surveys." In JSM Proceedings, Survey Research Methods Section, American Statistical Association, Vancouver, BC, July 31-August 5, 2010. Alexandria, VA: American Statistical Association. 91–105. Available at: http://ww2.amstat.org/sections/SRMS/Proceedings/y2010/Files/306005_55654.pdf (accessed February 2017).

Bates, N., J. Dahlhamer, and E. Singer. 2008. "Privacy Concerns, Too Busy, or Just Not Interested: Using Doorstep Concerns to Predict Survey Nonresponse." *Journal of Official Statistics* 24: 591–612.

Biemer, P.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: 147–168. Doi: http://dx.doi.org/10.1111/j.1467-985X.2012.01058.x.

Bothwell, L. and S. Podolsky. 2016. "The Emergence of the Randomized, Controlled Trial." *New England Journal of Medicine* 375: 501–504. Doi http://dx.doi.org/10.1056/NEJMp1604635.

Bradburn, N. 1978. "Respondent Burden." In JSM Proceedings, Survey Research Methods Section, American Statistical Association, San Diego, California, August 14–17, 1978. Alexandria, VA: American Statistical Association. 35–40. Available at: http://ww2.amstat.org/sections/SRMS/Proceedings/papers/1978_007.pdf (accessed February 2017).

U.S. Census Bureau. 2014. Design and Methodology: American Community Survey. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html (accessed February 2017).

Dyer, W. 2004. "Contact History Instrument." In IBUC Proceedings, International Blaise Users Conference, Qubec, Canada, September 22–24, 2004. International Blaise Users Group. 35–61. Available at: www.blaiseusers.org/2004/papers/03.pdf (accessed February 2017).

Ferguson, T. 1967. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press.

Fricker, S., T. Yan, and S. Tsai. 2014. "Response Burden: What Predicts It and Who is Burdened Out?" In AAPOR Proceedings, American Association for Public Opinion Research, Anaheim, California, May 15–18, 2014. Oakbrook Terrace, IL: American Association for Public Opinion Research. 4568–4577. Available at: https://www.bls.gov/osmr/pdf/st140170.pdf (accessed February 2017).

Griffin, D. 2013. "Effect of Changing Call Parameters in the American Community Survey's Computer Assisted Telephone Interviewing Operation." American Community Survey Research and Evaluation Report Memorandum Series ACS13-RER-17. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Griffin_03.pdf (accessed February 2017).

Griffin, D. 2014. "Reducing Respondent Burden in the American Community Survey's Computer Assisted Personal Visit Interviewing Operation – Phase 2 Results." American Community Survey Research and Evaluation Report Memorandum Series ACS14-RER-07. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2014/acs/2014_Griffin_01.pdf (accessed February 2017).

Griffin, D. and T. Hughes. 2013. "Analysis of Alternative Call Parameters in the American Community Survey's Computer Assisted Telephone Interviewing." American Community Survey Research and Evaluation Report Memorandum Series ACS13-RER-11. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Griffin_02.pdf (accessed February 2017).

Griffin, D. and D. Nelson. 2014. "Reducing Respondent Burden in the American Community Survey's Computer Assisted Personal Visit Interviewing Operation – Phase 1 Results (Part 2)." American Community Survey Research and Evaluation Report Memorandum Series ACS14-RER-22. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2014/acs/2014_Griffin_02.pdf (accessed February 2017).

Griffin, D., E. Slud, and C. Erdman. 2015. "Reducing Respondent Burden in the American Community Survey's Computer Assisted Personal Visit Interviewing Operations – Phase 3 Results." American Community Survey Research and Evaluation Report Memorandum Series ACS14-RER-28-R1. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2015/acs/2015_Griffin_01.pdf (accessed February 2017).

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

Groves, R.M., J. Wagner, and E. Peytcheva. 2007. "Use of Interviewer Judgments about Attributes of Selected Respondents in Post-Survey Adjustment for Unit Nonresponse: An Illustration with the National Survey of Family Growth." In JSM Proceedings, Survey Research Methods Section, American Statistical Association, Salt Lake City, Utah, July 29-August 2, 2007. Alexandria, VA: American Statistical Association. 3428–3431. Available at: http://ww2.amstat.org/sections/SRMS/Proceedings/y2007/Files/JSM2007-000782.pdf (accessed February 2017).

Hedlin, D., Dale, T., Haraldsen, G., and Jones, J. eds. 2005. *Developing Methods for Assessing Perceived Response Burden*. Research Report. Stockholm: Statistics Sweden, Oslo: Statistics Norway, and London: Office for National Statistics. Available at: http://ec.europa.eu/eurostat/documents/64157/4374310/10-DEVELOPING-METHODS-FOR-ASSESSING-PERCEIVED-RESPONSE-BURDEN.pdf/1900efc8-1a07-4482-b3c9-be88ee71df3b (accessed February 2017).

Hughes, T., E. Slud, R. Ashmead, and R. Walsh. 2016. "Results of a Field Pilot to Reduce Respondent Contact Burden in the American Community Survey's Computer Assisted Personal Interviewing Operation." American Community Survey Research and Evaluation Report Memorandum Series #ACS16-RER-07. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2016/acs/2016_Hughes_01.pdf (accessed February 2017).

Klein, J. and M. Moeschberger. 2003. *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd ed. Springer-Verlag.

Kreuter, F., K. Olson, K.J. Wagner, T. Yan, T. Ezzati-Rice, C. Casas-Cordero, M. Lemay, A. Peytchev, R. Groves, and T. Raghunathan. 2010. "Using Proxy Measures and Other Correlates of Survey Outcomes to Adjust for Non-response: Examples from Multiple Surveys." *Journal of the Royal Statistical Society*: Series A 173: 389–407. Doi: http://dx.doi.org/10.1111/j.1467-985X.2009.00621.x.

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: 169–189. Doi: http://dx.doi.org/10.1111/j.1467-985X.2012.01080.x.

Maitland, A., C. Casas-Cordero, and F. Kreuter. 2009. "An Evaluation of Nonresponse Bias using Paradata from a Health Survey." In JSM Proceedings, Survey Research Methods Section, American Statistical Association, Washington, DC, August 1–6, 2009. Alexandria, VA: American Statistical Association. 370–378. Available at: http://ww2.amstat.org/sections/SRMS/Proceedings/y2009/Files/303004.pdf (accessed February 2017).

Mills, G. 2016. "Simulated Effects of Changing Calling Parameters and Workload Size on Computer Assisted Telephone Interview Productivity in the American Community Survey." American Community Survey Research and Evaluation Report Memorandum Series #ACS16-RER-22. Washington, DC: U.S. Census Bureau. Available at: https://census.gov/content/dam/Census/library/working-papers/2016/acs/2016_Mills_02.pdf (accessed August 2017).

Olson, K. and R. Groves. 2012. "An Examination of Within-Person Variation in Response Propensity over the Data Collection Field Period." *Journal of Official Statistics* 28: 29–51.

Schouten, B., F. Cobben, and J. Bethlehem. 2009. "Indicators for the Representativeness of Survey Response." *Survey Methodology* 35: 101–113.

Sharp, L. and J. Frankel. 1983. "Respondent Burden: A Test of Some Common Assumptions." *Public Opinion Quarterly* 47: 36–53. Doi: https://doi.org/10.1086/268765.

Slud, E. 1998. "Predictive Models for Decennial Census Household Response." In JSM Proceedings, Survey Research Methods Section, American Statistical Association, Dallas, Texas, August 9–13, 1998. Alexandria, VA: American Statistical Association. 272–277. Available at: http://ww2.amstat.org/sections/SRMS/Proceedings/papers/1998_043.pdf (accessed February 2017).

Slud, E. 1999. "Analysis of 1990 Decennial Census Checkin-Time Data." In FCSM Proceedings Federal Committee on Statistical Methodology, Washington, DC, November 15–17, 1999. Available at: https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/05/IX-B_Slud_FCSM1999.pdf (accessed August 2017).

Slud, E. and C. Erdman. 2013. "Adaptive Curtailment of Survey Follow-up Based on Contact History Data." In FCSM Proceedings Federal Committee on Statistical Methodology, Washington, DC, November 4–6, 2013. 1–11. Available at: https://s3.amazonaws.com/sitesusa/wpcontent/uploads/sites/242/2014/05/B1_Slud_2013FCSM.pdf (accessed August 2017).

Virgile, M. 2015. "Measurement Error in American Community Survey Paradata and 2014 Redesign of the Contact History Instrument." Center for Statistical Research and Methodology Report Series (Survey Methodology #RSM2016-01). Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/srd/papers/pdf/RSM2016-01.pdf (accessed February 2017).

Wagner, J. and T. Raghunathan. 2010. "A New Stopping Rule for Surveys." *Statistics in Medicine* 29: 1014–1024. Doi: http://dx.doi.org/10.1002/sim.3834.

Zelenak, M.F. and M. Davis. 2013. "Impact of Multiple Contacts by Computer-Assisted Telephone Interview and Computer-Assisted Personal Interview on Final Interview Outcome in the American Community Survey." American Community Survey Research and Evaluation Report Memorandum Series #ACS13-RER-08. Washington, DC: U.S. Census Bureau. Available at: www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Zelenak_01.pdf (accessed February 2017).

Zelenak, M.F. 2014. "Reducing Respondent Burden in the American Community Survey's Computer Assisted Personal Visit Interviewing Operation – Phase 1 Results (Part 1)." American Community Survey Research and Evaluation Report Memorandum Series #ACS14-RER-06. Washington, DC: U.S. Census Bureau. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2014/acs/2014_Zelenak_01.pdf (accessed February 2017).