**NCSES Response to OMB Terms of Clearance for the
2019 National Survey of College Graduates**

28 February 2019

On 1 February 2019, the Office of Management and Budget (OMB) approved the collection of the 2019 National Survey of College Graduates (NSCG) under the following terms:

> *Within four weeks, NCSES will submit as a nonsubstantive change request updating Appendix H (Adaptive Survey Design Experiment) by adding a more complete description of the specific circumstances (values of monitored metrics) that will trigger the activation or deactivation of adaptive design interventions. Where specific triggering values of the monitored metrics can't be determined in advance of data collection, NCSES should explain in as much detail as possible how the metrics will be monitored and how decisions on interventions will be made. OMB approves NCSES starting production activities in the interim.*

In response to these terms of clearance, NCSES is submitting a revised version of Appendix H (attached), which includes an updated section on the "Adaptive Design Data Collection Intervention Schedule and Intervention Criteria" (see page 7). In addition to the detailed description provided in the text, Table H.1 has also been updated with more specific criteria for determining when interventions should be made (see page 11).

APPENDIX H – REVISED 2/28/19

2019 NSCG Adaptive Survey Design Experiment Goals, Interventions, Monitoring Metrics, and Potential Intervention Points

2019 NSCG Adaptive Design Experiment Goals, Interventions, and Monitoring Metrics

The 2019 NSCG Adaptive Design Experiment ("2019 Experiment") will be structured largely the same as the 2015 and 2017 NSCG Adaptive Design Experiments. Just as in those years, we will have experimental groups for the new sample cases (8,000) and the returning sample cases (10,000) with control groups identified for comparative purposes. Improvements will come from two directions for the 2019 Experiment:

1) Cases will be identified for interventions based on their ability to reduce the root mean squared error (RMSE) for key variables in the NSCG. Additionally, we will expand the data monitoring metrics that we implement during data collection to include evaluating the stability of survey estimates.
2) We will automate both the identification and selection of cases for interventions, as well as the delivery of the intervention file directly to the data collection modes. This will reduce the number of handoffs required to enact an intervention, making the implementation of adaptive design more efficient.

In 2015, NCSES and the Census Bureau worked to develop flow processing capabilities for the entire survey, with editing, weighting, and imputation occurring at time points during the data collection period as opposed to waiting until after data collection was over to perform the data processing. For the 2019 Experiment, we will be implementing simplified versions of flow processing to allow us to examine differences between the treatment and control groups not only with respect to representativeness and response rate, but also regarding stability of estimates and the effect of our nonresponse adjustment. These types of metrics will be considered as contributing factors in our decisions to make interventions.

Additionally, we will use past rounds of the NSCG to impute responses for nonrespondents throughout data collection, along with the propensity to respond given the application of particular data collection features and the cost of those features. These simulations will allow us to determine which features are most effective at reducing the RMSE of key estimates while understanding their effect on response rates and budget. We can use these simulations to evaluate the 2019 NSCG to see if the effects of data collection features are relatively stable over time.

The second improvement will continue the automation of the data analytic and business rule execution that was *ad hoc* in nature in the adaptive design experiments from previous cycles. While some monitoring metrics, including R-indicators, were run on an automated basis, specific decisions about when and where interventions should actually occur were the result of extended conversations and incremental data analysis. While these steps were important in the early stages of adaptive design, and for understanding how large interventions would be, adaptive design cannot be implemented in a standardized, repeatable production setting while maintaining such an extremely hands-on approach. For the 2019 Experiment, we will review the analytical questions that arose during past adaptive design decision meetings and attempt to automate these types of analyses in conjunction with the data monitoring metrics.

In a general sense, the goal of the 2019 Experiment is to evaluate new methods for case identification for interventions, expand usage of and access to data monitoring metrics, and develop a baseline level of comfort with automated interventions for adaptive design in a production setting.

The remainder of this appendix discusses several reasonable adaptive design goals, what interventions would allow the NSCG to achieve those goals, and what monitoring metrics would inform those interventions. As noted earlier, the 2019 Experiment will be structured largely the same as the 2015 and 2017 Experiments, and so the goals listed below are like the goals pursued as part of both the 2015 and 2017 Experiments. The major difference is that, instead of focusing on R-indicators, which only require frame data and response indicators, the selection criteria for interventions in the 2019 NSCG will utilize historical and current response data to intervene on cases that will reduce the RMSE of key survey estimates. However, both R-indicators and RMSE of key estimates can be used to reduce the risk of nonresponse bias in estimates and balance cost, so this change represents an expanding evaluation of monitoring metrics, without losing sight of our main adaptive design goals.

Goal 1: Balance Sample / Reduce Nonresponse Bias
Sampling balancing and/or reducing nonresponse bias relate to maintaining data quality in the face of shrinking budgets and falling response rates. Nonresponse bias arises when the outcomes of interest (the survey estimates) for respondents are different from those of nonrespondents. This difference results in a bias because the resulting estimates only represent a portion of the total target population. Surveys often try to correct for this *after* data collection using weighting, post-stratification, or other adjustments. Adaptive design interventions during data collection attempt to correct for nonresponse bias *during* data collection by actually changing the respondent population to be more balanced on frame characteristics related to response and outcome measures.

While discussing R-indicators, Schouten et al., provides reasons why balancing on variables related to response status and outcome variables is desirable. "In fact, we view the R-indicator as a lack-of-association measure. The weaker the association the better, as this implies that there is no evidence that nonresponse has affected the composition of the observed data." [3] This suggests that "selective forces…are absent in the selection of respondents" out of the sample population [2], and so nonresponse approaches missing at random, reducing the risk of nonresponse bias.

Interventions: Interventions are used to change the type or quantity of contacts targeted at specific subgroups or individuals. Interventions that will be considered for inclusion in the 2019 Experiment include:

- Sending an unscheduled mailing to sample persons;

- Sending cases to computer assisted telephone interviews (CATI) prior to the start of production CATI nonresponse follow up (NRFU), to target cases with an interviewer-assisted mode rather than limiting contacts to self-response modes;

- Putting CATI cases on hold, to reduce contacts in interviewer-assisted modes, while still requesting response in self-response modes;

- Withholding paper questionnaires while continuing to encourage response in the web mode to reduce the operational and processing costs associated with certain groups of cases;

- Withholding web invites to discourage response in certain groups of cases, while still allowing these cases to respond using previous invitations;

- Sending paper questionnaires to web nonrespondents earlier than the scheduled mail date to provide two modes of self-response rather than one; and

- Changing the CATI call time prioritization to increase or decrease the probability a case is called during a specific time.

Monitoring Methods:
- Root Mean Squared Error of Key Estimates
- R-indicators [2], [3], [4];
- Mahalanobis Distance or other distance measure [5];
- Response influence [6]; and
- Uncertainty/influence of imputed y-values [7].

We used R-indicators in the 2013 and 2015 Experiments and used a modified version of an R-indicator, and individual balancing propensity score, in the 2017 effort. As a metric, R-indicators were useful for measuring response balance, and served their purpose as a proof of concept for data monitoring. However, employing more metrics during data collection allows us to assess the usefulness of each monitoring metric and provides more confidence that data collection interventions were targeted in the most efficient way possible. That is, if R-indicators identify *subgroups* that should be targeted to increase response balance, and another metric (e.g., balancing propensity, response influence, Mahalanobis distance, etc.) identifies *specific cases* in those subgroups that also are likely to have an effect on nonresponse bias, then we have more confidence that those identified cases are the optimal cases for intervention, both from a response balance and nonresponse bias perspective.

Goal 2: Increase Timeliness of Data Collection
Analysts and other data users that need relevant, up-to-date information to build models, investigate trends, and write policy statements rely on timely survey data. NCSES specifically focused on timeliness as a goal for the 2013 NSCG [4] and reduced the length of time from the beginning of data collection to the time of data release from 28 months to 12 months. This required a reduction in the data collection from ten months to six months. In the future, NCSES is interested in further reducing data collection, specifically, from six months to five months.

Interventions: Interventions will attempt to either encourage response to the NSCG earlier than the standard data collection pathway or will be used to stop data collection if new respondents are not changing key estimates. This could be achieved by introducing modes earlier than the standard data collection pathway, sending reminders that elicit response more quickly, or

stopping data collection for all or a portion of cases and reallocating resources. Possible interventions include:

- Sending cases to CATI prior to the start of production CATI NRFU, to target cases with an interviewer-assisted mode rather than limiting contacts to self-response modes;

- Sending paper questionnaires to web nonrespondents earlier than the scheduled mail date to provide two modes of self-response rather than one;

- Sending email reminders earlier than the scheduled dates in data collection; and

- Stopping data collection for the sample or for subgroups given a sufficient level of data quality. For example, we could stop data collection if:
    - key estimates have stabilized, and standard errors fall within acceptable ranges, or
    - the coverage ratio for a subgroup of interest reaches a pre-determined threshold.

Monitoring Methods:
- Propensity to Respond by Modes [8];
- Change Point Analysis [9];
- Stability of Estimates [10]; and
- Coverage Ratios.

Ongoing NSCG research conducted by Chandra Erdman and Stephanie Coffey [8] could inform appropriate times to introduce new modes to cases ahead of the standard data collection schedule. Another possibility involves exploring change point analysis. If respondents per day as a metric changes over time, showing fewer responses in a given mode, there may be cause to introduce a new mode ahead of schedule. In addition, we will be able to calculate key estimates on a weekly or semi-weekly basis. As a result, we will be able to track stability of estimates during data collection to identify times when the data collection strategy has peaked, resulting in fewer responses or similar information that was already collected.

<u>Goal 3: Reduce Cost</u>
Controlling costs are always a survey management goal. More recently however, "the growing reluctance of the household population to survey requests has increased the effort that is required to obtain interviews and, thereby, the costs of data collection…[which] has threatened survey field budgets with increased risk of cost overruns" [10]. As a result, controlling cost is an important part of adaptive design. By allowing survey practitioners to reallocate resources during the data collection period, surveys can make tradeoffs to prioritize cost savings over other goals.

Interventions: Interventions will be used to encourage survey response via the web while discouraging response in more expensive modes (mail, CATI), or to eliminate contacts that may be ineffective. Possible interventions include:

- Putting CATI cases on hold, to reduce contacts in interviewer-assisted modes, while still requesting response in self-response modes;

- Withholding paper questionnaires while continuing to encourage response by web to reduce the operational and processing costs associated with certain groups of cases;

- Withholding web invites to discourage response from certain groups of cases, while still allowing these cases to respond using previous invitations;

- Prioritizing or deprioritizing cases in CATI during certain call times to increase or decrease the probability a case is called during a specific time frame without having to stop calling the case entirely; and

- Stopping data collection for the sample or for subgroups if key estimates and their standard errors have stabilized.

Monitoring Methods:
- Root Mean Squared Error of Key Estimates;

- R-indicators;

- Mahalanobis Distance or other distance measure;

- Response influence;

- Uncertainty/influence of imputed y-values;

- Stability of estimates; and

- Numbers of trips to locating.

The same indicators that are valuable for monitoring data quality also could measure survey cost reduction. If cases are in over-represented subgroups, or have low response influence, we may want to reduce or eliminate contacts on those cases.

In addition, the key estimates valuable to increasing timeliness, are also valuable for controlling cost. When estimates stabilize and their standard errors fall within acceptable limits for subgroups or the entire survey, new respondents are providing similar information to that which we have already collected. If continuing data collection would have little effect on estimates and their standard errors, stopping data collection to all or subgroups of cases would be an efficient way to control costs.

Another potential cost-saving intervention would be to limit the number of times a case could be sent to locating. If we have no contact information for a case, or previously attempted contact information has not been useful for obtaining contact, a case is sent to locating where researchers attempt to identify new, more up-to-date contact information. This operation can be time intensive, especially for cases repeatedly sent to locating. We could track the number of times a case is sent to interactive locating, or the length of time it spends in locating. Cases repeatedly sent to locating and cases that spend a large amount of time being researched may not be ultimately productive cases. Reallocating effort spent on these cases to those in locating for a

fewer number of times may be a sensible cost-saving measure that allows us to attempt contact on more cases, rather than spending large amounts of time (money) on the same cases.

Adaptive Design Data Collection Intervention Schedule and Intervention Criteria
To provide insight on the way that adaptive design criteria will be applied in the determination of interventions for the 2019 NSCG adaptive design experiment, NCSES is submitting a summary of the goals for this experiment, as well as a conceptual example below. Additionally, NCSES is submitting a table documenting the adaptive design intervention schedule and criteria (Table H.1.).

All sample cases will be monitored beginning at week 1. Adaptive interventions will be reviewed and implemented as needed at weeks 4, 5, 6, 8, 10, 12, 14, 16, 18, 20, 23, and 24 of the data collection period. As part of the adaptive design experiment, we have identified certain adaptive interventions that may be implemented depending upon the case monitoring results that could help the NSCG meet its data collection goals. The decision to implement an adaptive intervention will be based on the evaluation of specific criteria associated with the data collection metrics. The specific criteria are described generally below, and the specifics are provided in Table H.1.

Primary Data Collection Monitoring and Interventions:
We will use predictive models that estimate: (a) response by data collection mode and phase; (b) cost of particular data collection operations; and (c) the RMSE of key survey variables, so that interventions can focus on cost reduction without increasing the RMSE of survey estimates The goal of this experiment is to determine the most cost-effective pathway for individual cases and provide information on where survey data collection costs could be reduced without harming data quality. The knowledge gained from this experiment could help the NSCG allocate data collection resources differently in future rounds of the survey.

The interventions considered for a given week are designed to maintain or improve the RMSE of key survey estimates while reducing cost where feasible. Intuitively, we do not want to apply an expensive data collection feature (like telephone calls) to a case if we think that case would respond to a less expensive data collection feature, or if their nonresponse would have no effect on the final survey estimates. Here is an example of how this would work at week 12 in the NSCG:

- At week 12, all nonrespondent cases with a good telephone number would normally be sent to the CATI operation to receive outbound telephone calls.
- To determine the predicted effect of introducing CATI, we will do the following:
    - From prior rounds of data collection, we estimate the additional data collection costs that would be incurred for each nonrespondent case by sending it to CATI.
    - From prior rounds and the current round of data collection, we predict whether a case that is to be sent to CATI will respond in CATI, versus responding in other modes, or being a nonrespondent.

- o From prior rounds and the current round of data collection, we can also predict the survey outcome for key survey estimates (employment status, science and engineering indicator, and salary) in the NSCG for open cases predicted to respond (by any mode) before the end of data collection (using the actual responses for sample members that have responded).

These three models help us understand, given the baseline data collection operation, what the remainder of the survey will cost, who will respond, and how they will respond.

- Next, we select some percentage, say 10%, of cases who are predicted to be nonrespondents in CATI and simulate keeping those cases in the lower-effort data collection pathway, receiving web invites and reminders, but no outbound telephone calls.
  - o We re-estimate the data collection costs that would be incurred for this new pathway. This will result in cost savings.
  - o We re-predict, based on not sending these cases to CATI, whether the cases will respond, and if so, in what mode.
  - o We re-predict the survey outcome for key survey estimates for those cases to respond (by any mode) before the end of data collection.

These three models help us understand, under an alternate strategy (when 10% of likely nonrespondents in CATI are not sent to CATI), what the remainder of the survey will cost, who will respond, and how they will respond.

- At this point, we have two sets of predicted survey responses. We can estimate the RMSE of the "baseline" data collection strategy, and the RMSE of the "alternate" data collection strategy.
  - o If the alternate strategy results in lower costs without increasing the RMSE,[1] we repeat this process with a larger percentage of cases, say 15%. We continue repeating this whole process until we hit a point where, regardless of the cost savings, we are increasing the RMSE too much, and worsening data quality.
  - o If the alternate data collection operation increases the RMSE, it is not considered for implementation. At this point, we may try a smaller percentage of cases (5%), or, rather than considering likely nonrespondents, consider withholding cases from CATI if they are likely to respond in web. The process starts over.

In this way, we are evaluating many different potential data collection strategies to see their cost and quality properties. While cost savings are always a goal in data collection operations, it is important to also maintain data quality. Additionally, these estimations will be automated so that a variety of data collection pathways are evaluated without human intervention with respect to cost and RMSE. In this way, at each intervention point, we will be examining both cost and

---

[1] In general, we do not want the RMSE for key estimates to increase very much, even if we are achieving significant cost savings. At the same time, the additional complexity of carrying out an intervention (in the telephone centers and the mailout operations center) and the additional effort required by survey contractor staff may outweigh the potential cost savings of a potential alternate strategy, if the cost savings are predicted to be small. As a result, we have created thresholds different from just "more than a 0% cost savings" and "no increase in RMSE" in order to have an alternate strategy be considered viable. Those thresholds are noted in Table H.1.

quality properties of different data collection features (like sending cases to CATI early, or withholding mailed reminders), and we can select an alternate data collection pathway when it is predicted to be beneficial.

Because the NSCG has a sequential design, there are time-varying cost and response properties to keep in mind – that is, not all interventions are available at all times. For example, you cannot introduce a web invite in week 16, because web invites were introduced in week 1, so they are no longer novel. During the later weeks of data collection, sending a web invite could be done as a way of reducing effort in CATI, but not for introducing a new mode. This has a direct effect on adaptive design. Early in the data collection, the adaptive interventions generally attempt to actively improve the target metric (RMSE or R-indicator) by increasing response among cases selected for intervention. Therefore, we will generally be identifying cases who should receive more data collection resources. During the middle of the data collection, some of the interventions continue to improve response for cases that require more data collection resources, for example with extra questionnaire mailings to the specific groups. However, other interventions reduce effort for cases, and these interventions can be applied to cases that are either equally likely to respond to a more or a less expensive data collection feature, or those cases that are just highly unlikely to respond. Finally, near the end of the data collection, using metrics such as the number of trips to locating, response propensities, and the number of call attempts, the interventions focus on controlling data collection costs.

Secondary Data Collection Monitoring:
Because we are evaluating a different data collection intervention strategy than what was carried out in prior adaptive design experiments, we will *also* create and monitor R-indicators as they were implemented in prior rounds of the survey. The propensity models underlying the NSCG R-indicators have been validated over the past three data collection cycles and include variables that are highly correlated with survey outcomes, giving us confidence that our interventions are improving data collection, even if the metric we use is only a proxy for nonresponse bias. This process will allow us to compare, during and after data collection, whether the two different intervention frameworks select similar or disparate cases. Similar case selection would further support the idea that R-indicators are useful metrics for monitoring and intervention in surveys that have rich auxiliary frame data, like the NSCG. For this secondary monitoring metric, we will utilize the thresholds of the unconditional category level partial R-indicators that were used in the prior rounds of the NSCG [11].

The list of potential interventions for each week is shown in Table H.1., which includes information about metrics and criteria used. Additionally, a flowchart view of the potential data collection interventions illustrates which interventions are available each week (see Appendix H.2).

References:

[1] Coffey, S. (2014, April). "Report for the 2013 National Survey of College Graduates Methodological Research Adaptive Design Experiment." Census Bureau Memorandum for NCSES.

[2] Schouten, B., Cobben, F., Bethlehem, J. (2009, June). "Indicators for representativeness of survey response." *Survey Methodology*, 35.1, 101-113.

[3] Schouten, B., Shlomo, N., Skinner, C. (2011). "Indicators for monitoring and improving representativeness of response." *Journal of Official Statistics*, 27.2, 231-253.

[4] Coffey, S., Reist, B., White, M. (2013). "Monitoring Methods for Adaptive Design in the National Survey of College Graduates (NSCG)." *2013 Joint Statistical Meeting Proceedings,* Survey Research Methods Section. Alexandria, VA: American Statistical Association.

[5] de Leon A.R., Carriere K.C. (2005). "A generalized Mahalanobis distance for mixed data." *Journal of Multivariate Analysis,* 92, 174-185.

[6] Särndal, C., Lundström, S. (2008). "Assessing auxiliary vectors for control of nonresponse bias in the calibration estimator." *Journal of Official Statistics,* 24, 167-191.

[7] Wagner, J. (2014). "Limiting the Risk of Nonresponse Bias by Using Regression Diagnostics as a Guide to Data Collection**."** Presentation at the 2014 Joint Statistical Meetings. August 2014.

[8] Erdman, C., Coffey, S. (2014). "Predicting Response Mode During Data Collection in the NSCG." Presentation at the 2014 Joint Statistical Meetings. August 2014.

[9] Killick, R., Eckley, I. (2014). "Changepoint: An R Package for Changepoint Analysis." Downloaded from http://www.lancs.ac.uk/~killick/Pub/KillickEckley2011.pdf on August 8, 2014.

[10] Groves, R.M., Heeringa, S. (2006). "Responsive design for household surveys: tools for actively controlling survey errors and costs." *Journal of the Royal Statistical Society Series A: Statistics in Society,* 169, 439-457.

[11] Coffey, S., Miller, P., Reist, B. "Interventions On-Call: Dynamic Adaptive Design in the 2015 National Survey of College Graduates." *Journal of Survey Statistics and Methodology*. Under Review.

Table H.1. Potential Intervention Points

Week	Production Operation Description	Adaptive Design Interventions	How to determine to intervene using cost and RMSE as the quality metrics?	How to determine to intervene using R-indicators as the quality metric?	Other contributing factors
1	Week 1 Web Invite, Incentives (If Appropriate)	No interventions.	N/A	N/A	N/A
2	Week 2 Reminder, Questionnaire Mailing (If Mail Preference)	No interventions.	N/A	N/A	N/A
4-23	Various	Activate cases in CATI early or take cases in CATI off hold	If the alternate strategy is predicted to result in $\geq$ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by $\geq$ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator < -0.01, intervene.[2]	If these subgroups are low interest groups (e.g., non-S&E), we may not intervene. If the subgroups are very large and we do not want to move all cases to CATI, use response propensity for these cases, and move over only the "higher" propensity cases.
		Put cases in CATI on hold	If the alternate strategy is predicted to result in $\geq$ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by $\geq$ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If key estimates of interest have not stabilized in the experimental group, we may not intervene.
4 - 23	Various	Send an off-production-path questionnaire	If the alternate strategy is predicted to result in $\geq$ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by $\geq$ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator < -0.01 *and* a case has a higher probability to respond by paper than by web, intervene.	If these cases are in over-represented groups or if they are in low interest groups (e.g., non-S&E), we may not intervene.

---

[2]All R-indicator thresholds documented here were used in adaptive design experiments in past rounds of the NSCG. For more detail on their usefulness, see [11].

			If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the most over-represented subgroups are not much different from other groups, we may not intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.
5 6 12	Weeks 5, 6, 12, Reminder Email	Withhold email reminder	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the most over-represented subgroups are not much different from other groups, we may not intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.
5	Week 5 Reminder Letter	Withhold reminder contact	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the most over-represented subgroups are not much different from other groups, we may not intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.
6	Week 6 Reminder Postcard	Withhold reminder contact	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the most over-represented subgroups are not much different from other groups, we may not intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.
8	Week 8 Questionnaire with Web Invite	Withhold questionnaire but not web invite	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the probability of responding by mail > the probability of responding by web, we may not intervene in all the cases in these subgroups. If the most over-represented subgroups are not much different from other groups, we may not intervene.
10	Week 10 Reminder Email	Withhold email reminder	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the most over-represented subgroups are not much different from other groups, we may not intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.
12	Week 12 Pressure Sealed, Perforated Reminder (Start of	Withhold reminder contact	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data	If the most out-of-balance variable (based on the variable level unconditional	If the most over-represented subgroups are not much different from other groups, we may not

	CATI)		collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.
16	Week 16, Postcard Reminder	Withhold reminder contact	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the most over-represented subgroups are not much different from other groups, we may not intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.
18	Week 18, Web Invite (Prior Round Respondents)	Withhold reminder contact	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the most over-represented subgroups are not much different from other groups, we may not intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.
18	Week 18 Questionnaire with Web Invite (Prior Round Nonrespondents)	Withhold questionnaire but not web invite	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the probability of responding by mail > the probability of responding by web, we may not intervene in all the cases in these subgroups. If the most over-represented subgroups are not much different from other groups, we may not intervene.
20	Week 20, Web Invite, new sample, Priority envelope, questionnaire	Withhold questionnaire but not web invite	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the probability of responding by mail > the probability of responding by web, we may not intervene in all the cases in these subgroups. If the most over-represented subgroups are not much different from other groups, we may not intervene.
23	Week 23 Last Chance Email	Withhold email reminder	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a	If the most over-represented subgroups are not much different from other groups, we may not intervene.

			increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If key estimates of interest have not stabilized for the experimental group, we may not intervene.
23	Week 23, Web Invite	Withhold reminder contact	If the alternate strategy is predicted to result in ≥ 5% savings in remaining data collection costs without increasing the predicted RMSE of final estimates by ≥ 5%, intervene.	If the most out-of-balance variable (based on the variable level unconditional partial R-indicators) has a subgroup where the category-level unconditional partial R-indicator > 0.01, intervene.	If the most over-represented subgroups are not much different from other groups, we may not intervene. If key estimates of interest have not stabilized for the experimental group, we may not intervene.