Memorandum

United States Department of Education
Institute of Education Sciences
National Center for Education Statistics

DATE: July 6, 2017

TO: Robert Sivinski, OMB

THROUGH: Kashka Kubzdela, OMB Liaison, NCES

FROM: David Richards, BPS:12/17 Project Officer, NCES
Tracy Hunt-White, Team Lead, Postsecondary Longitudinal and Sample Surveys, NCES

SUBJECT: 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) Incentive Boost Change Request (OMB# 1850-0631 v.15)

The 2012/17 Beginning Postsecondary Students Longitudinal Study (BPS:12/17) is conducted by the National Center for Education Statistics (NCES), within the U.S. Department of Education (ED). BPS is designed to follow a cohort of students who enroll in postsecondary education for the first time during the same academic year, irrespective of the date of high school completion. Data from BPS are used to help researchers and policymakers better understand how financial aid influences persistence and completion, what percentages of students complete various degree programs, what are the early employment and wage outcomes for certificate and degree attainers, and why students leave school. The request to conduct the BPS:12/17 full-scale data collection was approved by OMB in December 2016 (1850-0631 v.10) with change requests approved in January, February, and April 2017 (OMB# 1850-0631 v.11-14). This request is to: (1) use specific incentive strategies with the main BPS:12/17 sample, based on recent modeling using study nonrespondents, and (2) make minor modifications to the BPS:12/17 instrument facsimile, specifically to the instructions displayed on a form used to collect information when a response fails the security question. This request does not introduce changes to respondent burden or the cost to the federal government.

<u>BPS:12/17 responsive design plan</u>

The first stage of the BPS:12/17 responsive design plan was a calibration sample, completed and described in a previous change request (OMB# 1850-0631 v.13). This document, for the current change request, describes phase one of the second stage of the BPS:12/17 responsive design plan, the selection and targeting of cases for an additional monetary incentive. In August 2017 we will update OMB again, at the start of the second phase of stage two, when cases will be selected for an offer to complete an abbreviated interview.

This change memorandum summarizes results of the modeling and targeting activities performed in accordance with the procedures described in the approved Supporting Statement Part B (OMB# 1850-0631 v.14). Attachment A, appended in this document, provides an excerpt from Part B that describes the responsive design plan, including the results of the BPS:12/14 responsive design approach (Section 4.a) and the BPS:12/17 plan (Section 4.b). Attachment B, also appended in this document, presents the variables used in the BPS:12/17 responsive design plan models, along with the relevant analyses results.

While similar in approach, the BPS:12/17 design includes unique features as compared to BPS:12/14: different experiments during the calibration sample, a special protocol for one subgroup (double nonrespondents), targeting selected cases that would reduce nonresponse bias within institution type (or sector) estimates, and a different series of interventions during main sample data collection. First, the calibration sample tests the efficacy of a prepaid incentive and the efficacy of offering a higher incentive amount versus an abbreviated interview. Secondly, for double nonrespondents (those who have responded to neither the NPSAS:12 nor the BPS:12/14 student interviews), BPS:12/17 includes a special data collection protocol that potentially includes a larger incentive amount, as well as an accelerated timeline for data collection interventions. Lastly, BPS:12/17 uses an *importance* score (which considers both likelihood to respond and likelihood of contributing to bias) to select nonrespondents most worth targeting within institution types. The BPS:12/17 plan includes two points in the main sample data collection at which nonresponding sample members will be targeted in an effort to obtain a response from them, the first one using a $45 incentive boost, and the second offering an abbreviated survey.

Targeting cases for a monetary incentive boost

In mid-June, 2017, eligible survey nonrespondents were selected for bias likelihood modeling. The allocation of sample members for modeling is shown in Table 1. In total 15,202 current nonrespondents were included in the bias likelihood model. Results were combined with *a priori* response propensity modeling data, constructed from NPSAS:12 and BPS:12/14 data, to produce the importance measure.

Table 1. Allocation of sample for response propensity and bias likelihood modeling

BPS:12/17 sample	Number of cases	Included in *a priori* propensity model	Included in bias-likelihood model	Eligible for targeting
Total	33,728			
Deceased	3	No	No	No
Double-Nonrespondents	3,271	Yes	No	No
Control cases[1]	3,152	Yes	Yes	No
Current respondents[2]	12,100	Yes	Yes	No
Current nonrespondents[2]	15,202	Yes	Yes	Yes

[1] Note that 463 more control cases are double nonrespondents.
[2] Response status through 6/28/2017 for cases that are not deceased, double nonrespondents, or controls.

As described in Attachment A, below, the BPS:12/17 responsive design plan aims to reduce the impact of unit nonresponse *within institution type (or sector)*. Institution types are grouped as follows:

Institution Group A: Public less-than 2-year, Public 2-year

Institution Group B: Public 4-year non-doctorate-granting; Public 4-year doctorate-granting; Private nonprofit less than 4-year; Private nonprofit 4-year non-doctorate-granting; Private nonprofit 4-year doctorate-granting

Institution Group C: Private for profit less-than-2-year

Institution Group D: Private for profit 2-year

Institution Group E: Private for profit 4-year

The individual nonresponse rates within these institution groups at the conclusion of the BPS:12/14 student interview were reviewed and compared, and are shown here in Table 2. Our targeting approach is to select cases with high importance scores in proportion to the institution group nonresponse rates in BPS:12/14, in order to reduce nonresponse bias by increasing the responses of these targeted cases within each of the institution groups. Proportional targeting begins by taking the total number of nonrespondent cases by institution group and subtracting the control cases and exclusions. The resulting number is multiplied by the nonresponse rate for that sector in BPS:12/14 (e.g., multiplying 4,281 by 35.6 percent for institution group A). The resulting numbers of targeted cases to be offered the $45 incentive are shown in Table 2, by institution group. The total of 4,195 targeted nonrespondents is approximately 35 percent of the total available current nonrespondents.

Table 2. BPS:12/14 final unweighted nonresponse rates and targeted cases by institution group

Institution group	BPS:12/14 final unweighted nonresponse rate (percent)	BPS:12/17 eligible nonrespondent cases (number)	Cases to target (number)
A	35.6	4,281	1,523
B	22.3	3,469	773
C	47.4	323	153
D	40.8	1,100	450
E	38.4	3,374	1,296
Total	33.4	12,547	4,195

Note: BPS:12/17 eligible nonrespondents (less control cases and exclusions) as of 6/29/17.

While response rates are not a good indicator of nonresponse bias across studies (Groves and Peytcheva, 2008), different response rates within a study have been observed to covary with survey estimates (e.g., Peytchev, 2013), which is indicative of a link between response rates and nonresponse bias. The described approach aims to allocate greater effort to the sectors with lower response rates, for which there is increased risk of nonresponse bias.

As described below in Attachment A, propensity scores above high and low cutoffs (as determined by a review of the predicted distribution) will be excluded as potential targets during data collection. The proposal to exclude cases from targeting was adopted in BPS:12/14 as a way to be thoughtful and efficient with the expenditure of project resources (i.e., the available budget for incentives and data collection activities) by not expending additional resources on sample members who are either very likely to respond without additional treatment or very unlikely to respond no matter what treatments are offered. Statistical processes (such as outlier analysis or distance to the mean value) do not lend themselves well to the determination of which cases to exclude based on resource considerations, because excluded cases are not outliers in a statistical distribution sense. For instance, k-means clustering was explored as a technique to partition observations. However, results showed that this method would lead to the exclusion of more cases on the upper end and fewer cases on the lower end than would be excluded by visual inspection, even when three clusters were specified to approximate the top and bottom exclusions groups with a middle target group. For example, for institution group A, 28 percent of the cases (25 percent from the top and 3 percent from the bottom) would be excluded using k-means clustering, versus 17 percent (7 percent from the top and 10 percent from the bottom) when using a visual inspection approach. Hence too few high propensity cases would be excluded and too many low propensity cases would be included for targeting, resulting in a less effective use of study resources.

The exclusion of cases with high and low propensity scores is a somewhat subjective decision, intended to focus limited resources where they are most likely to be effective. Exclusion is not strictly necessary; all cases could be included in the targeting process. However, doing so would not be the most effective use of study resources.

Cases excluded from the top and bottom have been chosen differently because cases at either end of the response propensity score distribution behave differently. While cases at the bottom may still respond, they are substantially less likely to do so, regardless of treatment. Therefore the bottom 10 percent of cases within each institution group is excluded from targeting. In contrast, the top proportion exclusion cut-offs were determined via visual examination of each institution group's scatterplot of *a priori* propensity scores (see Attachment C below in this document). Each institution group appeared to have a small grouping of cases with higher response likelihoods that were separated from the rest of the cases by natural gaps in the data. As these gaps were at different response-propensity levels for each institution group, each institution group was assigned a slightly different cut-off value for the cases with higher response likelihoods.

Data collection with these 4,195 sample members identified for targeting is scheduled to begin after the 14th week of data collection, on or about July 5, 2017. With the original baseline incentive of $30 plus the increased monetary incentive of $45, these targeted nonrespondents will be offered a total of $75 to complete the survey. As described below in Attachment A, during week 21 of data collection, modeling and targeting will be repeated and cases will be selected to be offered an abbreviated interview to further encourage their completion of the survey. Data collection activities for non-targeted cases will continue as planned.

Changes to the instrument facsimile (Appendix G)

We have revised a security question to improve respondent comprehension of this item's purpose. The item collects contact information for respondents whose answer to the security challenge question RESPCONF3, which asks them to select the school they were attending in the 2011-12 academic year, is different from that expected based on prior BPS data. In such cases, contact information is requested as a second attempt to verify that the correct individual is attempting to complete the survey, and to follow up with these individuals when needed. Specifically, the following information is requested: name of school attended in 2011-12, email address, phone numbers, and mobile number for text messages. Wording was revised in RESPCONF3 to clarify to the respondent why these data are being requested. The wording revision, which is included on page G-18 in the revised Appendix G, is as follows:

- PREVIOUSLY APPROVED: Based on your response, it seems you may not be eligible for this study. We will review your responses and may need to contact you again if we determine that you are eligible to participate in this survey.
- REVISED: In order to verify that we are surveying the correct person, **please provide at least 3 pieces of information.** Without this information, we are unable to confirm your identity and you will not be able to proceed in the survey.

In addition, we corrected Appendix G to remove reference to two variables that were removed as part of OMB# 1850-0631 v.11. Specifically, we removed reference to B17RESPCONF4 and SECFAILTEXT that previously appreared on G-2 and G-6. The wording for these items had been removed but incidental reference to them had not been properly deleted from Appendix G.

References

Groves, R. & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, 72 (2), 167–189.

Peytchev, A. (2013). Consequences of survey nonresponse. *The Annals of the American Academy of Political and Social Science*, *645*(1), 88–111.

Attachment A. Excerpt from the approved Supporting Statement Part B (OMB# 1850-0631 v.14)

4. Tests of Procedures and Methods

The design of the BPS:12/17 full-scale data collection—in particular, the use of responsive design principles to reduce bias associated with nonresponse—expands on data collection experiments designed for several preceding NCES studies and, particularly, on the responsive design methods employed in BPS:12/14. Section B.4.a below provides an overview of the responsive design methods employed for BPS:12/14, section B.4.b provides a description of the proposed methods for BPS:12/17, and section B.4.c describes the tests that will be conducted through the BPS:12 PETS pilot study.

BPS:12/14 Full Scale[1]

The BPS:12/14 full-scale data collection combined two experiments in a responsive design (Groves and Heeringa 2006) in order to examine the degree to which targeted interventions could affect response rates and reduce nonresponse bias. Key features included a calibration sample for identifying optimal monetary incentives and other interventions, the development of an importance measure for use in identifying nonrespondents for some incentive offers, and the use of a six-phase data collection period.

Approximately 10 percent of the 37,170 BPS:12/14 sample members were randomly selected to form the *calibration sample*, with the remainder forming the *main sample*, although readers should note that respondents from the calibration and main sample were combined at the end of data collection. Both samples were subject to the same data collection activities, although the calibration sample was fielded seven weeks before the main sample.

First Experiment: Determine Baseline Incentive. The first experiment with the calibration sample, which began with a web-only survey at the start of data collection (**Phase 1**), evaluated the baseline incentive offer. In order to assess whether or not baseline incentive offers should vary by likelihood of response, an *a priori* predicted probability of response was constructed for each calibration sample member. Sample members were then ordered into five groups using response probability quintiles and randomly assigned to one of eleven baseline incentive amounts ranging from $0 to $50 in five dollar increments. Additional information on how the *a priori* predicted probabilities of response were constructed is provided below.

For the three groups with the highest predicted probabilities of response, response rates for a given baseline incentive offer ($0 to $25) were statistically higher than response rates for the next lowest incentive amount up to $30. In addition, response rates for incentives of $35 or higher were not statistically higher than response rates at $30. For the two groups with the lowest predicted probabilities of response, the response rate at $45 was found to be statistically higher than the response rate at $0, but the finding was based on a small number of cases. Given the results across groups, a baseline incentive amount of $30 was set for use with the main sample. Both calibration and main sample nonrespondents at the end of Phase 1 were moved to **Phase 2** with outbound calling; no changes were made to the incentive level assigned at the start of data collection.

Second Experiment: Determine Monetary Incentive Increase. Phase 3, a second experiment implemented with the calibration sample, after the first 28 days of Phase 2 data collection, determined the

[1] This section addresses the following BPS terms of clearance: (1) From OMB# 1850-0631 v.8: “OMB approves this collection under the following terms: At the conclusion of each of the two monetary incentive calibration activities, NCES will meet with OMB to discuss the results and to determine the incentive amounts for the remaining portion of the study population. Further, NCES will provide an analytical report back to OMB of the success, challenges, lessons learned and promise of its approach to addressing non-response and bias via the approach proposed here. The incentive levels approved in this collection do not provide precedent for NCES or any other Federal agency. They are approved in this specific case only, primarily to permit the proposed methodological experiments.”; and (2) From OMB# 1850-0631 v.9: “Terms of the previous clearance remain in effect. NCES will provide an analytical report back to OMB of the success, challenges, lessons learned and promise of its approach to addressing non-response and bias via the approach proposed here. The incentive levels approved in this collection do not provide precedent for NCES or any other Federal agency. They are approved in this specific case only, primarily to permit the proposed methodological experiments.”

additional incentive amount to offer the remaining nonrespondents with the highest "value" to the data collection, as measured by an "importance score" (see below). During Phase 3, 500 calibration sample nonrespondents with the highest importance scores were randomly assigned to one of three groups to receive an incentive boost of $0, $25, or $45 in addition to the initial offer.

Across all initial incentive offers, those who had high importance scores but were in the $0 incentive boost group had a response rate of 14 percent, compared to 21 percent among those who received the $25 incentive boost, and 35 percent among those who received the $45 incentive boost. While the response rate for the $25 group was not statistically higher than the response rate for the $0 incentive group, the response rate for the $45 group was statistically higher than the response rates of both the $25 and the $0 groups. Consequently, $45 was used as the additional incentive increase for the main sample.

Importance Measure. Phases 1 and 3 of the BPS:12/14 data collection relied on two models developed specifically for this collection. The first, an *a priori* response propensity model, was used to predict the probability of response for each BPS:12/14 sample member prior to the start of data collection (and assignment to the initial incentive groups). Because the BPS:12/14 sample members were part of the NPSAS:12 sample, predictor variables for model development included sampling frame variables and NPSAS:12 variables including, but not limited to, the following:

- responded during early completion period,
- interview mode (web/telephone),
- ever refused,
- call count, and
- tracing/locating status (located/required intensive tracing).

The second model, a bias-likelihood model, was developed to identify those nonrespondents, at a given point during data collection, who were most likely to contribute to nonresponse bias. At the beginning of Phase 3, described above, and of the next two phases – local exchange calling (Phase 4) and abbreviated interview for mobile access (Phase 5) – a logistic regression model was used to estimate, not predict, the probability of response for each nonrespondent at that point. The estimated probabilities highlight individuals who have underrepresented characteristics among the respondents at the specific point in time. Variables used in the bias-likelihood model were derived from base-year (NPSAS:12) survey responses, school characteristics, and sampling frame information. It is important to note that paradata, such as information on response status in NPSAS:12, particularly those variables that are highly predictive of response but quite unrelated to the survey variables of interest, were excluded from the bias-likelihood model. Candidate variables for the model included:

- highest degree expected,
- parents' level of education,
- age,
- gender,
- number of dependent children,
- income percentile,
- hours worked per week while enrolled,
- school sector,
- undergraduate degree program,
- expected wage, and
- high school graduation year.

Because the variables used in the bias-likelihood model were selected due to their potential ability to act as proxies for survey outcomes, which are unobservable for nonrespondents, the predicted probabilities from the bias-likelihood model were used to identify nonrespondents in the most underrepresented groups, as defined by the variables used in the model. Small predicted probabilities correspond to nonrespondents in the most underrepresented groups, i.e. most likely to contribute to bias, while large predicted probabilities identify groups that are, relatively, well-represented among respondents.

The importance score was defined for nonrespondents as the product of a sample member's *a priori* predicted probability of response and one minus the sample member's predicted bias-likelihood probability. Nonrespondents with the highest calculated importance score at the beginning of Phases 3, 4, and 5, were considered to be most likely to contribute to nonresponse bias and, therefore, were offered the higher monetary incentive increase (Phase 3), were sent to field and local exchange calling (Phase 4), and were offered an abbreviated interview (Phase 5). An overview of the calibration and main sample data collection activities is provided in table 5.

Table 5. Summary of start dates and activities for each phase of the BPS:12/14 data collection, by sample

Phase	Start date		Activity	
	Calibration subsample	Main subsample	Calibration subsample	Main subsample
1	2/18/2014	4/8/2014	Begin web collection; Randomize calibration sample to different baseline incentives (experiment #1)	Begin web collection; baseline incentives determined by results of first calibration experiment
2	3/18/2014	5/6/2014	Begin CATI collection	Begin CATI collection
3	4/8/2014	5/27/2014	Randomize calibration sample nonrespondents to different monetary incentive increases (experiment #2)	Construct importance score and offer incentive increase to select nonrespondents; incentive increase determined by results of second calibration experiment
4	5/6/2014	6/24/2014	Construct importance score and identify select nonrespondents for Field/local exchange calling for targeted cases	Construct importance score and identify select nonrespondents for Field/local exchange calling for targeted cases
5	7/15/2014	9/2/2014	Construct importance score and identify select nonrespondents for abbreviated interview with mobile access	Construct importance score and identify select nonrespondents for abbreviated interview with mobile access
6	8/12/2014	9/30/2014	Abbreviated interview for all remaining nonrespondents	Abbreviated interview for all remaining nonrespondents

CATI = computer-assisted telephone interviewing

Impact on Nonresponse Bias. As all BPS:12/14 sample members were submitted to the same data collection procedures, there is no exact method to assess the degree to which the responsive design reduced nonresponse bias relative to another data collection design that did not incorporate responsive design elements. However, a post-hoc analysis was implemented to compare estimates of nonresponse bias to determine the impact of the responsive design. Nonresponse bias estimates were first created using all respondents and then created again by reclassifying targeted respondents as nonrespondents. This allows examination of the potential bias contributed by the subset of individuals who were targeted by responsive design methods although this is not a perfect design as some of these individuals would have responded without interventions. The following variables were used to conduct the nonresponse bias analysis:[2]

- Region (categorical);
- Age as of NPSAS:12 (categorical);
- CPS match as of NPSAS:12 (yes/no);
- Federal aid receipt (yes/no);
- Pell Grant receipt (yes/no);
- Pell Grant amount (categorical);

[2] For the continuous variables, except for age, categories were formed based on quartiles.

- Stafford Loan receipt (yes/no);
- Stafford Loan amount (categorical);
- Institutional aid receipt (yes/no);
- State aid receipt (yes/no);
- Major (categorical);
- Institution enrollment from IPEDS file (categorical);
- Any grant aid receipt (categorical); and
- Graduation rate (categorical).

For each variable listed above, nonresponse bias was estimated by comparing estimates from base-weighted respondents with those of the full sample to determine if the differences were statistically significant at the 5 percent level. Multilevel categorical terms were examined using indicator terms for each level of the main term. The relative bias estimates associated with these nonresponse bias analyses are summarized in Table 6.

The mean and median percent relative bias are almost universally lowest across all sectors when all respondents are utilized in the bias assessment. The overall percentage of characteristics with significant bias is lowest when all respondents are utilized but the percentage of characteristics with significant bias is lowest in seven of the ten sectors when responsive design respondents are excluded. However, the percentage of characteristics with significant bias is affected by sample sizes and as there are approximately 5,200 respondents who were ever selected under the responsive design, the power to detect a bias that is statistically different from zero is higher when using all respondents versus a smaller subset of those respondents in a nonresponse bias assessment. Consequently, the mean and median percent relative bias are better gauges of how the addition of selected responsive design respondents impacts nonresponse bias.

Given that some of the 5,200 selected respondents would have responded even if they had never been subject to responsive design, it is impossible to attribute the observed bias reduction solely to the application of responsive design methods. However, observed reduction of bias is generally quite large and suggests that responsive design methods may be helpful in reducing nonresponse bias.

Table 6. Summary of responsive design impact on nonresponse bias, by institutional sector: 2014

Nonresponse bias statistics[1]	Overall	Public less-than-2-year	Public 2-year	Public 4-year non-doctorate-granting	Public 4-year doctorate-granting	Private nonprofit less-than-4-year	Private nonprofit 4-year non-doctorate-granting	Private nonprofit 4-year doctorate-granting	Private for-profit less-than-2-year	Private for-profit 2-year	Private for-profit 4-year
All respondents											
Mean percent relative bias across characteristics	10.7	7.8	6.4	8.9	4.2	12.9	7.1	4.7	13.0	9.6	7.2
Median percent relative bias across characteristics	6.3	5.1	3.8	4.6	2.7	8.8	4.0	3.8	7.5	5.4	5.5
Percentage of characteristics with significant bias	62.1	36.4	29.0	43.2	33.3	7.4	32.6	29.3	25.0	8.3	11.7
Respondents excluding those selected for responsive design											
Mean percent relative bias across characteristics	22.9	12.5	12.8	13.6	4.3	29.0	16.5	4.7	12.1	18.2	11.6
Median percent relative bias across characteristics	12.4	9.7	5.0	7.8	3.0	16.5	4.2	3.6	4.7	9.0	8.1
Percentage of characteristics with significant bias	73.6	16.7	26.9	55.6	37.5	32.4	31.8	30.2	17.1	5.7	9.7

[1] Relative bias and significance calculated on respondents vs. full sample.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2012/14 Beginning Postsecondary Students Longitudinal Study (BPS:12/14).

BPS:12/17 Full Scale

The responsive design methods proposed for BPS:12/17 expand and improve upon the BPS:12/14 methods in three key aspects:

1) Refined targeting of nonresponding sample members so that, instead of attempting to reduce unit nonresponse bias for national estimates only, as in BPS:12/14, the impact of unit nonresponse on the bias is reduced for estimates *within institutional sector*.
2) Addition of a special data collection protocol for a hard-to-convert group: NPSAS:12 study member double-interview nonrespondents.
3) Inclusion of a randomized evaluation designed to permit estimating the difference between unit nonresponse bias arising from application of the proposed responsive design methods and unit nonresponse bias arising from not applying the responsive design methods.

As noted previously, the responsive design approach for the BPS:12/14 full scale included (1) use of an incentive calibration study sample to identify optimal monetary incentives, (2) development of an importance measure for identifying nonrespondents for specific interventions, and (3) implementation of a multi-phase data collection period. Analysis of the BPS:12/14 case targeting indicated that institution sector dominated the construction of the importance scores; meaning that nonrespondents were primarily selected by identifying nonrespondents in the sectors with the lowest response rates. For the BPS:12/17 full scale we are building upon the BPS:12/14 full scale responsive design but, rather than selecting nonrespondents using the same approach as in BPS:12/14, we propose targeting nonrespondents within:

- Institution Sector – we will model and target cases within sector groups in an effort to equalize response rates across sectors.
- NPSAS:12 study member double interview nonrespondents – we will use a calibration sample to evaluate two special data collection protocols for this hard-to-convert group, including a special baseline protocol determined by a calibration sample and an accelerated timeline.

We have designed an evaluation of the responsive design so that we can test the impact of the targeted interventions to reduce nonresponse bias versus not targeting for interventions. For the evaluation, we will select a random subset of all sample members to be pulled aside as a *control* sample that will not be eligible for intervention targeting. The remaining sample member cases will be referred to as the *treatment* sample and the targeting methods will be applied to that group.

In the following sections, we will describe the proposed importance measure, sector grouping, and intervention targeting, then describe the approach for the pre-paid and double nonrespondent calibration experiments, and outline how these will be implemented and evaluated in the BPS:12/17 full scale data collection.

The importance measure. In order to reduce nonresponse bias in survey variables by directing effort and resources during data collection, and to minimize the cost associated with achieving this goal, three related conditions have to be met: (1) the targeted cases must be drawn from groups that are under-represented on key survey variable values among those who already responded, (2) their likelihood of participation should not be excessively low or high (i.e., targeted cases who do not respond cannot decrease bias; targeting only high propensity cases can potentially increase the bias of estimates), and (3) targeted cases should be numerous enough to impact survey estimates within domains of interest. While targeting cases based on response propensities may reduce nonresponse bias, bias may be unaffected if the targeted cases are extremely difficult to convert and do not respond to the intervention as desired.

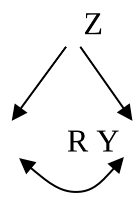
One approach to meeting these conditions is to target cases based on two dimensions: the likelihood of a case to contribute to nonresponse bias if not interviewed, and the likelihood that the case could be converted to a respondent. These dimensions form an importance score, such that:

$$I_{ij} \propto U_{ij} \cdot P(R)_i$$

Where I is the calculated importance score, U is a measure of under-representativeness on key variables that reflects their likelihood to induce bias if not converted, and $P(R)$ is the predicted final response propensity, across i sample members and j data collection phases with responsive design interventions.

The importance score will be determined by the combination of two models: a response propensity model and a bias-likelihood model. Like BPS:12/14, the response propensity component of the importance score is being calculated in advance of the start of data collection. The representativeness of key variables, however, can only be determined during specific phases of the BPS:12/17 data collection, with terms tailored to BPS:12/17. The importance score calculation needs to balance two distinct scenarios: (1) low propensity cases that will likely never respond, irrespective of their underrepresentation, and (2) high propensity cases that, because they are not underrepresented in the data, are unlikely to reduce bias. Once in production, NCES will provide more information about the distribution of both propensity and representation from the BPS:12/17 calibration study, which will allow us to explore linear and nonlinear functions that optimize the potential for nonresponse bias and available incentive resources. We will share the findings with OMB at that time.

Bias-likelihood (U) model. A desirable model to identify cases to be targeted for intervention would use covariates (Z) that are strongly related to the survey variables of interest (Y), to identify sample members who are under-represented (using a response indicator, R) with regard to these covariates. We then have the following relationships, using a single Z and Y for illustration:



Nonresponse bias arises when there is a relationship between R and Y. Just as in adjustment for nonresponse bias (see Little and Vartivarian, 2005), a Z-variable cannot be effective in nonresponse bias reduction if corr(Z,Y) is weak or nonexistent, even if corr(Z,R) is substantial. That is, selection of Z-variables based only on their correlation with R may not help to identify cases that contribute to nonresponse bias. The goal is to identify sample cases that have Y-variable values that are associated with lower response rates, as this is one of the most direct ways to reduce nonresponse bias in an estimate of a mean.

The key Z-variable selection criterion should then be association with Y. Good candidate Z-variables would be the Y-variables or their proxies measured in a prior wave and any correlates of change in estimates over time. A second set of useful Z-variables would be those used in weighting and those used to define subdomains for analysis – such as demographic variables. This should help to reduce the variance inflation due to weighting and nonresponse bias in comparisons across groups. Key, however, is the exclusion of variables that are highly predictive of R, but quite unrelated to Y. These variables, such as the number of prior contact attempts and prior refusal, can dominate in a model predicting the likelihood of participation and mask the relationship of Z variables that are associated with Y.

Prior to the start of later phases of data, when the treatment interventions will be introduced, we will conduct multiple logistic regressions in order to predict the survey outcome (R) through the current phase of collection using only substantive and demographic variables and their correlates from NPSAS:12 and the sampling frame (Z), and select two-way interactions. For each sector grouping (see table 8 below), a single model will be fit. The goal of this model is not to maximize the ability to predict survey response ($\hat{p}$), but to obtain a predicted likelihood of a completed interview reducing nonresponse bias if successfully interviewed. Because of this key difference, we use $(1 - \hat{p})$ to calculate a case-level prediction representing bias-likelihood, rather than response propensity.

Variables to be used in the bias-likelihood model will come from base-year survey responses, institution characteristics, and sampling frame information[3] (see table 7). It is important to note that paradata, particularly those variables that are highly predictive of response, but quite unrelated to the survey variables of interest, will be excluded from the bias-likelihood model.

Table 7. Candidate variables for the bias likelihood model

Variables	
Race	Attendance intensity
Gender	Highest level of education ever expected
Age	Dependent children and marital status
Sector*	Federal Pell grant amount
Match to Central Processing System	Direct subsidized and unsubsidized loans
Match to Pell grant system	Total federal aid
Total income	Institutional aid total
Parent's highest education level	Degree program

* Variable to be included in bias likelihood model for targeting sample members from public 4-year and private nonprofit institutions (sector group B in table 8).

Response propensity (P(R)) model. Prior to the start of BPS:12/17 data collection, a response propensity model is being developed to predict likelihood to respond to BPS:12/17 based on BPS:12/14 data and response behavior. NCES will share the model with OMB when finalized and prior to implementation. The model will use variables from the base NPSAS:12 study as well as BPS:12/14 full scale that have been shown to predict survey response, including, but not limited to:

- responded during early completion period,
- response history,
- interview mode (web/telephone),
- ever refused,
- incentive amount offered,
- age,
- gender,
- citizenship,
- institution sector,
- call count, and
- tracing/locating status (located/required intensive tracing).

We will use BPS:12/14 full scale data to create this response propensity model as that study was similar in design and population to the current BPS:12/17 full scale study (note that BPS:12/17 did not have a field test that could be leveraged, and the pilot study was too limited in size and dissimilar in approach and population to be useful for this purpose).

Targeted interventions. In the BPS:12/14 responsive design approach, institution sector was the largest factor in determining current response status. For BPS:12/17 full scale, individuals will be targeted within groupings of institution sectors in an effort to equalize response rates across the sector groups. Designed to reduce the final unequal weighting effect, targeting within the groups will allow us to fit a different propensity or bias likelihood model for each group while equalizing response rates across groups.

Targeting within sector groups is designed to reduce nonresponse bias within specific sectors rather than across the aggregate target population. The five sector groupings (Table 8) were constructed by first identifying sectors with historically low response rates, as observed in BPS:12/14 and NPSAS:12, and, second, assigning the sectors with the lowest participation to their own groups. The remaining sectors were then combined into groups consisting of multiple sectors. The private for profit sectors (groups C, D, and E) were identified to have low response rates. Public less-than-2-year and public 2-year institutions (group A) were combined as they were similar, and because the public less-than-2-year sector was too small to act as a distinct group. Public 4-year and private nonprofit institutions (sector group B) remained combined as they

[3] Key variables will use imputed data to account for nonresponse in the base year data.

have not historically exhibited low response rates (nonetheless, cases within this sector group are still eligible for targeting; the targeting model for sector group B will include sector as a term to account for differences between the sectors).

Table 8. Targeted sector groups

Sector Group	Sectors	Sample Count
A	1: Public less-than-2-year 2: Public 2-year	205 10,142
B	3: Public 4-year non-doctorate-granting 4: Public 4-year doctorate-granting 5: Private nonprofit less than 4-year 6: Private nonprofit 4-year nondoctorate 7: Private nonprofit 4-year doctorate-granting	1,829 3,398 334 2,283 2,602
C	8: Private for profit less-than-2-year	1,463
D	9: Private for profit 2-year	3,132
E	10: Private for profit 4-year	8,340

All NPSAS:12 study members who responded to the NPSAS:12 or BPS:12/14 student interviews (hereafter called previous respondents) will be initially offered a $30 incentive, determined to be an optimal baseline incentive offer during the BPS:12/14 Phase 1 experiment with the calibration sample. Following the $30 baseline offer, two different targeted interventions will be utilized for the BPS:12/17 responsive design approach:

- First Intervention (Incentive Boost): Targeted cases will be offered an additional $45 over an individual's baseline incentive amount. The $45 amount is based on the amount identified as optimal during Phase 3 of the BPS:12/14 calibration experiment.
- Second Intervention (Abbreviated Interview): Targeted cases will be offered an abbreviated interview at 21 weeks (note that all cases will be offered abbreviated interview at 31 weeks).

Before each targeted intervention, predicted bias-likelihood values and composite propensity scores will be calculated for all interview nonrespondents. The product of the bias-likelihood and response propensity will be used to calculate the target importance score described above. Propensity scores above high and low cutoffs, determined by a review of the predicted distribution, will be excluded as potential targets during data collection[4].

Pre-paid calibration experiment. It is widely accepted that survey response rates have been in decline in the last decade. Incentives, and in particular prepaid incentives, can often help maximize participation. BPS will test a prepaid incentive, delivered electronically in the form of a PayPal[5] payment, to selected sample members. Prior to the start of full-scale data collection, 2,970 members of the previous respondent main sample will be identified to participate in a calibration study to evaluate the effectiveness of the pre-paid PayPal offer. At the conclusion of this randomized calibration study, NCES will meet with OMB to discuss the results of the experiment and to seek OMB approval through a change request for the pre-paid offer for the remaining nonrespondent sample. Half of the calibration sample will receive a $10 pre-paid PayPal amount and an offer to receive another $20 upon completion of the survey ($30 total). The other half will receive an offer for $30 upon completion of the survey with no pre-paid amount. At six weeks the response rates for the two approaches will be compared to determine if the pre-paid offer should be extended to the main sample. For all monetary incentives, including prepayments, sample members have the option of receiving disbursements through PayPal or in the form of a check.

[4] These adjustments will help ensure that currently over-represented groups, high propensity/low importance cases, and very-difficult-to-convert nonrespondents are not included in the target set of nonrespondents. The number of targeted cases will be determined by BPS staff during the phased data collection and will be based on the overall and within sector distributions of importance scores.

[5] A prepaid check will be mailed to sample members who request it. Sample members can also open a PayPal account when notified of the incentive. Any prepaid sample member who neither accepts the prepaid PayPal incentive nor check would receive the full incentive amount upon completion by the disbursement of their choice (i.e. check or PayPal).

During the calibration phase in March 2017, PayPal compliance notified RTI that three sample members designated to be given a pre-paid incentive were flagged as persons possibly sanctioned by U.S. Department of the Treasury's Office of Foreign Assets Control (OFAC). To comply with OFAC sanctions and to ensure the BPS:12/17 PayPal account remained in good standing, RTI began implementing methods to identify sample members who may match to those listed on OFAC's Specially Designated Nationals and Blocked Persons (SDN) list. Programmatic matching, using methods recommended in the OFAC's Web-based Sanction List Search tool, was performed on the entire BPS:12/17 fielded sample (n=33,750). This matching process resulted in 345 potential matches between the BPS:12/17 sample and the SDN list. The 345 cases were manually reviewed against additional sources of information. Of these cases, 315 individuals were ruled out as matches to individuals on the OFAC SDN list. The remaining 30 cases in the main and calibration samples could not be ruled out as sanctioned individuals. To comply with OFAC requirements and to avoid compliance issues with PayPal, the 315 individuals will be offered an incentive by check only. The remaining 30 cases will not be fielded in BPS:12/17 and will be excluded from the survey. Of those excluded, 3 cases were in the calibration sample, and 27 cases were in the main sample.

After the calibration experiment detailed above, the calibration sample will join with the main sample to continue data collection efforts. These are described in detail below, summarized in table 9, and their timeline is shown graphically in figure 1.

Table 9. Timeline for previous respondents

Phase	Start date		Activity	
	Calibration sample	Main sample	Calibration sample	Main sample
PR-1	Week 0	Week 7	Begin data collection; calibration sample for $10 pre-paid offer versus no pre-paid offer	Begin data collection; make decision on implementation of pre-paid offer based on results of calibration
PR-2	Week 14	Week 14	Target treatment cases for incentive boost	Target treatment cases for incentive boost
PR-3	Week 21	Week 21	Target treatment cases for early abbreviated interview	Target treatment cases for early abbreviated interview
PR-4	Week 31	Week 31	Abbreviated interview for all remaining nonrespondents	Abbreviated interview for all remaining nonrespondents

Special data collection protocol for double nonrespondents. Approximately 3,280 sample members (group 4 in table 2) had sufficient information in NPSAS:12 to be classified as a NPSAS:12 study member but have neither responded to the NPSAS:12 student interview nor the BPS:12/14 student interview (henceforth referred to as *double nonrespondents*). In planning for the BPS:12/17 collection, we investigated characteristics known about this group, such as the distribution across sectors, our ability to locate them in prior rounds, and their estimated attendance and course-taking patterns using PETS:09. We found that while this group constitutes approximately 10 percent of the sample, 58 percent of double nonrespondents were enrolled within the private for-profit sectors in NPSAS:12. We found that over three-quarters of double nonrespondents had been contacted—yet had not responded. We also found, using a proxy from the BPS:04 cohort, that double nonrespondents differed by several characteristics of prime interest to BPS, such as postsecondary enrollment and coursetaking patterns. We concluded that double nonrespondents could contribute to nonresponse bias, particularly in the private for-profit sector.

While we were able to locate approximately three-quarters of these double nonrespondents in prior data collections, we do not know their reasons for refusing to participate. Without knowing the reasons for refusal, the optimal incentives are difficult to determine. In BPS:12/14, due to the design of the importance score which excluded the lowest propensity cases, nonrespondents who were the most difficult to convert were not included in intervention targeting. As a result, very few of the double nonrespondents were ever exposed to incentive boosts or early abbreviated interviews in an attempt to convert them. In fact, after examining BPS:12/14 data, we found that less than 0.1 percent were offered more than $50 dollars and only 3.6 percent were offered more than $30. Similarly, we do not know if a shortened abbreviated interview would improve response rates for this group. Therefore, we propose a calibration sample with an experimental design that evaluates the efficacy of additional incentive versus a shorter interview. The results of the experiment will inform the main sample.

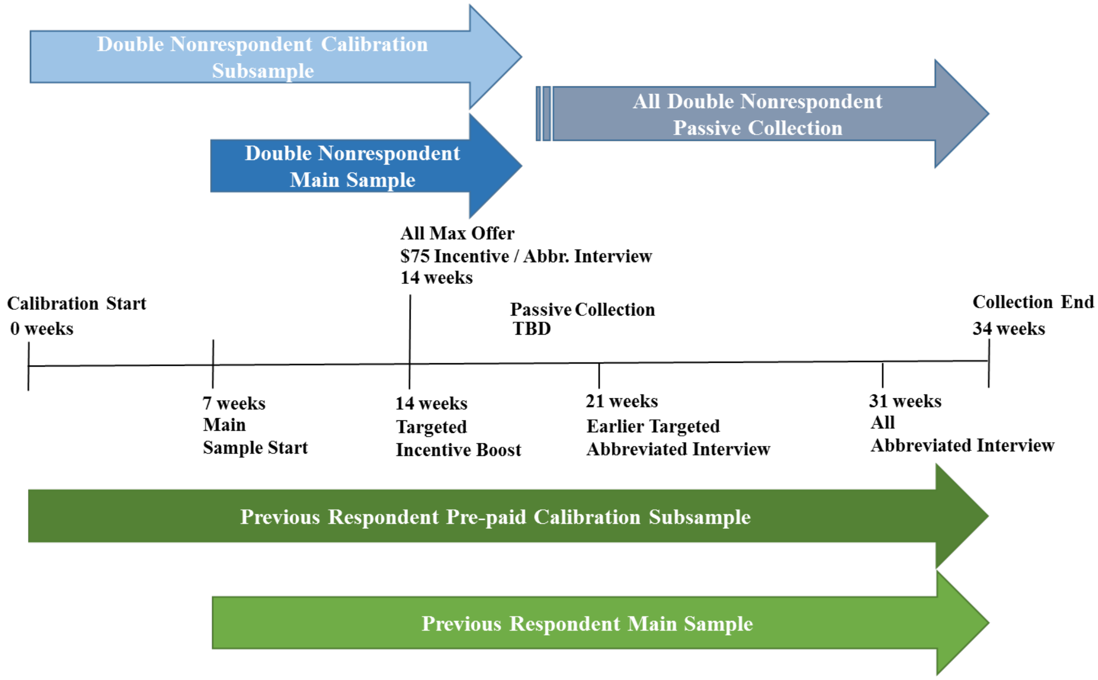
Specifically, we propose fielding a calibration sample, consisting of 869 double nonrespondents, seven weeks ahead of the main sample to evaluate the two special data collection protocols for this hard-to-convert group: a shortened interview vs. a monetary incentive. A randomly-selected half of the calibration sample will be offered an abbreviated interview along with a $10 pre-paid PayPal amount and an offer to receive another $20 upon completion of the survey ($30 total). The other half will be offered the full interview along with a $10 pre-paid PayPal amount[6] and an offer to receive another $65 upon completion of the survey ($75 total). At six weeks, the two approaches will be compared using a Pearson Chi-squared test to determine which results in the highest response rate from this hard-to-convert population and should be proposed for the main sample of double nonrespondents. If both perform equally, we will select the $30 total baseline along with the abbreviated interview. Regardless of the selected protocol, at 14 weeks into data collection, all remaining nonrespondents in the double nonrespondent population will be offered the maximum special protocol intervention consisting of an abbreviated interview and $65 upon completion of the interview to total $75 along with the $10 pre-paid offer. In addition, at a later phase of data collection, we will move this group to a passive status by discontinuing CATI operations and relying on email contacts. The timeline for double nonrespondents is summarized in table 10 and figure 1.

Table 10. Timeline for NPSAS:12 study member double interview nonrespondents

Phase	Start date		Activity	
	Calibration sample	Main sample	Calibration sample	Main sample
DNR-1	Week 0	Week 7	Begin data collection; calibration sample for baseline special protocol (full interview and $75 total vs. abbreviated interview and $30 total)	Begin data collection; baseline special protocol determined by calibration results (full interview and $75 total vs. abbreviated interview and $30 total)
DNR-2	Week 14	Week 14	Offer all remaining double nonrespondents $75 incentive and abbreviated interview	Offer all remaining double nonrespondents $75 incentive and abbreviated interview
DNR-3	Week TBD	Week TBD	Move to passive data collection efforts for all remaining nonrespondents; time determined bases on sample monitoring	Move to passive data collection efforts for all remaining nonrespondents; time determined bases on sample monitoring

Figure 1. Data collection timeline

[6] All double nonrespondents will be offered the $10 pre-paid PayPal amount in an attempt to convert them to respondents. For all monetary incentives, including prepayments, sample members have the option of receiving disbursements through PayPal or in the form of a check.



Evaluation of the BPS:12/17 Responsive Design Effort. The analysis plan is based upon two premises: (1) offering special interventions to some, targeted, sample members will increase participation in the aggregate for those sample members and (2) increasing participation among the targeted sample members will produce estimates with lower bias than if no targeting were implemented. In an effort to maximize the utility of this research, the analysis of the responsive design and its implementation will be described in a technical report that includes these two topics and their related hypotheses described below. We intend to examine these aspects of the BPS:12/17 responsive design and its implementation as follows:

1. *Evaluate the effectiveness of the calibration samples in identifying optimal intervention approaches to increase participation.*

A key component of the BPS:12/17 responsive design is the effectiveness of the changes in survey protocol for increasing participation. The two calibration experiments examine the impact of proposed features – a pre-paid PayPal offer for previous respondents and two special protocols for double nonrespondents.

Evaluation of the experiments with calibration samples will occur during data collection so that findings can be implemented in the main sample data collection. Approximately six weeks after the start of data collection for the calibration sample, response rates for the calibration *pre-paid offer* group versus the *no pre-paid offer* group for previous respondents will be compared using a Pearson chi-square test. Similarly the double nonrespondent group receiving the abbreviated interview plus the total $30 offer will be compared to the group receiving the abbreviated interview plus the total $75 offer.

2. *Evaluate sector group level models used to target cases for special interventions*

To maximize the effectiveness of the BPS:12/17 responsive design approach, targeted cases need to be associated with survey responses that are underrepresented among the respondents, and the targeted groups need to be large enough to change observed estimates. In addition to assessing model fit metrics and the effective identification of cases contributing to nonresponse bias for each of the models used in the importance score calculation, the distributions of the targeted cases will be reviewed for key variables, overall and within sector, prior to identifying final targeted cases. Again, these key variables include base-year survey responses, institution characteristics, and sampling frame information as shown in table 7. During data collection, these reviews will help ensure that the cases most likely to decrease bias are targeted and that project resources are used efficiently. After data collection, similar summaries will be used to describe the composition of the targeted cases along dimensions of interest.

The importance score used to select targeted cases will be calculated based on both the nonresponse bias potential and on an *a priori* response propensity score. To evaluate how well the response propensity measure predicted actual response, we will compare the predicted response rates to observed response rates at the conclusion of data collection. These comparisons will be made at the sector group level as well as in aggregate.

3. *Evaluate the ability of the targeted interventions to reduce unit nonresponse bias through increased participation.*

To test the impact of the targeted interventions to reduce nonresponse bias versus not targeting for interventions, we require a set of similar cases that are held aside from the targeting process. A random subset of all sample members will be pulled aside as a control sample that is not eligible for intervention targeting. The remaining sample member cases will be referred to as the treatment sample and the targeting methods will be applied to that group. Sample members will be randomly assigned to the control group within each of the five sector groups. In all, the control group will be composed of 3,606 individuals (approximately 721 per sector group) who form nearly 11 percent of the total fielded sample.

For evaluation purposes, the targeted interventions will be the only difference between the control and treatment samples. Therefore both the control and treatment samples will consist of previous round respondents and double nonrespondents, and they will both be involved in the calibration samples and will both follow the same data collection timelines.

The frame, administrative, and prior-round data used in determining cases to target for unit nonresponse bias reduction can, in turn, be used to evaluate (1) unit nonresponse bias in the final estimates and (2) changes in unit nonresponse bias over the course of data collection. Unweighted and weighted (using design weights) estimates of absolute nonresponse bias will be computed for each variable used in the models:

$$\left| \bar{y}_r - \bar{y}_s \right|$$

where $\bar{y}_r$ is the respondent mean and $\bar{y}_s$ is the full sample mean. Bias estimates will be calculated separately for treatment and control groups and statistically compared under the hypothesis that treatment interventions yields estimates with lower bias.

BPS:12/17 Responsive Design Research Questions. With the assumption that increasing the rate of response among targeted, underrepresented cases will reduce nonresponse bias, the BPS:12/17 responsive design experiment will explore the following research questions which may be stated in terms of a null hypothesis as follows:

- <u>Research question 1</u>: Did the *a priori* response propensity model predict overall unweighted BPS:12/17 response?
  - H_0: At the end of data collection, there will be no association between *a priori* propensity predictions and observed response rates.
- <u>Research question 2</u>: Does one special protocol increase response rates for double nonrespondents versus the other protocol?
  - H_0: At the end of the double nonrespondent calibration sample, there will be no difference in response rates between a $75 baseline offer along with a full interview and a $30 baseline offer along with an abbreviated interview.
- <u>Research question 3</u>: Does a $10 pre-paid PayPal offer increase early response rates?
  - H_0: At the end of the previous respondent calibration sample, there will be no difference in response rates between cases that receive a $10 pre-paid PayPal offer and those that do not.
- <u>Research question 4</u>: Are targeted respondents different from non-targeted respondents on key variables?

- o H_0: Right before the two targeted interventions, and the end of data collection, there will be no difference between targeted respondents and non-targeted and never-targeted respondents in weighted or unweighted estimates of key variables not included in the importance score calculation.

- Research question 5: Did targeted cases respond at higher rates than non-targeted cases?
  - o H_0: At the end of the targeted interventions, and at the end of data collection, there will be no difference in weighted or unweighted response rates between the treatment sample and the control sample.

- Research question 6: Did conversion of targeted cases reduce unit nonresponse bias?
  - o H_0: At the end of data collection, there will be no difference in absolute nonresponse bias of key estimates between the treatment and control samples.

Power calculations. The first step in the power analysis was to determine the number of sample members to allocate to the control and treatment groups. For each of the five institution sector groups, roughly 721 sample members will be randomly selected into the control group that will not be exposed to any targeted interventions. The remaining sample within each sector group will be assigned to the treatment group. We will then compare absolute measures of bias between the treatment and control groups under the hypothesis that the treatments, that is, the targeted interventions, reduce absolute bias. As we will be comparing absolute bias estimates, which range between zero and one, a power analysis was conducted using a one-sided, two-group chi-square test of equal proportions with unequal sample sizes in each group. The absolute bias estimates will be weighted and statistical comparisons will take into account the underlying BPS:12/17 sampling design; therefore, the power analysis assumes a relatively conservative design effect of 3. Table 11 shows the resulting power based on different assumptions for the base absolute bias estimates.

Table 11. Power for control versus treatment comparisons across multiple assumptions

				Alpha	**0.05**	**0.05**	**0.05**
				Treatment Abs. Bias	**0.4**	**0.2**	**0.125**
				Control Abs. Bias	**0.5**	**0.3**	**0.2**
Sector Group	**Sectors**	**Total Count**	**Control Sample**	**Treatment Sample**	**Unequal Group Power**	**Unequal Group Power**	**Unequal Group Power**
A	1: Public less-than-2-year 2: Public 2-year	10,345	721	9,624	0.915	0.966	0.924
B	3: Public 4-year non-doctorate-granting 4: Public 4-year doctorate-granting 5: Private nonprofit less than 4-year 6: Private nonprofit 4-year nondoctorate 7: Private nonprofit 4-year doctorate-granting	10,445	721	9,724	0.916	0.966	0.924
C	8: Private for-profit less-than-2-year	1,463	722	741	0.718	0.819	0.727
D	9: Private for-profit 2-year	3,132	721	2,411	0.864	0.935	0.876
E	10: Private for-profit 4-year	8,340	721	7,619	0.911	0.963	0.920

NOTE: After sampling for BPS:12/17, three cases were determined to be deceased, and have been removed from the power calculations. In addition, this table does not include 30 cases that were excluded based on matching to the OFAC SDN list.

The final three columns of table 11 show how the power estimates vary depending upon the assumed values of the underlying absolute bias measures, and the third to last column specifically shows the worst case scenario where the bias measures are 50 percent. The overall control sample size is driven by sector group C which has the lowest available sample, and for some bias domains we may need to combine sectors C and D for analysis purposes. Given the sensitivity of power estimates to assumptions regarding the underlying treatment effect, there appears to be sufficient power to support the proposed calibration experiment across a wide range of possible scenarios.

After the assignment of sample members to treatment and control groups, we will construct the two calibration samples: 1) previous respondents and 2) double nonrespondents. The calibration sample of previous respondents (n=2,970) will be randomly split into two groups, with 1,486 sample members in the treatment group and 1,484 in the control group[7]. One group will receive a $10 pre-paid offer while the other will not receive the pre-paid offer. For a power of 0.80, a confidence level of 95 percent, and given the sample within each condition, the experiment of pre-paid amounts should detect a 5.0 percentage point difference in response rate using Pearson's chi-square[8]. This power calculation assumes a two-sided test of proportions as we are uncertain of the effect of offering the pre-paid incentive. In addition, for the power calculation, an initial baseline response rate of 36.0 percent was selected given that this was the observed response rate after six weeks of data collection for the BPS:12/14 full scale study.

Similarly, we will randomly split the calibration sample of double nonrespondents (n=869)[9] into two groups where one group (n=435) will receive a $30 baseline offer with an abbreviated interview, while the other group (n=434) will be offered the full interview and a total of $75. Using a power of 0.80, a confidence level of 95 percent, and given sample available within each condition, the experiment among double nonrespondents should detect a 5.0 percent point difference in response rate using Pearson's chi-square. This power calculation assumes a two-sided test of proportions, as we have no prior data for which special protocol will perform better with respect to response rate. For this calculation, we assumed the six week response rate for the protocol with the lower response rate would be 5.0 percent. For a test between two proportions, the power to detect a difference is dependent on the initially assumed baseline response rate. In the proposed scenario, with a one-sided test, the baseline response rate would be the response rate for the approach with the lower response rate at six weeks. We assumed that this response rate would be 5% for power calculations. If the six week response rate differs from 5% then the power of the test to detect a 5% difference in response rates would fluctuate as shown in table 12 below.

Table 12. Power to detect 5 percent difference in response rate based on different baseline response rates

Alpha 0.05, Sample Size per group = 435, Detectable Difference 5%	
Assumed Response Rate	**Power to Detect Difference**
1%	98%
5%	80%
10%	61%

[7] After 30 cases were excluded from the BPS:12/17 sample to comply with OFAC sanctions, two cases that were originally assigned to the previous respondent control group were removed.
[8] Calculated using SAS Proc Power. https://support.sas.com/documentation/cdl/en/statugpower/61819/PDF/default/statugpower.pdf
[9] After 30 cases were excluded from the BPS:12/17 sample to comply with OFAC sanctions, one case that was originally assigned to the double nonrespondent treatment group that received the $75 offer for a full interview was removed.

Attachment B. *A priori* propensity and bias-likliehood model variables

***A priori* propensity model variables**

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	1.6798	0.0822	418.0648	<.0001
AGE		1	-0.0168	0.00210	64.5708	<.0001
GENDER	2	1	0.1411	0.0290	23.7028	<.0001
citizen	2	1	0.1890	0.0882	4.5928	0.0321
citizen	3	1	-1.0143	0.1245	66.4115	<.0001
citizen	-9	1	-0.4604	0.0856	28.9074	<.0001
emailcount		1	0.0926	0.0234	15.6244	<.0001
early_comp	1	1	0.5189	0.0321	261.1316	<.0001
ever_any_ref	1	1	-0.2182	0.0496	19.3658	<.0001
SECTOR10	1	1	0.0792	0.1674	0.2237	0.6362
SECTOR10	3	1	-0.0405	0.0691	0.3437	0.5577
SECTOR10	4	1	0.1681	0.0593	8.0302	0.0046
SECTOR10	5	1	-0.2284	0.1363	2.8074	0.0938
SECTOR10	6	1	0.0930	0.0667	1.9457	0.1631
SECTOR10	7	1	0.3833	0.0691	30.7782	<.0001
SECTOR10	8	1	-0.4017	0.0695	33.3545	<.0001
SECTOR10	9	1	0.0431	0.0520	0.6859	0.4076
SECTOR10	10	1	0.00710	0.0382	0.0345	0.8526
NPSAS12RESP	0	1	-2.4365	0.0605	1623.6972	<.0001
addr_update	1	1	1.6727	0.1045	256.2360	<.0001
inc_amount		1	-0.0128	0.000814	245.9575	<.0001
leaf2[1]	1	1	0.5232	0.1144	20.9059	<.0001
leaf3[2]	1	1	-0.6254	0.0793	62.2445	<.0001

[1] - Leaf2 = Interaction of NPSAS12RESP (0-nonrespondent), Age (<20), Sector10 (4,6,7)

[2] - Leaf3 = Interaction of NPSAS12RESP (0-nonrespondent), Age (>=20), Addre_update (0-not updated)

Bias-likelihood model variables, institution groups A-E

Bias-likelihood model variables: institution group A (as of 6/27/17)

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.0259	0.1042	96.8996	<.0001
RACE	2	1	0.0516	0.0641	0.6469	0.4212
RACE	3	1	0.0685	0.0606	1.2743	0.2590
RACE	4	1	0.2107	0.1102	3.6574	0.0558
RACE	5	1	-0.5767	0.2608	4.8891	0.0270
RACE	6	1	-0.2460	0.3109	0.6261	0.4288
RACE	7	1	0.1392	0.1172	1.4111	0.2349
GENDER	2	1	0.5671	0.0443	163.8508	<.0001
AGE3	2	1	0.00223	0.0602	0.0014	0.9705
AGE3	3	1	0.1405	0.0907	2.3975	0.1215
INCPS_n	0	1	-0.2485	0.0669	13.7962	0.0002
PELLMATCH_n	0	1	-0.2125	0.1130	3.5362	0.0600
CINCOME4	2	1	0.0783	0.0611	1.6395	0.2004
CINCOME4	3	1	0.1480	0.0658	5.0683	0.0244
CINCOME4	4	1	0.1691	0.0857	3.8978	0.0483
PAREDUC	0	1	-0.0478	0.1095	0.1906	0.6624
PAREDUC	1	1	0.0934	0.0806	1.3407	0.2469
PAREDUC	3	1	0.2578	0.1042	6.1242	0.0133
PAREDUC	4	1	0.0875	0.0820	1.1382	0.2860
PAREDUC	5	1	0.2139	0.0657	10.6101	0.0011
PAREDUC	6	1	0.1623	0.0700	5.3685	0.0205
PAREDUC	7	1	0.1234	0.0933	1.7516	0.1857
PAREDUC	8	1	0.1476	0.2112	0.4886	0.4846
PAREDUC	9	1	0.3356	0.2349	2.0409	0.1531
ATTNPTRN	2	1	0.0355	0.0546	0.4216	0.5161
ATTNPTRN	3	1	0.0367	0.0571	0.4132	0.5203
HIGHLVEX	1	1	-0.4387	0.5332	0.6769	0.4107
HIGHLVEX	2	1	-0.0901	0.1267	0.5063	0.4768
HIGHLVEX	4	1	0.0837	0.0561	2.2239	0.1359
HIGHLVEX	6	1	0.1739	0.0660	6.9412	0.0084
HIGHLVEX	7	1	0.1821	0.1020	3.1877	0.0742
HIGHLVEX	8	1	0.3745	0.1126	11.0570	0.0009
DEPNUMCH3	1	1	-0.1619	0.0929	3.0346	0.0815
DEPNUMCH3	2	1	-0.3020	0.1105	7.4630	0.0063
DEPNUMCH3	3	1	-0.6218	0.1392	19.9523	<.0001
pellamt3	0	1	0.2925	0.1182	6.1279	0.0133
pellamt3	2	1	0.1101	0.0651	2.8604	0.0908
totaid4	0	1	-0.4324	0.0949	20.7543	<.0001
totaid4	2	1	0.1320	0.0703	3.5225	0.0605
totaid4	3	1	0.0846	0.1133	0.5583	0.4550
totaid4	4	1	0.00603	0.2254	0.0007	0.9787
tfedaid4	0	1	0.3489	0.1208	8.3443	0.0039
tfedaid4	2	1	-0.0326	0.0728	0.1999	0.6548
tfedaid4	3	1	0.0111	0.1021	0.0117	0.9138
tfedaid4	4	1	0.000997	0.1745	0.0000	0.9954
instamt3	0	0	0	.	.	.
instamt3	2	1	0.1322	0.0676	3.8240	0.0505
instamt3	3	1	0.4358	0.4216	1.0681	0.3014
UGDEG	1	1	0.0155	0.0865	0.0319	0.8582
UGDEG	3	1	-0.0570	0.1180	0.2336	0.6289
UGDEG	4	1	0.2495	0.1656	2.2702	0.1319
JOBHOUR4	1	1	0.1178	0.0658	3.2093	0.0732
JOBHOUR4	3	1	0.0715	0.0632	1.2803	0.2579
JOBHOUR4	4	1	-0.1056	0.0697	2.2977	0.1296

Bias-likelihood model variables: institution group B (as of 6/27/17)

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-0.9033	0.1999	20.4103	<.0001
RACE	2	1	-0.1231	0.0690	3.1817	0.0745
RACE	3	1	-0.0702	0.0657	1.1442	0.2848
RACE	4	1	-0.0154	0.0844	0.0335	0.8548
RACE	5	1	-0.0660	0.2257	0.0856	0.7699
RACE	6	1	-0.9153	0.3764	5.9141	0.0150
RACE	7	1	-0.0692	0.1020	0.4607	0.4973
GENDER	2	1	0.3135	0.0419	56.0557	<.0001
AGE3	2	1	-0.1817	0.1078	2.8421	0.0918
AGE3	3	1	-0.00858	0.1860	0.0021	0.9632
INCPS_n	0	1	-0.2609	0.0669	15.2235	<.0001
PELLMATCH_n	0	1	-0.0510	0.1175	0.1885	0.6642
CINCOME4	2	1	0.1008	0.0860	1.3745	0.2410
CINCOME4	3	1	0.2008	0.0864	5.4082	0.0200
CINCOME4	4	1	0.2432	0.1003	5.8758	0.0154
PAREDUC	0	1	-0.0622	0.1537	0.1640	0.6855
PAREDUC	1	1	0.2473	0.1213	4.1539	0.0415
PAREDUC	3	1	0.2475	0.1239	3.9910	0.0457
PAREDUC	4	1	0.2172	0.0936	5.3850	0.0203
PAREDUC	5	1	0.2774	0.0779	12.6857	0.0004
PAREDUC	6	1	0.2612	0.0697	14.0573	0.0002
PAREDUC	7	1	0.4257	0.0772	30.3694	<.0001
PAREDUC	8	1	0.4816	0.1081	19.8550	<.0001
PAREDUC	9	1	0.3486	0.1171	8.8627	0.0029
ATTNPTRN	2	1	-0.1255	0.1193	1.1070	0.2927
ATTNPTRN	3	1	0.0530	0.0652	0.6598	0.4166
HIGHLVEX	1	1	1.6163	1.3201	1.4991	0.2208
HIGHLVEX	2	1	0.6535	0.3265	4.0068	0.0453
HIGHLVEX	4	1	0.1979	0.1736	1.2995	0.2543
HIGHLVEX	6	1	0.3032	0.1757	2.9790	0.0844
HIGHLVEX	7	1	0.3551	0.1826	3.7832	0.0518
HIGHLVEX	8	1	0.4328	0.1831	5.5889	0.0181
DEPNUMCH3	1	1	-0.0900	0.1976	0.2078	0.6485
DEPNUMCH3	2	1	0.0950	0.2412	0.1552	0.6936
DEPNUMCH3	3	1	0.0822	0.2994	0.0753	0.7838
pellamt3	0	1	-0.0412	0.1261	0.1067	0.7440
pellamt3	2	1	0.1503	0.0780	3.7104	0.0541
totaid4	0	1	-0.5655	0.1069	27.9891	<.0001
totaid4	2	1	0.1169	0.0883	1.7496	0.1859
totaid4	3	1	0.1592	0.0923	2.9741	0.0846
totaid4	4	1	0.1893	0.1009	3.5200	0.0606
tfedaid4	0	1	0.5269	0.1275	17.0899	<.0001
tfedaid4	2	1	0.00438	0.0744	0.0035	0.9531
tfedaid4	3	1	0.00981	0.0844	0.0135	0.9075
tfedaid4	4	1	-0.2456	0.0888	7.6503	0.0057
instamt3	0	0	0	.	.	.
instamt3	2	1	0.0577	0.0623	0.8590	0.3540
instamt3	3	1	0.0719	0.0701	1.0541	0.3046
UGDEG	1	1	-0.0149	0.2250	0.0044	0.9474
UGDEG	3	1	0.1665	0.1162	2.0531	0.1519
UGDEG	4	1	0.2233	0.4325	0.2667	0.6056
JOBHOUR4	1	1	0.1811	0.0488	13.7845	0.0002
JOBHOUR4	3	1	-0.1396	0.1023	1.8625	0.1723
JOBHOUR4	4	1	-0.1347	0.1120	1.4465	0.2291
SECTOR10	4	1	-0.0236	0.0650	0.1319	0.7165

SECTOR10	5	1	-0.3808	0.1605	5.6308	0.0176
SECTOR10	6	1	-0.0440	0.0783	0.3164	0.5738
SECTOR10	7	1	0.0504	0.0759	0.4419	0.5062

Bias-likelihood model variables: institution group C (as of 6/27/17)

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.5376	1.0567	2.1171	0.1457
RACE	2	1	0.3631	0.1820	3.9777	0.0461
RACE	3	1	0.2402	0.1639	2.1474	0.1428
RACE	4	1	0.3421	0.5193	0.4339	0.5101
RACE	5	1	-0.1430	0.6052	0.0559	0.8132
RACE	6	1	-0.7105	1.1265	0.3978	0.5283
RACE	7	1	-0.3348	0.3727	0.8072	0.3689
GENDER	2	1	0.4980	0.1737	8.2240	0.0041
AGE3	2	1	-0.1916	0.1577	1.4767	0.2243
AGE3	3	1	0.0203	0.2223	0.0083	0.9274
INCPS_n	0	1	-0.5218	0.3645	2.0498	0.1522
PELLMATCH_n	0	1	0.4108	0.5478	0.5623	0.4533
CINCOME4	2	1	0.0646	0.1548	0.1744	0.6763
CINCOME4	3	1	0.1513	0.1944	0.6058	0.4364
CINCOME4	4	1	-0.2507	0.3805	0.4340	0.5100
PAREDUC	0	1	-0.4540	0.3024	2.2540	0.1333
PAREDUC	1	1	-0.0469	0.1948	0.0579	0.8099
PAREDUC	3	1	-0.0328	0.2788	0.0138	0.9064
PAREDUC	4	1	0.0794	0.2956	0.0722	0.7881
PAREDUC	5	1	0.6824	0.2095	10.6073	0.0011
PAREDUC	6	1	0.0693	0.2416	0.0823	0.7742
PAREDUC	7	1	-0.1917	0.3432	0.3121	0.5764
PAREDUC	8	1	0.6872	0.4894	1.9715	0.1603
PAREDUC	9	1	0.6493	0.8171	0.6313	0.4269
ATTNPTRN	2	1	0.1293	0.2676	0.2335	0.6289
ATTNPTRN	3	1	-0.0533	0.2079	0.0658	0.7976
HIGHLVEX	1	1	27.7710	1514.9	0.0003	0.9854
HIGHLVEX	2	1	-0.0700	0.1794	0.1520	0.6966
HIGHLVEX	4	1	-0.2351	0.2118	1.2314	0.2671
HIGHLVEX	6	1	-0.0502	0.2652	0.0358	0.8500
HIGHLVEX	7	1	-0.1125	0.4761	0.0558	0.8133
HIGHLVEX	8	1	0.4040	0.3591	1.2663	0.2605
DEPNUMCH3	1	1	0.1950	0.1794	1.1820	0.2770
DEPNUMCH3	2	1	-0.1903	0.2477	0.5900	0.4424
DEPNUMCH3	3	1	-0.3923	0.2903	1.8252	0.1767
pellamt3	0	1	-0.0339	0.5156	0.0043	0.9476
pellamt3	2	1	0.0170	0.1681	0.0103	0.9192
totaid4	0	1	-0.5387	0.4242	1.6128	0.2041
totaid4	2	1	-0.1602	0.4150	0.1491	0.6994
totaid4	3	1	-0.0927	0.4265	0.0473	0.8279
totaid4	4	1	-0.3059	0.4412	0.4805	0.4882
tfedaid4	0	1	1.1994	0.5028	5.6904	0.0171
tfedaid4	2	1	0.0424	0.4173	0.0103	0.9190
tfedaid4	3	1	0.3168	0.4491	0.4974	0.4806
tfedaid4	4	1	0.4032	0.4646	0.7530	0.3855
instamt3	0	0	0	.	.	.
instamt3	2	1	0.2422	0.5088	0.2267	0.6340
instamt3	3	1	-12.9676	859.4	0.0002	0.9880
UGDEG	1	1	0.1396	0.9703	0.0207	0.8856
UGDEG	4	1	-13.0402	730.2	0.0003	0.9858
JOBHOUR4	1	1	0.5424	0.2246	5.8290	0.0158
JOBHOUR4	3	1	0.2143	0.2000	1.1484	0.2839

JOBHOUR4	4	1	0.5010	0.2259	4.9158	0.0266

Bias-likelihood model variables: institution group D (as of 6/27/17)

Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.4541	0.2974	23.8999	<.0001
RACE	2	1	0.1647	0.1301	1.6027	0.2055
RACE	3	1	0.1177	0.1044	1.2709	0.2596
RACE	4	1	0.9713	0.3514	7.6386	0.0057
RACE	5	1	-0.1897	0.3118	0.3700	0.5430
RACE	6	1	0.4252	0.4200	1.0249	0.3114
RACE	7	1	0.1457	0.2566	0.3227	0.5700
GENDER	2	1	0.5159	0.0949	29.5229	<.0001
AGE3	2	1	-0.0346	0.1015	0.1159	0.7336
AGE3	3	1	-0.0879	0.1474	0.3558	0.5508
INCPS_n	0	1	0.4875	0.2953	2.7248	0.0988
PELLMATCH_n	0	1	0.1894	0.2931	0.4175	0.5182
CINCOME4	2	1	-0.0604	0.1029	0.3450	0.5569
CINCOME4	3	1	-0.0652	0.1258	0.2682	0.6045
CINCOME4	4	1	-0.0843	0.2247	0.1408	0.7075
PAREDUC	0	1	0.00581	0.1699	0.0012	0.9727
PAREDUC	1	1	-0.00187	0.1290	0.0002	0.9884
PAREDUC	3	1	0.1721	0.1829	0.8856	0.3467
PAREDUC	4	1	0.3769	0.1725	4.7730	0.0289
PAREDUC	5	1	0.1259	0.1334	0.8913	0.3451
PAREDUC	6	1	-0.0343	0.1624	0.0446	0.8327
PAREDUC	7	1	-0.4234	0.2592	2.6671	0.1024
PAREDUC	8	1	-0.4108	0.4325	0.9021	0.3422
PAREDUC	9	1	-0.3658	0.5917	0.3822	0.5364
ATTNPTRN	2	1	-0.0225	0.1773	0.0160	0.8992
ATTNPTRN	3	1	0.0254	0.1302	0.0381	0.8452
HIGHLVEX	1	1	27.0084	896.8	0.0009	0.9760
HIGHLVEX	2	1	-0.2313	0.1217	3.6128	0.0573
HIGHLVEX	4	1	-0.1071	0.1160	0.8530	0.3557
HIGHLVEX	6	1	0.0266	0.1485	0.0321	0.8577
HIGHLVEX	7	1	0.2814	0.2588	1.1818	0.2770
HIGHLVEX	8	1	0.7320	0.2764	7.0145	0.0081
DEPNUMCH3	1	1	0.1106	0.1255	0.7770	0.3781
DEPNUMCH3	2	1	0.0227	0.1599	0.0201	0.8871
DEPNUMCH3	3	1	-0.2762	0.2150	1.6504	0.1989
pellamt3	0	1	0.0296	0.2884	0.0105	0.9183
pellamt3	2	1	0.1604	0.1056	2.3069	0.1288
totaid4	0	1	-0.4105	0.2966	1.9158	0.1663
totaid4	2	1	-0.0606	0.2927	0.0429	0.8359
totaid4	3	1	-0.2265	0.2948	0.5904	0.4422
totaid4	4	1	-0.1481	0.3037	0.2379	0.6257
tfedaid4	0	1	0.9515	0.3307	8.2795	0.0040
tfedaid4	2	1	0.6438	0.2999	4.6100	0.0318
tfedaid4	3	1	0.5898	0.3115	3.5842	0.0583
tfedaid4	4	1	0.6985	0.3156	4.8976	0.0269
instamt3	0	0	0	.	.	.
instamt3	2	1	0.2055	0.2434	0.7129	0.3985
instamt3	3	1	0.2135	0.4995	0.1827	0.6691
UGDEG	1	1	0.0972	0.1040	0.8732	0.3501
UGDEG	3	1	-0.0377	0.4899	0.0059	0.9387
UGDEG	4	1	-12.7628	431.7	0.0009	0.9764
JOBHOUR4	1	1	0.1293	0.1546	0.6994	0.4030
JOBHOUR4	3	1	-0.1920	0.1378	1.9407	0.1636
JOBHOUR4	4	1	0.0344	0.1314	0.0687	0.7932

Bias-likelihood model variables: institution group E (as of 6/27/17)

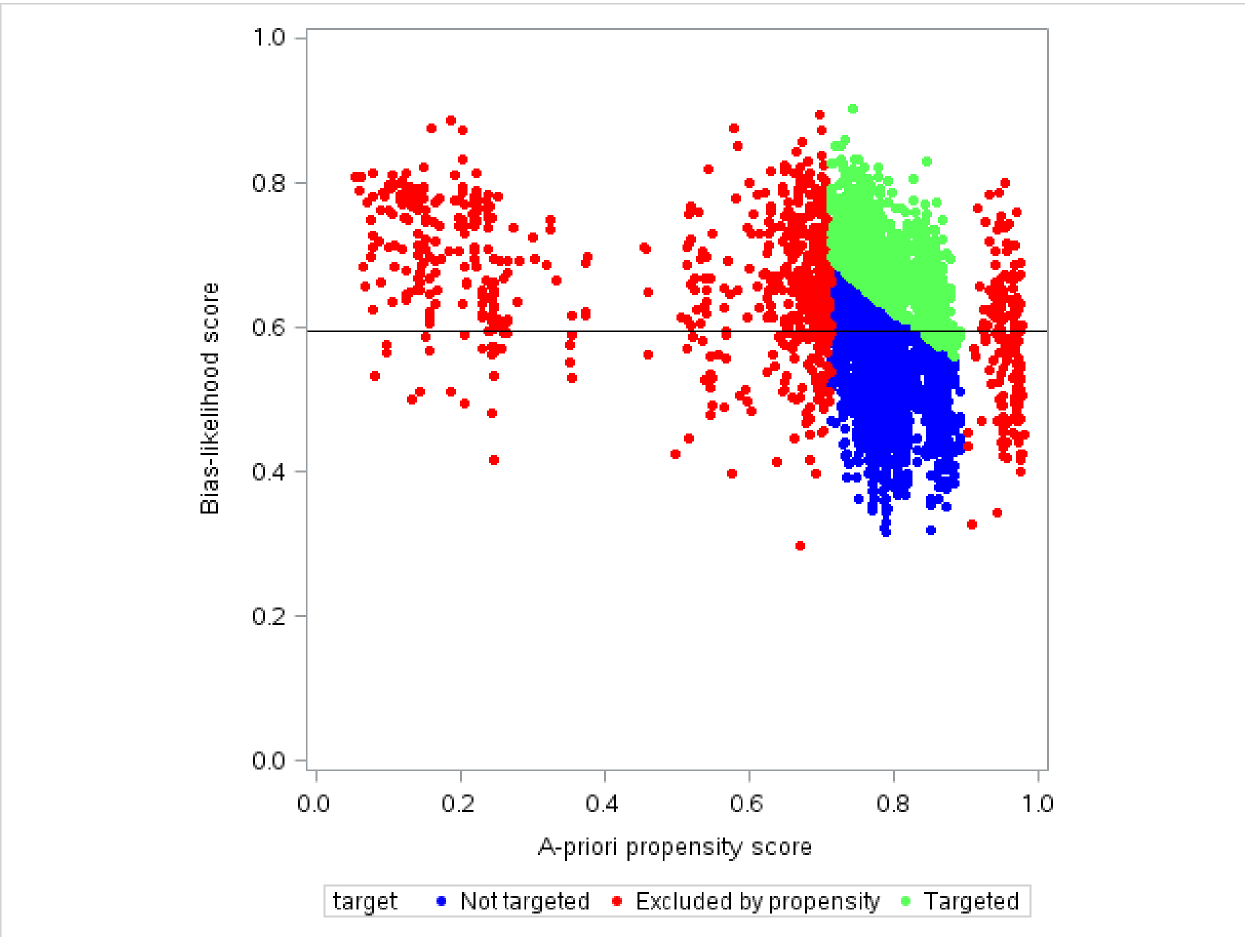
Analysis of Maximum Likelihood Estimates						
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.2124	0.1629	55.3651	<.0001
RACE	2	1	0.1782	0.0658	7.3430	0.0067
RACE	3	1	0.0973	0.0667	2.1259	0.1448
RACE	4	1	0.0800	0.1700	0.2216	0.6378
RACE	5	1	-0.1326	0.2409	0.3031	0.5820
RACE	6	1	-0.1828	0.2871	0.4053	0.5244
RACE	7	1	0.2014	0.1259	2.5578	0.1098
GENDER	2	1	0.5171	0.0543	90.8133	<.0001
AGE3	2	1	-0.00467	0.0663	0.0050	0.9438
AGE3	3	1	0.0323	0.0853	0.1432	0.7051
INCPS_n	0	1	-0.1100	0.1529	0.5180	0.4717
PELLMATCH_n	0	1	0.0224	0.1486	0.0227	0.8802
CINCOME4	2	1	0.2224	0.0627	12.5701	0.0004
CINCOME4	3	1	0.1706	0.0739	5.3358	0.0209
CINCOME4	4	1	0.2516	0.1208	4.3400	0.0372
PAREDUC	0	1	-0.0179	0.1109	0.0259	0.8721
PAREDUC	1	1	0.1033	0.0834	1.5337	0.2156
PAREDUC	3	1	0.2747	0.1158	5.6266	0.0177
PAREDUC	4	1	0.1976	0.0992	3.9654	0.0464
PAREDUC	5	1	0.1937	0.0767	6.3762	0.0116
PAREDUC	6	1	0.00922	0.0855	0.0116	0.9141
PAREDUC	7	1	-0.0368	0.1257	0.0858	0.7696
PAREDUC	8	1	0.2788	0.2256	1.5280	0.2164
PAREDUC	9	1	-0.0503	0.2814	0.0320	0.8580
ATTNPTRN	2	1	-0.1639	0.0797	4.2235	0.0399
ATTNPTRN	3	1	0.0376	0.0653	0.3320	0.5645
HIGHLVEX	1	1	0.5323	1.5769	0.1139	0.7357
HIGHLVEX	2	1	0.0699	0.1579	0.1959	0.6580
HIGHLVEX	4	1	-0.0465	0.0776	0.3600	0.5485
HIGHLVEX	6	1	-0.1184	0.0876	1.8275	0.1764
HIGHLVEX	7	1	-0.2827	0.1428	3.9174	0.0478
HIGHLVEX	8	1	-0.1235	0.1749	0.4985	0.4802
DEPNUMCH3	1	1	-0.2512	0.0836	9.0167	0.0027
DEPNUMCH3	2	1	-0.1245	0.0960	1.6802	0.1949
DEPNUMCH3	3	1	-0.2038	0.1081	3.5543	0.0594
pellamt3	0	1	0.1347	0.1481	0.8276	0.3630
pellamt3	2	1	0.2055	0.0653	9.9133	0.0016
totaid4	0	1	-0.5926	0.1630	13.2216	0.0003
totaid4	2	1	-0.1318	0.1503	0.7687	0.3806
totaid4	3	1	0.00193	0.1562	0.0002	0.9902
totaid4	4	1	0.0347	0.1620	0.0460	0.8302
tfedaid4	0	1	-11.5054	227.9	0.0025	0.9597
tfedaid4	2	1	0.2443	0.1544	2.5058	0.1134
tfedaid4	3	1	0.2914	0.1563	3.4769	0.0622
tfedaid4	4	1	0.2332	0.1625	2.0579	0.1514
instamt3	0	1	12.2012	227.9	0.0029	0.9573
instamt3	2	1	-0.0509	0.1705	0.0890	0.7654
instamt3	3	1	-0.2658	0.2545	1.0909	0.2963
UGDEG	1	1	-0.1437	0.1057	1.8471	0.1741
UGDEG	3	1	0.0435	0.0636	0.4687	0.4936
UGDEG	4	1	0.3694	0.6645	0.3089	0.5783
JOBHOUR4	1	1	0.2120	0.0909	5.4368	0.0197
JOBHOUR4	3	1	0.0618	0.0779	0.6303	0.4273
JOBHOUR4	4	1	0.0730	0.0674	1.1722	0.2790

<u>Attachment C. Targeting of current nonrespondents by institution group</u>

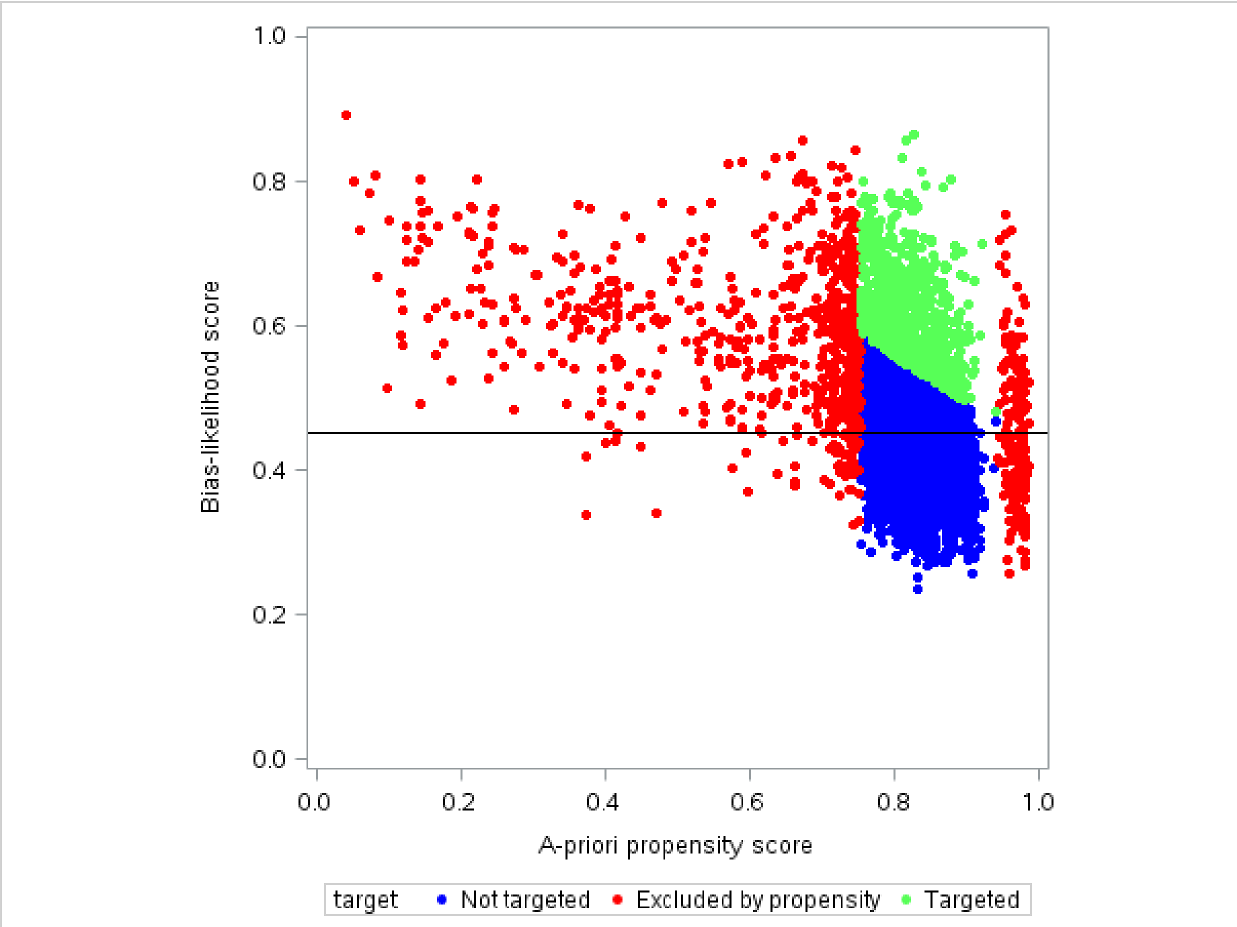
The following graphs plot the all current nonrespondents within each intituion group that are eligible for targeting (excludes controls groups, double nonrespondent, or deceased). In each scatterplot, the red dots show top and bottom response propensity scores that are removed from consideration for targeting as these cases are highly likely to respond without additional efforts or highly unlikely to respond even with additional efforts. The upper cut-off is based on visual examination of each scatterplot by institution group.

The green dots display the targeted cases. These cases have the highest importance scores (high values of *a-priori* propensity times bias-likelihood shown in the upper right corner) among the remaining eligible cases. The blue dots show the remaining eligible cases that will not be targeted. The horizontal line on each scatterplot represents the median bias-likelihood for that institution group.

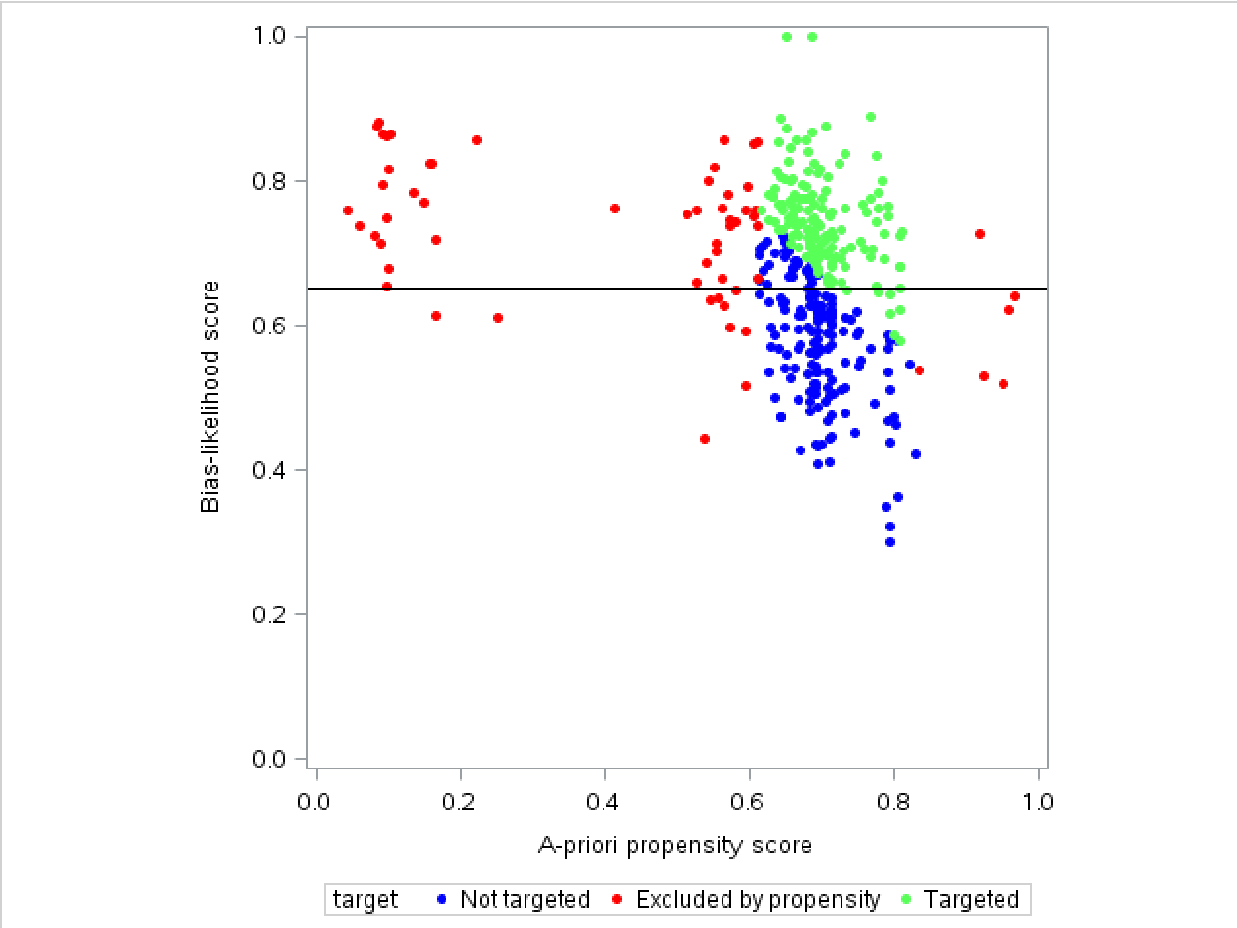
Targeting: Institution Group A



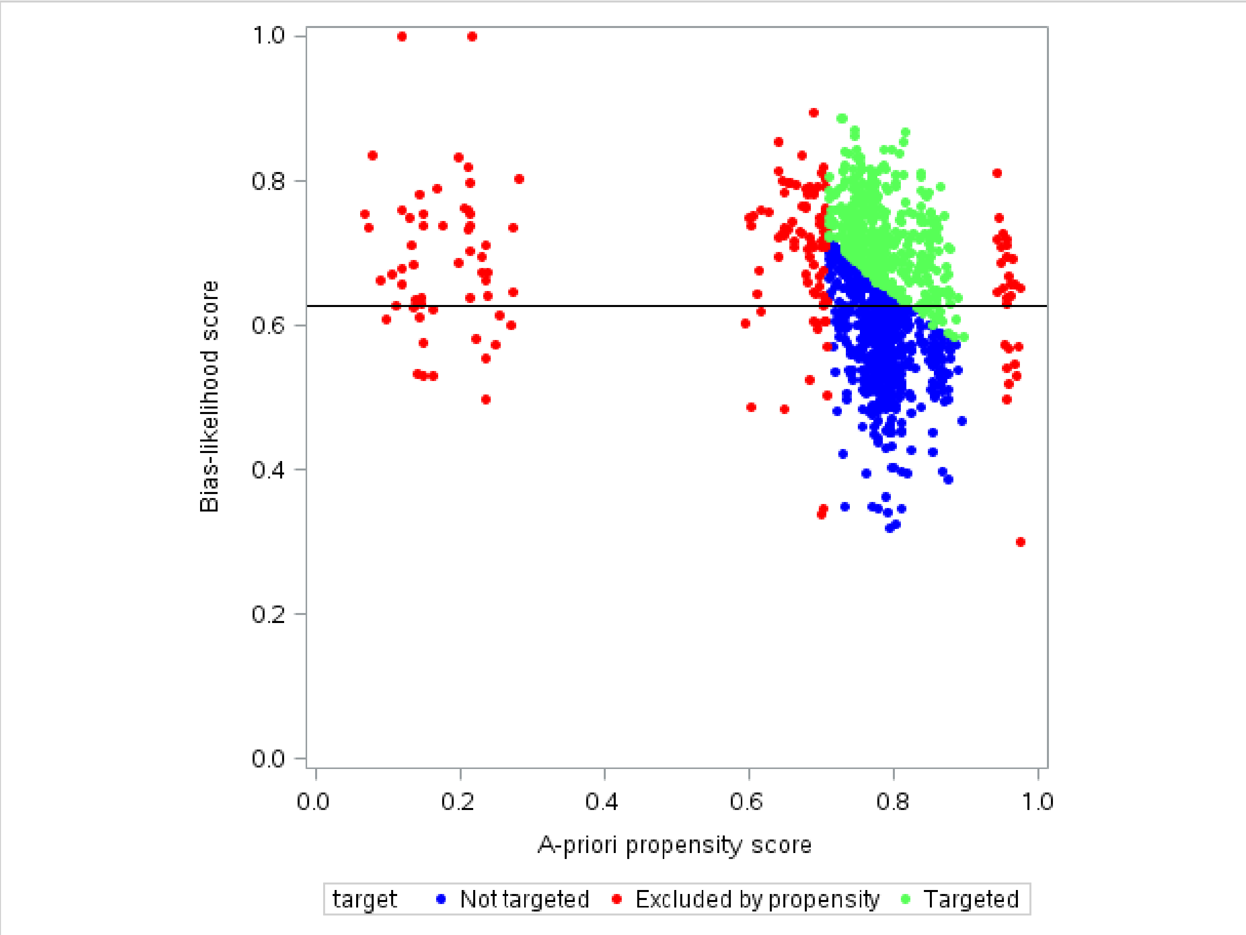
Targeting: Institution Group B



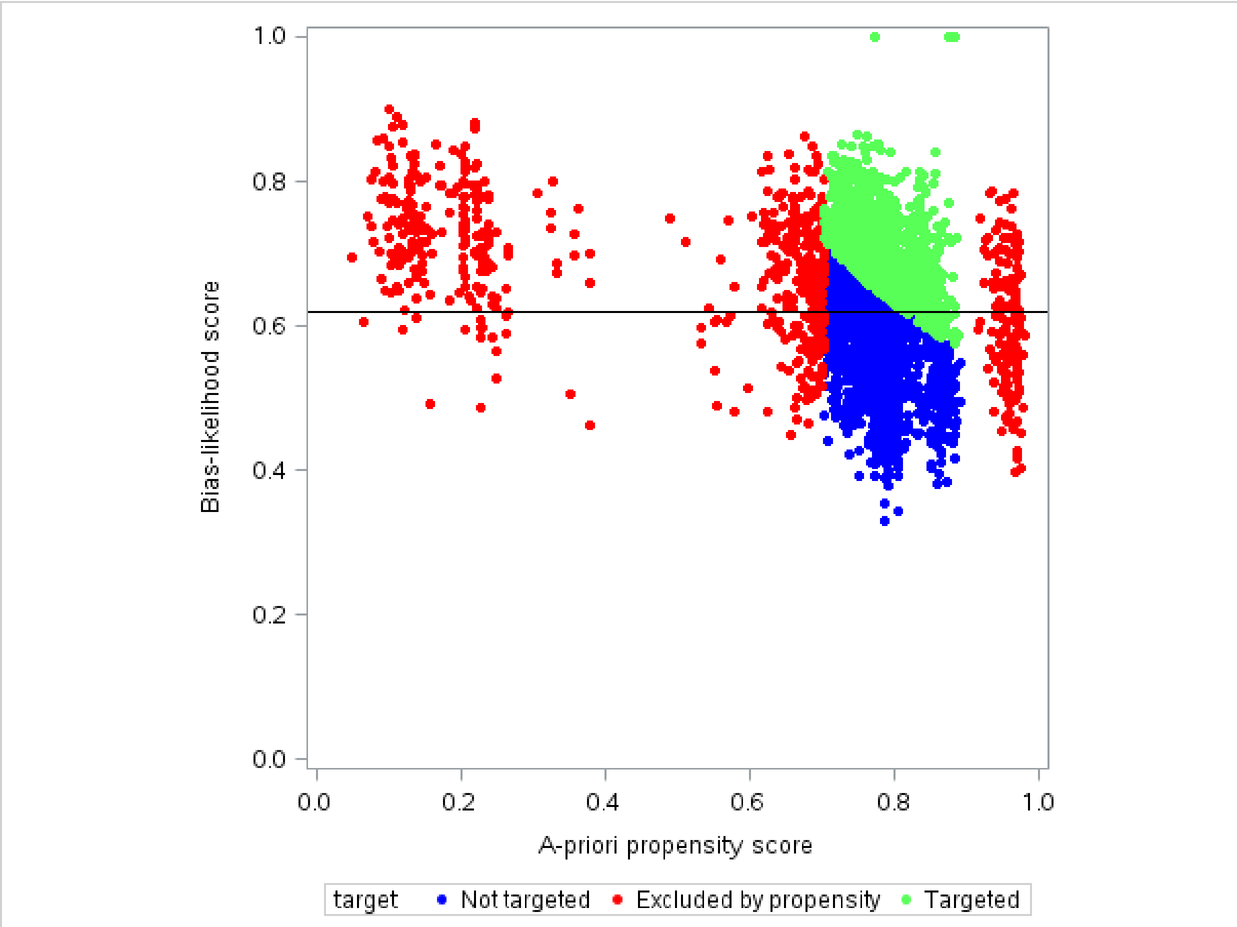
Targeting: Institution Group C



Targeting: Institution Group D



Targeting: Institution Group E



Targeting: Overall - All Institution Groups