Adaptive Curtailment of Survey Followup Based on Contact History Data

Eric Slud[1,2] **and Chandra Erdman**[1]

[1] U.S. Census Bureau, CSRM, 4600 Silver Hill Rd., Washington, DC 20933
[2] Mathematics Department, University of Maryland, College Park, MD 20742

Abstract. Several major Census Bureau household surveys collect a common form of contact-history paradata in the Contact History Instrument (CHI) and other similar systems. At each attempt at household contact, computer-assisted telephone (CATI) or personal (CAPI) interviewers enter various codes for interviewee reluctance in a standard format to aid in planning for subsequent contact attempts. The American Community Survey (ACS) in particular, in its CATI phase, schedules household call attempts with regard to previous nonproductive calls and reluctance. Analyses of ACS contact histories have recently been conducted (Zelenak and Davis 2013, Griffin and Hughes 2013), to explore CATI curtailment policies that can minimize the perceived respondent burden of households due to multiple contact attempts while losing as few interviews as possible.

This paper describes simple methods that can support adaptive design to yield optimal interview rates subject to constraints on survey followup. These methods make minimal use of models and consist of three primary elements: (i) Interview yields per contact attempt are constructed as functions of the number of contact attempts, on mutually exclusive subgroups of the survey population. (ii) Subgroups defined from contact-history variables are chosen to achieve clear separation of the curves defined in (i). (iii) Policies designed to curtail followup differentially in the subgroups found in (ii) are compared with respect to an array of costs, including measures of perceived respondent burden which themselves vary in separate paradata-defined groups, under simple models based on tabulated groupwise response behavior. No single policy can simultaneously minimize all costs, but this kind of analysis allows us to identify decision-theoretically admissible policies and to choose policies that most efficiently trade off the most essential cost elements. The steps described above are illustrated using ACS CATI data, in light of current ACS design aimed at minimizing perceived respondent burden.

Key words: adaptive design, discrete hazard function, loss function, respondent burden.

This paper describes research and analysis of the authors, and is released to inform interested parties and encourage discussion. Results and conclusions are the authors' and have not been endorsed by the Census Bureau.

1 Introduction

Methodological research is ongoing at the Census Bureau to collect as much survey data as possible while reducing costs expressed not only in terms of administrative overhead and personnel costs but also of perceived respondent burden. Like other censuses and surveys, the American Community Survey (ACS) relies on multiple contact attempts through multiple modes to secure survey data (Griffin and Hughes 2013). Within ACS and many other Census Bureau surveys, paradata concerning the series of attempts to obtain interviews from the selected households are now routinely collected. These paradata are being studied for their usefulness either in targeting or curtailing additional interviewer followup of sampled households according to data-defined categories. Future surveys and data-collection operations are likely to include elements of targeting and of curtailed followup as part of adaptive design.

Efforts to utilize contact-history paradata within Census Bureau administered surveys include research

projects to model survey response propensity as a function of paradata and known address information (Erdman and Morris 2013) and also include research like that of the present paper to study the effects of administrative policies which would curtail repeated contact attempts according to various possible plans. The time-dependent propensity models are very similar in spirit to event-history models developed in biostatistical contexts (Andersen et al. 1993, Klein and Moeschberger 2003) to describe failure rates in terms of possibly time-dependent risk factors. In other contexts, response-rate models may simply be logistic regression models with time-dependent covariates, as in Slud and Kedem 1994. In the sample-survey context, the time variable is an operational time, the number of contact attempts for a specific sampled household. Survey response propensity models making using of time-varying contact-history variables are being developed (Olson and Groves 2012, West and Groves 2013) and applied in the context of adaptive design in national household surveys (Erdman and Morris 2013). However, in the present research we remain as model-free as possible, attempting in an ACS data set to create mutually exclusive population subgroups defined in terms of cumulative contact-history counts, to separate interview completion rates as far as possible. Our method is first to calculate overall interview-completion rates for eligible households by group, and then to display group differences as a function of number of contact attempts.

The idea that we draw from biostatistics is the *life table method* (Klein and Moeschberger 2003), which is essentially to compute discrete time-specific rates of event (interview) occurrence. Letting the time variable t index successive contact attempts from now on, we refer to all eligible households for which neither an interview nor the end of contact attempts has occurred before t as being *at risk* at t. The discrete time-specific hazard rate function for a specific population group G is then defined as

$$r_G(t) \;=\; \#(\text{households } \textbf{interviewed} \text{ at } t)/\#(\text{households } \textbf{at risk} \text{ at } t) \tag{1}$$

By adopting specific policies of adaptive targeting and curtailment of followup, a survey incurs economic costs as well as non-economic costs such as those due to statistical and nonsampling errors. These costs can be linearly aggregated as components of a loss-function of the type studied in statistical decision theory (Ferguson 1967). We find in Section 2 a useful partition of the sampled at-risk survey population into subgroups defined from contact-history variables. The objective of these subgroups was, without becoming too small, to achieve reasonable separation of rates of interview completion in eligible cases. In Section 3 we confirm that these subgroups also maintain a reasonable separation of discrete hazard functions. As a by-product of this partition, we develop a new approach, described in Section 4, to measuring incremental respondent-burden as a function of t (number of contact attempts). A framework for assessing the respondent burden incurred under several distinct policies for curtailing ACS CATI followup is introduced in Section 5. An analysis of the trade-off between several categories of costs for policies of ACS CATI curtailment was provided by Griffin (2013) and Griffin and Hughes (2013) and will not be repeated here. But broadly speaking, we view the idea of minimizing linear combinations of loss-function elements, with coefficients reflecting the relative importance of each cost component, as the right way to proceed to design effective adaptive policies.

1.1 Data Setting

For general orientation to the ACS sampling design, see Census Bureau (2009). For present purposes, the important thing to understand is that the CATI phase of ACS data-collection applies to almost all 'mailable' ACS-sampled housing unit addresses for which a response was not received within one month and for which at least one telephone number is known. Of those sampled housing units that do not supply a sufficiently complete CATI interview, a fraction of about 1/3 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 data analyzed in this study consist of preprocessed CATI status and history files and CAPI history files covering the ACS monthly samples for which contacts began during the period June 2011–Feb. 2012. The data set consisted of 1,097,985 housing-unit records, of which 307,054 were only in CAPI, 600,203 only in CATI, and 190,728 were in both.

In each of the CATI and CAPI data stages, some cases are deemed ineligible. In CATI, ineligibility corresponds to a final status-code "Other", which means for the most part that a sampled 'mailable' case did not

have a correct or usable telephone number and so should not have been included in CATI. In CAPI, a final status code of "Type C" non-interview denotes ineligibility, with the interpretation that such a case is not a residential housing unit. Ineligible cases in CATI which are subsampled to CAPI are for the most part exactly like cases directly sampled into CAPI without being included in CATI. In the June 2011-Feb. 2012 data set discussed here, there were 326,705 eligible CATI cases, of which 41,819 were subsampled into CAPI. There were a total of 458,171 CAPI-eligible cases, of which 274,772 were only in CAPI and 142,339 were CATI-'Other' cases. Throughout the following analyses, all rates of interviews for CATI and CAPI are calculated only with respect to the universe of eligible cases in these ACS modes.

Important variables for each case in our CATI analyses were: final status of CATI case (Interview, termination because of reaching maximum number of allowed calls or of Hangups or of Refusals, or "Other," which signifies ineligibility for CATI), total numbers of calls resulting in Hangups, number of calls until the first, second, third and fourth Hangups; total number of calls resulting in Refusal or Interview, and the numbers of calls until the first and second Refusal; number of calls until a "Sufficient Partial" Interview if one occurs (i.e., a partial interview containing enough information to be deemed a Response or Interview); and number of calls resulting in a request for Callback, as well as number of calls until first such request. Important variables for each case in our CAPI analyses were: final status of CAPI case (Interview, Non-Interview, or Ineligible), total numbers of CAPI contact attempts (visits) resulting in a Hard Refusal, number of visits until first Hard Refusal; total numbers of visits resulting in Soft Refusals as well as numbers of visits until respectively first or second or third Soft Refusal; and total numbers of visits resulting in Reluctance as well as numbers of visits until respectively first or second or third visit with Reluctance. Differences between Hard and Soft Refusals and Reluctance are discussed in Section 2.2.

The current and contemplated curtailment policies for attempted contacts in CATI are expressed in terms of total number of calls, initial length of string of calls with no contact, number of Hangups, and number of Refusals within a single case. The corresponding control variables for policies governing CAPI visits are: number of visits, numbers of Hard and Soft Refusals, and number of visits resulting in Reluctance. One purpose of this research was to see whether any other control variables are potentially useful. As we describe next, we did find one such variable in CATI in the indicator of whether a household ever requested a Callback.

2 Definition of Population Subgroups

2.1 Subgroups of CATI Cases

The CATI cases in the data set were classified into groups defined by the cumulative numbers of Hangups (`HUP`), Refusals (`REF`) and Callback requests (`CB`) they generated in CATI. The corresponding group sizes and interview rates, after excluding the CATI ineligibles, are displayed in Tables 1 and 2. Note that in the CATI data set, call attempts ended with very few exceptions when the total number of Refusals reached 2 or when any Hangup or Refusal followed 3 earlier Hangups.

Table 1: Sizes of CATI eligible-case subgroups in June 2011–Feb. 2012 data defined by cross-classifying total numbers of Hangups (`HUP`), Refusals (`REF`) and Callbacks (`CB`).

	`CB` = 0				`CB` = 1+			
`HUP`	`REF` =	0	1	2	`REF` =	0	1	2
0		171,300	10,494	9,431		46,730	9,066	5,791
1		15,681	7,107	1,605		13,651	5,590	1,421
2		4,439	2,203	458		4,675	2,169	482
3+		6,542	1,450	1		4,980	1,438	1

Table 2 shows that the `CB` = 1+ groups had materially higher CATI interview rates, which had not been known before. In addition, the cases in (`HUP`, `REF`, `CB`) groups which did not provide CATI interviews and were sampled in CAPI have small differences in CAPI interview rates, as shown in Table 3. In the CAPI rate calculate of Table 3, the eligibility for CATI is ignored, but CAPI ineligibles are excluded.

Table 2: CATI-eligible interview rates in subgroups defined by (HUP, REF, CB) as in Table 1. Estimated rates for the few cases with HUP = 3+ and REF = 2 are unstable and are omitted.

	CB = 0			CB = 1+		
HUP	REF = 0	1	2	REF = 0	1	2
0	0.625	0.602	0.037	0.999	0.607	0.102
1	0.466	0.108	0.030	0.578	0.230	0.089
2	0.312	0.077	0.031	0.456	0.176	0.093
3+	0.069	0.015	*	0.168	0.061	*

Table 3: Group sizes and CAPI interview rates for cases that are in both CAPI and CATI, for groups defined by counters of HUP, REF, and CB in CATI.

[h]

		CB = 0		CB = 1+	
Group	Definition	Size	Rate	Size	Rate
G1	HUP=0, REF=0	147,296	0.954	8917	0.942
G2	HUP=1, REF=0	4,609	0.929	2647	0.944
G3	HUP=0, REF=1	2,167	0.915	1285	0.914
G4	HUP=2+, REF=0	1,377	0.929	1098	0.943
G5	HUP=1, REF=1	2,051	0.914	1371	0.928

The groups displayed in Table 3 are further subdivided according to CB = 0 versus CB = 1+. Since these groups are defined from cumulative counts of events observed within the contact history, they can also be used to define time-varying reluctance status for use as time-dependent covariates either in interview-propensity models or in measures of time-varying rate of change of respondent burden, as described in Section 4.

2.2 Subgroups of CAPI Cases

Since Table 3 indicates that CATI (HUP, REF, CB) status is by itself not terribly important for CAPI interview rates, we next consider a breakdown into groups that are particularly tailored to CAPI. Within the CAPI contact-history data set, there are *coded Refusals* which are definitive Refusals that we we refer to here as Hard Refusals; there are also *noncoded Refusals* that allow field representatives to categorize Non-Interviews through a variety of codes that we refer to here as incidents of expressed Reluctance; and within the set of Non-Interviews coded as episodes of Reluctance, those coded as *Firm Reluctance* were particularly noteworthy and are referred to in this paper as Soft Refusals.

Restricting attention to the 458,171 eligible CAPI cases in the June 2011–Feb. 2012 data set, we first separated out cases with at least one Hard Refusal, of which there are only 8,279 and which have the strikingly low CAPI-eligible interview rate of 0.227, as compared with an overall CAPI-eligible interview rate of 0.954. We then partitioned the CAPI cases that have no Hard Refusals according to the presence and number of visits with expressed Reluctance and, in particular, of Soft Refusals. We did this by creating a REFCOD score from 1 to 5, as follows. REFCOD = 5 indicates that a case has at least one visit resulting in a Hard Refusal. Among the cases with incidents of Reluctance, REFCOD = 1 indicates that a case has at most one CAPI visit resulting in Reluctance, while for those with at least 2 occasions of Reluctance, REFCOD = 2, 3, or 4 respectively indicates that the case has 0, 1, or at least 2 CAPI visits recorded as Soft Refusals. The numbers of CAPI cases in these five REFCOD groups and their corresponding eligible interview rates are given in Table 4.

The successive REFCOD groups 1-5 have clearly decreasing interview rate. We decomposed the REFCOD groups into subgroups of CAPI-only cases and cases that were in both CATI and CAPI, finding that these two groups hardly differed in CAPI interview rate within each REFCOD. Moreover, since we saw in Table 3 that the detailed (HUP, REF) CATI classification did not do much to separate CAPI interview rates, it does not make sense to divide the CATI-CAPI REFCOD subgroups further according to final counts of HUP and REF in

Table 4: REFCOD CAPI-eligible group sizes and CAPI interview rates.

REFCOD =	1	2	3	4	5
# cases	421,652	18,286	6,183	3,771	8,279
Interview Rate	0.970	0.952	0.872	0.811	0.277

CATI. Nevertheless, we did find some separation of CAPI interview rates within the large REFCOD = 2 and CATI-CAPI group between those cases with final CATI status of hitting Refusal-Max or Call-Max versus those with final status "Other" (CATI-ineligibility). We summarize our final separation of the CAPI cases into eleven groups labeled by an index Gcod, with definitions, group sizes and interview rates in Table 5.

Table 5: CAPI subgroup definitions, showing numbers of eligible cases, Interviews, and interview rates. CATI-St is final CATI status: Max for Refusal-Max or Call-Max, else Other.

Gcod	CAPI codes	in-CATI	CATI-St	Eligible	Interviews	Rate
0	$\leq$ 1 Reluctant	No	*	255,693	248,584	0.972
1	$\leq$ 1 Reluctant	Yes	Other	130,728	126,951	0.971
2	$\leq$ 1 Reluctant	Yes	Max	35,231	33,654	0.955
3	2+ Reluctant, 0 Soft	No	*	9,898	9,427	0.952
4	2+ Reluctant, 0 Soft	Yes	*	8,388	7,979	0.951
5	2+ Reluctant, 1 Soft	No	*	3,151	2,787	0.884
6	2+ Reluctant, 1 Soft	Yes	*	3,032	2,603	0.859
7	2+ Reluctant, 2+ Soft	No	*	1,940	1,597	0.823
8	2+ Reluctant, 2+ Soft	Yes	*	1,831	1,463	0.799
9	Hard	No	*	4,090	1,162	0.284
10	Hard	Yes	*	4,189	1,134	0.271

Thus, both within the CATI and CAPI data-collection modes in ACS, it is not hard to find a set of mutually exclusive case-subgroups that have visibly different interview rates. These subgroups can further be defined dynamically as of any specified contact attempt, from cumulative count and status variables in the earlier contact history. The dynamic subgroups become especially useful for adaptive design when their discrete hazard functions (1) remain separated across wide intervals of contact-attempt numbers.

In this paper, we have not drawn attention to the statistical significance of differences between groupwise interview rates. Precise calculations of standard errors would require replication-based estimates of variances of survey-weighted totals. However, we have maintained group sizes of at least about 1,800 with the idea that the half-width of the 95% confidence interval for a binomial proportion of at least 0.80 with sample size $n \geq 1800$ is $2\{(0.8)(0.2)/1800\}^{1/2} = 0.018$, so that overall rate differences of 0.02 are at least barely significant. When such differences are later found to be fairly consistent across contact attempts in the next Section, the group differences become more compelling.

3 Contact-attempt Specific Interview Rates

In CATI, we considered the ten subgroups formed by cross-classifying the groups G1-G5 defined in Section 2.1 with the indicator CB = 1+. Since these subgroups are defined completely in terms of cumulative HUP, REF and CB counts, they can also be defined for each call-number t by using the same definition expressed in terms of the cumulative counts up to call $t-1$. For these dynamically-defined subgroups, the group-specific discrete hazard curves (1) are displayed in Figure 1. Although there is some crossing of the group-specific curves, this figure shows that the ten dynamically-defined subgroups largely separate the time- (call-number) specific ACS interview rates in CATI. The picture is fairly striking not only because the separation of rates by group is mostly consistent across time but also because of the dramatic decrease of call-yield over time, suggesting that not much would be lost if the maximum number of CATI attempts were

curtailed. (The latter conclusion was already reached in Zelenak and Davis 2013.) The picture also indicates which subgroups of cases defined by contact history are best abandoned, and which are worth pursuing. In this way, model-free data analysis can inform adaptive design.

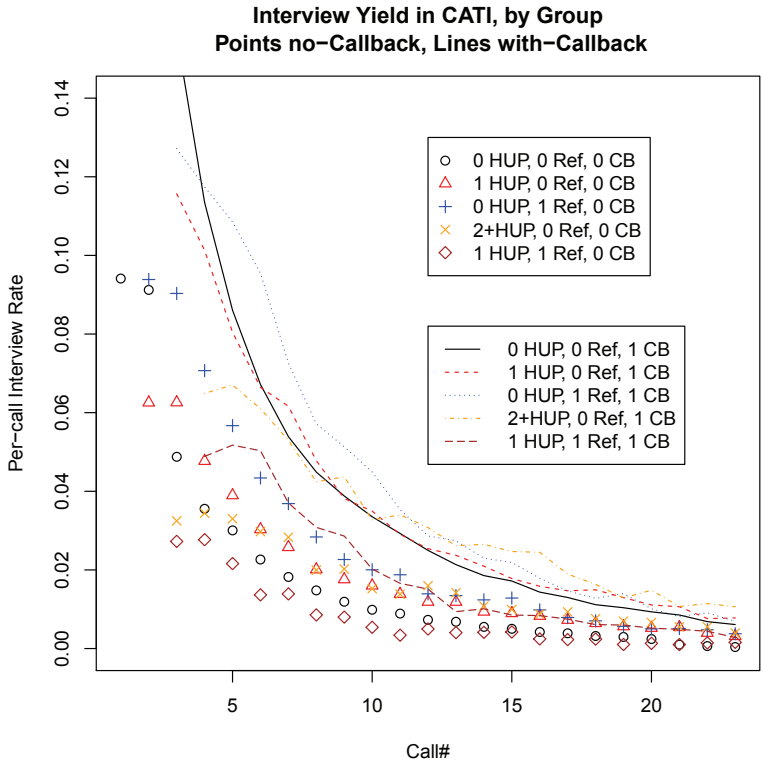

Figure 1: Call-specific Eligible Interview Rates by Group for ACS CATI Data, June 2011-Feb. 2012. The Groups are the same as `G1-G5` defined in Section 2.1, subdivided further by `CB` $= 0$ versus `CB` $= 1+$. Note that groups defined by counts $\geq t$ are empty before call t.

An analogous pattern emerges in CAPI for the subgroups described by the index `Gcod` defined in Table 5. Here also, as of each CAPI visit t, one may define the subgroups using the CAPI-history counts of Hard Refusals, visits with Reluctance, and Soft Refusals up through visit $t-1$ instead of the counts based on *all* CAPI visits as was done in calculating Table 5. The resulting picture in Figure 2 of discrete hazard functions (1) provides the visit-specific interview yields by group in CAPI. Note that just as the CATI "Other" final-status cases were excluded as ineligible in calculating the curves in Figure 1, the CAPI ineligible cases were excluded in the rates plotted in Figure 2. Almost as strongly as in Figure 1, the differences in interview rate between the major CAPI groups (those associated with different plotting symbols and colors in Figure 2) shown in Figure 2 are sustained across the range of CAPI visits. As expected from Table 5, the differences between curves for different CATI status within groups are unclear except in the largest group of cases (the blue curves) with at most 1 instance of Reluctance.

Examination and comparison of the rate curves displayed in Figures 1 and 2 lead to several remarks.

- The call-specific interview rate in CATI decays to 0 rapidly with increasing numbers of calls *in all CATI subgroups*. Therefore, economic considerations other than respondent burden would mandate an upper bound to the number of calls; moreover, the difference between rate-functions across groups suggests that there might be a benefit from choosing the upper bound to be different in different CATI subgroups. Thus, if the maximum for each group were to be assigned as the number of calls at which the rate function first dips below 0.02 per call, the maxima for groups `G1-G5` with `CB` $= 1+$ would respectively be roughly 14, 14, 16, 18, 10, and the maxima for groups `G1-G5` with `CB` $= 0$ would respectively be 7, 9, 10, 10, 6.

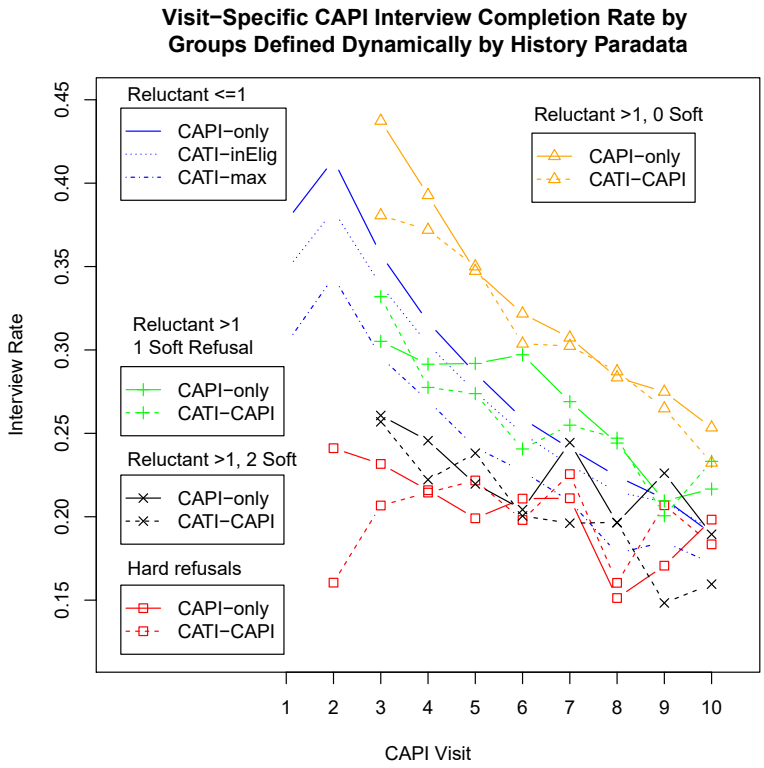

Figure 2: Visit-specific Eligible Interview Rates by Group for ACS CAPI Data, June 2011-Feb. 2012. Groups correspond to the `Gcod` values in Table 5.

- While the visit-specific rates of CAPI interviews do generally decrease, in all groups `Gcod0`–`Gcod10`, the final rate at 10 visits is generally at least about 2/3 as large as the rate at the first visit. In none of these groups does the rate fall below 0.15 by the 10^{th} visit.
- Research remains to be done on the extent to which additional variables (for example, neighborhood variables, housing type, etc.) known to the CAPI interviewers and their managers might further separate the interview rates and lead to efficient adaptive targeting of ACS cases. Research is also needed to determine whether the groups `Gcod0`–`Gcod10` are very different demographically, so that targeting by group might be inadvisable because of its effects on representativeness of the interviewed population. Moreover, CAPI visits involving actual contacts do carry a heavy respondent burden, so the measurement of costs related to burden should have a large impact on CAPI curtailment policy.
- Since overall CAPI visit-specific interview rates for eligible cases other than Hard Refusals are uniformly above 0.15 for up to 10 visits, the differences in discrete hazard functions pictured in Figure 2 must mean that the `Gcod` groups with lower curves require systematically more visits to achieve interviews. This could be translated both into economic and perceived-burden costs.

4 Time-varying Respondent Burden

The division of the sampled population into subgroups, both in CATI and CAPI, was done by choosing categories with increasing degrees of uncooperativeness as measured through counts of non-interview contact attempts recorded by the interviewers. The next step in our approach to the data analysis can be described as an attempt to define measures of potential interviewees' perceived burden reflecting each sampled household's duration within each such subgroup. Generally speaking, we expect the increase of perceived burden per contact attempt to be larger for a household at times where previous counts of noncooperative episodes are

larger. To explain this idea further, we restrict ourselves to the terminology of CATI, and we view each household's contact history as a trajectory through successive group labels marking level of noncooperation in previous attempts. As measured by counts of Hangups and Refusals, the trajectory shows increasing noncooperation, but in the CATI case, the count CB of requests for Callback is viewed as a positively cooperative signal. Generally, cases in the ten groups (G1-G5 x CB = 0 or 1+) at the t^{th} call can be assumed to perceive larger burden per call when they fall in groups for which the corresponding overall group interview rates given in Table 2 are smaller. However, the exact quantitative relationship between group interview rate and perceived burden is hypothetical so far, and remains to be specified.

One way to specify the perceived burden per call is to specify a table of relative burden values, such as

Group	CB = 0	CB = 1+
G1 (HUP=0, REF=0)	1.0	0.5
G2 (HUP=1, REF=0)	3.0	1.5
G3 (HUP=0, REF=1)	3.0	2.0
G4 (HUP=2+, REF=0)	4.0	2.0
G5 (HUP=1, REF=1)	8.0	4.0

The (purely hypothetical) entries in this table are interpreted as the increment of perceived burden experienced at any time t for which the cumulative CB and (HUP, REF) (implicit in the G1-G5 group definitions) at $t-1$ are as shown. For example, if at times 9-11 a case-record is in the status CB = 0, HUP = 1, REF = 0, (i.e., G2), the increment of respondent burden over those three calls is $3 \times 3 = 9$.

Two other was to quantify burden is to view the incremental burden per call in a given group G as as being given by the expressions $10 \times (1 - R_G)$ or $1/R_G$ in terms of the overall group interview rate R_G (given in the CATI case by selected entries in Table 2, and also in Table 5). (We refer to these two other hypothetical forms of perceived burden increments as H2 and H3.) In the CATI setting, we have altogether three alternative ways of measuring per-call perceived burden, which we summarize together with the CATI groupwise interview rates R_G in Table 6.

Table 6: Three hypothetical tables (in one line each, labeled H1, H2, H3) of groupwise per-call perceived burden in CATI, together with the groupwise interview rates R_G.

	CB = 0					CB = 1+				
	G1	G2	G3	G4	G5	G1	G2	G3	G4	G5
H1	1	3	3	4	8	0.5	1.5	2	2	4
H2	3.75	5.34	3.98	8.33	8.92	0.01	4.22	3.93	6.93	7.70
H3	1.60	2.15	1.66	5.99	9.26	1.00	1.73	1.65	3.26	4.35
R_G	0.62	0.47	0.60	0.17	0.11	1.00	0.58	0.61	0.31	0.23

For example, according to Table 6, a case for which 9 successive calls produced No Answer 3 times, then Hangup, then No Answer 2 times, then a Refusal, No Answer, and then Interview, would be coded with CB=0 and with group labels at the 9 calls of G1, G1, G1, G1, G2, G2, G2, G4, G4, and therefore would have measured burden according to the three possible measures H1, H2, or H3 respectively equal to $4 \times 1 + 3 \times 3 + 2 \times 4 = 21$, to $4 \times 3.75 + 3 \times 5.34 + 2 \times 8.33 = 47.68$, or to $4 \times 1.60 + 3 \times 2.15 + 2 \times 5.99 = 24.83$. Of course these burden measures are not comparable across different H cost tables, but only across cases for a single cost table.

Another, differently motivated, measure of respondent burden was used by Griffin and Hughes (2013) in a recent comparison of an array of twelve proposed control policies governing curtailment of followup in CATI. That measure was the number of calls placed following any incident of reluctance to respond. Another measure of burden that they and we have looked at in the same spirit is the number of productive calls (i.e., calls resulting in an answered telephone) following any initial show of reluctance.

The general objective of creating these measures of burden was to compare the outcomes of various curtailment policies (in CATI and CAPI, either separately or combined) with respect to overall yield of interviews,

yield of CATI versus CAPI interviews (which has an impact on cost), and perceived burden. The losses or costs which are computed for each curtailment policy based on all (CATI or CAPI or both) records are:

(a) Total number of CATI calls,
(b) Total number of CAPI visits,
(c) Number of households not providing a CATI or CAPI interview, and
(d) Total measure of perceived respondent burden.

In this list of outcomes, it is clear that policies curtailing the contact attempts would reduce losses in categories (a), (b) and (d) but increase them in (c). Defining the non-interview total (c) involves at least four methodological decisions: (i) that interviews obtained in CAPI after a number of visits greater than that allowed by a policy are deemed lost and recharacterized as Non-Interviews; (ii) that for CATI cases for which Sufficient Partials were received, each interview recorded with number of call attempts greater than that allowed by a policy is recharacterized as a CATI Non-Interview only if the policy terminates for that case before the call where the sufficient partial was achieved; (iii) that whenever a CATI curtailment policy recharacterizes a CATI interview as being lost, it is treated as being re-converted to a CAPI interview with probability equal to the rate at which cases in both CATI and CAPI (and eligible for CAPI) are converted to interviews in CAPI under whatever CAPI curtailment policy is in force; and (iv) that in the same setting as (iii), the CAPI visits for a lost CATI interview are imputed to be equal to the average CAPI visits for CAPI-eligible cases that are in both CATI and CAPI. Since the cases that actually yield CATI interviews under the current rules are likely to be somewhat more cooperative and responsive than those that do not, the costs and non-interview rates from alternative (more restrictive) CATI curtailment policies are likely to be somewhat less than is estimated by the current approach. For perspective, note that the proportion of CAPI interviews completed for cases that are also in CATI (and CAPI-eligible) was 0.9476 in this 2011-2012 data set, and the corresponding average number of CAPI visits was 3.02.

5 Economic Costs and Adaptive Design

It is beyond the scope of this paper to give further details of the cost analysis which was recently done (Griffin and Hughes 2013) to inform the ACS decision, implemented beginning with the March 2013 ACS monthly sample, to reduce two call parameters governing contact attempts in CATI, the maximum allowed numbers of calls and of consecutive unanswered calls before first contact. An example of the kinds of policies considered was the policy: terminate followup at the earliest occurrence of: 15 total calls, 12 initial unanswered calls, 2 Refusals, or a Refusal or Hangup following 2 Hangups.

The method of analysis by Griffin and Hughes (2013), which did make use of our data analysis and this research, was the one sketched at the end of the previous Section: to use each proposed curtailment policy to create a modified (artificial) ACS data set from the (CATI) cases in the June 2011–Feb. 2012 data set that we have used throughout this paper. For each proposed policy, the cases that gave an interview in the real data after the time at which the policy would have curtailed calls to them were treated as non-interviews, and deemed to be randomly subsampled into CAPI at a rate of 1 in 3. The costs (lost interviews, total calls and post-reluctance calls) in these artificial populations were tallied over all cases and compared.

More generally, the dynamic subdivision of a sampled population into subgroups with well separated contact-specific eligible interview rates has important and natural applications in adaptive design. Not every survey will have measures of respondent burden appearing as a primary cost, but all surveys will have many disparate and to some extent incommensurable cost elements: economic personnel and resource costs arising from multiple contact attempts, loss of accuracy due to sampled units not providing interview-responses, perhaps lower-quality data (e.g., due to more frequent omissions of survey questions) arising in some surveys from completed interviews achieved only after many contact attempts, etc. Stricter curtailment policies than those actually employed in collecting a survey data set can be investigated by recharacterizing interviews obtained after the stricter curtailment limits as being lost and tallying costs accordingly. Although it will not always be possible to realize in practice the hypothetical cost savings attained by carefully trading off various costs under these alternative curtailment scenarios, a recent report of Griffin (2013) documents that most of the expected cost and burden reductions anticipated from the recently implemented CATI call parameter changes are in fact being realized. In any case, the kind of cost-projection analysis sketched here provides

a framework for an analysis of estimated costs which is close to being fully design based, and in that sense reliable when based on large samples as in ACS.

6 Discussion

Like other statistical government agencies, the U.S. Census Bureau must remain attentive to the minimization of non-economic costs of running its household surveys, as well as those costs conventionally considered by survey methodologists. The less conventional costs include respondent burden as well as any nonsampling errors that result from undercoverage or from demographic and geographic imbalances that might result from curtailed efforts at making contact with sampled households.

The research described here provides a largely model-free framework for the development and analysis of survey costs associated with alternative curtailment strategies for the extended followup and contact attempts often used to obtain interviews in large household surveys.

Acknowledgments. The research described here was undertaken at the request of Debbie Griffin of the ACS office, by a team of researchers at the Census Bureau's Center for Statistical Research and Methodology (CSRM). That team consisted of ourselves along with Darcy Morris, Tom Petkunas, Josh Tokle, and Jerzy Wieczorek. The team worked together on data preprocessing and creation of flat-file extracts, and experimented with several alternative definitions of population subgroups and data displays for corresponding CATI interview-completion (and in a few cases, of CAPI interview-completion) rates. Further analysis and reporting was performed by the first author for CATI, and both authors jointly for CAPI. Original data files were supplied by Mary Frances Zelenak of the Census Bureau's Decennial Statistical Studies Division.

We are grateful to Debbie Griffin for many discussions which have educated us about issues in ACS related to this research, and to her and John Chesnut for comments on the paper that led to improvements.

References

Andersen, P., Borgan, Ø., Gill, R. and Keiding, N. (1993) *Statistical Models based on Counting Processes.* Springer-Verlag.

Census Bureau (2009), *Design and Methodology: American Community Survey*, http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology.pdf with additional online documentation at http://www.census.gov/acs/www/data_documentation/documentation_main/.

Erdman, C. and Morris, D. (2013) An Analysis of Adaptive Stopping Policies in the National Health Interview Survey, internal Census Bureau memorandum.

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

Griffin, D. (2013), Effect of Changing Call Parameters in the American Community Survey's Computer Assisted Telephone Interviewing Operation, *ACS Research and Evaluation Report Memorandum Series*, **#ACS13-RER-17**, U.S. Census Bureau.

Griffin, D. and Hughes, T. (2013), Analysis of Alternative Call Parameters in the American Community Survey's Computer Assisted Telephone Interviewing, *ACS Research and Evaluation Report Memorandum Series*, **#ACS13-RER-11**, U.S. Census Bureau.

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

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

Slud, E. (2012), Summary of Data Analysis to Inform Curtailment Policies for CATI Calls and CAPI Visits in ACS, Internal Memorandum Dec. 13, 2012, U.S. Census Bureau.

Slud, E. and Kedem, B. (1994) Partial likelihood analysis of logistic regression and autoregression, *Statistica Sinica* **4**, 89-106.

West, B. and Groves, R. (2013), A Propensity-Adjusted Interviewer Performance Indicator, *Public Opinion Quarterly*, **77**, 352-374.

Zelenak, M. and Davis, M. (2013), Impact of Multiple Contacts by Computer-Assisted Telephone Interview and Computer-Assisted Personal Interview on Final Interview Outcome in the American Community Survey, *ACS Research and Evaluation Report Memorandum Series*, **#ACS13-RER-08**, U.S. Census Bureau.