Memorandum

DATE: April 5, 2017, revised April 7, 2017

TO: Robert Sivinski and E. Ann Carson, 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) Incentives Change Request (OMB# 1850-0631 v.13)

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 and February 2017 (OMB# 1850-0631 v.11-12). This request is to (1) use specific incentive strategies with the main BPS:12/17 sample, based on the recent data collection results from the calibration sample, and (2) to alter recruitment plans for sample member matched to a sanctions list maintained by the U.S. Treasury's Office of Foreign Assets Controls (OFAC). This request does not introduce changes to survey content, respondent burden, or the cost to the federal government.

Responsive design plan and calibration sample experiments

The approach to the BPS:12:17 responsive design plan employs an early calibration sample to identify treatments in the main study, and targets nonrespondents for specialized treatments to reduce nonresponse bias. The targeting utilizes an *importance* measure for selecting nonrespondents with moderate likelihood to respond and high potential impact on bias. For reference, Attachment A includes an excerpt from Part B of the full-scale study submission that describes the responsive design plan. Within Attachment A, the features and results of the BPS:12/14 responsive design plan can be found in Section 4.a, and the BPS:12/17 plan can be found in Section 4.b

While similar in approach, the BPS:12/17 design includes unique features 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 institutional 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, and an accelerated timeline for data collection interventions. Lastly, using the *importance* score to select nonrespondents that have potential to reduce nonresponse bias within institutional sector estimates, the BPS:12/17 plan includes two points in the main sample data collection at which nonresponding sample members will be targeted – the first using a $45 incentive boost, and the second with an offer of an abbreviated survey.

Experiment #1 for previous respondents. Previous respondents are NPSAS:12 study members who responded to either or both the NPSAS:12 or BPS:12/14 student interviews. As demonstrated in previous rounds, this group can be motivated to complete an interview but may require a series of prompts and communications, such as emails, letters, and CATI outbound calling. As NCES continues to seek new ways to encourage completion, experiment #1 is designed to test the efficacy of a non-contingent $10 prepaid incentive delivered through PayPal. Upon completion of the interview, the respondent receives an additional $20 (for a total of $30). The previous respondents in the calibration sample (n=2,972) were randomly assigned to the two treatment groups: 1) a $10 pre-paid offer with $20 promise incentive, and 2) no pre-paid offer ($30 promise incentive).

Table 1 displays the results of the first experiment that examines the impact of a $10 pre-paid offer versus no pre-paid amount for previous respondents. A Pearson chi-squared test was used to measure statistical significance with the null hypothesis being that there was no difference in response rate between the two experiment conditions. This test was a two-sided test as we had no prior belief that one of the conditions would outperform the other. Using the numbers below, the *p*-value for the statistic was less than 0.0001 which allowed us to reject the null hypothesis and conclude that the response rates between the two groups significantly differ. The response rate for the experimental condition with no pre-paid offer was higher (34.9 percent) than the response rate of those offered a $10 pre-paid PayPal offer (28.0 percent).

Table 1. Results of pre-paid $10 experiment (#1)

| Calibration | Respondent | | Total | Response Rate |
	No	Yes		
No pre-paid	**968**	**518**	1,486	34.9%
Pre-paid	**1,070**	**416**	1,486	28.0%
Total	2,038	934	2,972	31.4%

Experiment #2 for double nonrespondents. Double nonrespondents are NPSAS:12 study members who responded to neither the NPSAS:12 nor the BPS:12/14 student interview. In this experiment, NCES attempted to determine if an additional incentive or a less burdensome interview would motivate a response. Experiment #2 was designed to test the impact of offering a $75 incentive for completion of the full interview response versus a $30 incentive for an abbreviated interview[1]. The double nonrespondents in the calibration sample (n=870) were randomly assigned to the two treatment groups.

Table 2 displays the results of the second experiment that examines the special protocol for double nonrespondents: a full interview with a $75 total incentive versus an abbreviated interview and $30 total incentive. Using a two-sided Pearson chi-squared test to measure statistical significance, the null hypothesis was no difference in response rate between the two experiment conditions. The *p*-value for the statistic was 0.013 which allowed us to reject the null hypothesis and conclude that the response rates

---

[1] All double nonrespondents were offered the prepaid $10 through PayPal.

between the two groups significantly differ. The response rate for the full interview with $75 total incentive was higher, (4.1 percent) than response rate (1.4 percent) of those offered the abbreviated interview with $30 total incentive.

Table 2. Results of larger incentive vs. shorter survey experiment (#2)

Calibration	Respondent		Total	Response Rate
	No	Yes		
Full/$75	**417**	**18**	435	4.1%
Abbrev./$30	**429**	**6**	435	1.4%
Total	846	24	870	2.8%

However, given the low overall response rate for this group after 4 weeks of data collection (2.8 percent), we conducted additional analyses to assess the impact of double nonrespondents on estimates. While only 10 percent of the entire sample (calibration and main sample), 58 percent of double nonrespondents were enrolled in private for-profit institutions in NPSAS:12. As a result of this clustering in the for-profit sectors, we attempted to assess potential impact of this group on nonresponse bias and on the precision of estimates.

For nonresponse bias, we know little about how double nonrespondents differ on key measures from previous respondents, but were able to estimate some differences using transcript data in the BPS:04 cohort. Though not perfectly analogous to the BPS:12 double nonrespondents, we used the PETS:09 data to create a proxy and to infer postsecondary attendance and coursetaking patterns. From this analysis, we concluded that double nonrespondents have higher rates of enrollment at 2-year and less-than-2-year institutions, enroll at for-profit institutions at higher rates, and are less likely to attain a degree or certificate. In summary, we conclude that double nonrespondents, modeled on the BPS:04 cohort, differ by several characteristics of prime interest to BPS:12/17, thereby potentially contributing to nonresponse bias.

We also assessed the increase in precision if we field double nonrespondents. We estimated the standard error of the sample proportion of respondents, attending private for-profit institutions, receiving Pell grants under two scenarios: a) pursuing double nonrespondents, and b) excluding double nonrespondents. We then calculated the estimated standard errors under these scenarios by taking the ratio of the standard error under scenario #1 to the standard error under scenario #2. To estimate standard errors, we had to make assumptions about the rate at which double nonrespondents and previous respondents would participate by the end of data collection. To convey the range of possibilities, we considered six different participation rates for double nonrespondents and five different participation rates for previous respondents - for a total of 30 different combinations. The percent reduction in standard error for each combination of rates is provided in the following table 3.

Table 3. Percent Reduction in Standard Error Arising from Pursuit of Double Nonrespondents, by Assumed Participation Rates, at For-Profit Institutions

Assumed participation rate of double nonrespondents	Assumed participation rate of previous respondents				
	55%	**60%**	**65%**	**70%**	**75%**
4%	0.5%	0.5%	0.4%	0.4%	0.4%
10%	1.2%	1.1%	1.1%	1.0%	0.9%
15%	1.8%	1.7%	1.6%	1.5%	1.4%
20%	2.4%	2.2%	2.1%	1.9%	1.8%
25%	3.0%	2.8%	2.6%	2.4%	2.2%
30%	3.5%	3.3%	3.0%	2.8%	2.7%

At the time of this assessment, the observed participation rate among double nonrespondents was 4.1 percent so we conservatively estimated a reduction in standard error of approximately 0.5 percent. The degree to which the standard error is reduced varies by participation rates. The precision of the sample proportion increases as the participation rate of double norespondents increases and decreases as participation rate of previous respondents increases. The reduction in precision as the previous respondent participation rate increases is due to the large disparity in the numbers of double nonrespondents (n=2,406) and prior respondents (n=27,507) in the main sample.

Recommendations for main sample

Based on the findings of the calibration experiments, we recommend dropping the offer of a prepaid incentive to either group (previous respondents and double nonrespondents)[2].

For double nonrespondents, we recommend fielding this subgroup in the main sample data collection with a $75 incentive offer upon completion of the full interview. After data collection and data processing, NCES will analyze the *actual* impact of double nonrespondents on nonresponse bias in order to inform future BPS studies, as well as the HSLS and B&B. If results from BPS:12/17 show that including them has no impact on the estimates, or if we are unable to obtain enough responses, then we can conclude future studies should not expend the effort or resources for participation of double nonrespondents. However, if results demonstrate that the inclusion of double nonrespondents has an impact on estimates, NCES can either include them, or warn analysts of the issue in the data documentation.

Data collection for the main sample is scheduled to begin on May 3, 2017. The initial contact letter, scheduled to be mailed on April 19, 2017, will need to incorporate the incentive amounts we will offer the main sample. Modeling for targeted treatment of the $45 incentive boost will be conducted around June 14, 2017. Modeling for the targeted offer of the abbreviate interview is scheduled for August 9, 2017. Before NCES uses these models to target sample members in the main sample, NCES will provide OMB with results of the modeling for OMB's review and approval. Data collection is scheduled to end on October 10, 2017.

Office of Foreign Assets Controls' (OFAC) Specially Designated Nationals List

Prior to the start of data collection for the main sample, BPS sample members will be matched to a federal database maintained by the U.S. Department of the Treasury's Office of Foreign Assets Controls (OFAC). OFAC administers and enforces economic and trade sanctions based on U.S. foreign policy and national

---

[2] The original BPS:12/17 responsive design plan, as written in Part B, gave all double nonrespondents the prepaid incentive. However, given the negative impact on response rates for previous respondents, NCES recommends not offering the prepaid offer to any group.

security goals. As part of its enforcement efforts, OFAC publishes a list of individuals and companies called the "Specially Designated Nationals List" or "SDN." Their assets are blocked and U.S. entities are prohibited from conducting trade or financial transactions with those on the list (https://www.treasury.gov/resource-center/sanctions/Pages/default.aspx).

In order to determine if there are any BPS:12/17 sample members to whom we cannot offer an incentive, we are matching the sample to SDN by name using the Jaro-Winkler and Soundex algorithms recommended by OFAC. To avoid over-matching, we will review the cases based on full name, date of birth, and address. The challenge is to ensure that each final match is the person on the SDN list, and not someone else, and that the matched cases are not unfairly identified as ineligible for an incentive. We are still determining the procedures to be used with the final matched sample members and will submit to OMB a change memo in the near future to clarify our approach. For now the final matches will not receive a survey request until an approach is determined.

Attachment A. Excerpt of Supporting Statement Part B

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[3]

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

---

[3] 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."

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 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:[4]

- 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);
- 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.

---

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

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).

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:

$$Z$$

$$R \quad Y$$
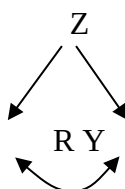
Nonresponse bias arises when there is a relationship between R and Y. Just as in adjustment for nonresponse bias (see Little and Vartivarian, 2005), a Z-variable cannot be effective in nonresponse bias reduction if corr(Z,Y) is weak or nonexistent, even if corr(Z,R) is substantial. That is, selection 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[5] (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

---

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

they 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	172 10,030
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	1,838 3,545 397 2,219 2,717
C	8: Private for profit less-than-2-year	1,450
D	9: Private for profit 2-year	2,989
E	10: Private for profit 4-year	8,401

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[6].

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[7] payment, to selected sample members. Prior to the start of full-scale data collection, 2,972 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.

---

[6] 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.

[7] 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).

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 870 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[8] 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
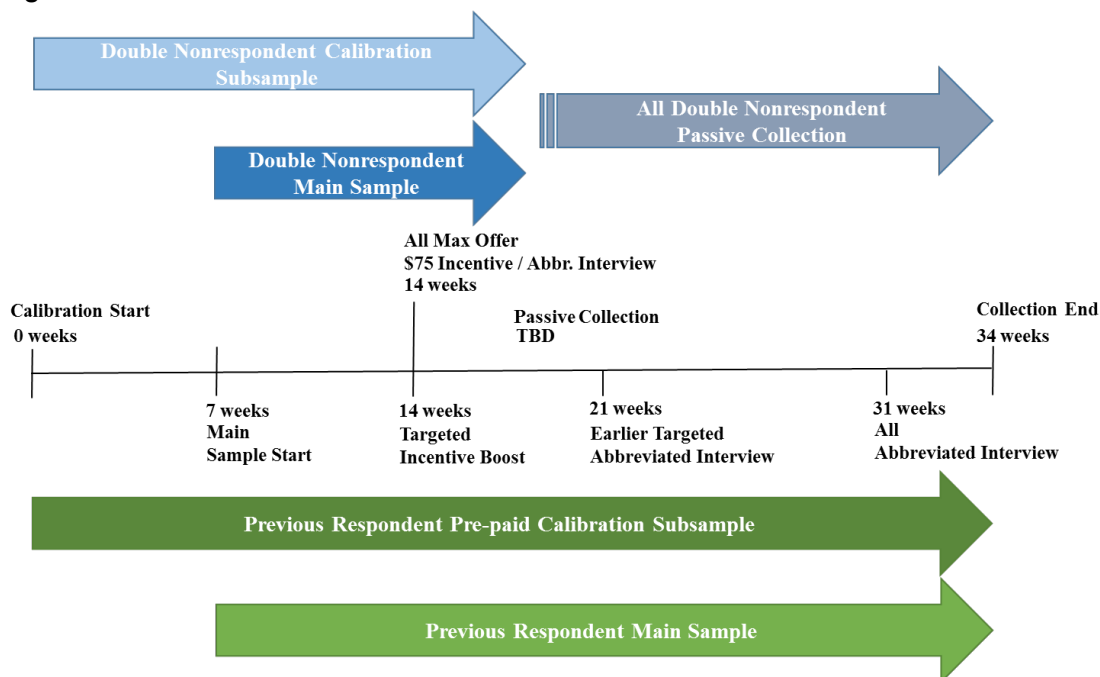
Evaluation of the BPS:12/17 Responsive Design Effort. The analysis plan is based upon two premises: (1) offering special interventions to some, targeted, sample members will increase participation in the

---

[8] 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.

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 3,615 individuals (723 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)

7

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:

- Research question 1: Did the *a priori* response propensity model predict overall unweighted BPS:12/17 response?

    o H_0: At the end of data collection, there will be no association between *a priori* propensity predictions and observed response rates.

- Research question 2: Does one special protocol increase response rates for double nonrespondents versus the other protocol?

    o 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.

- Research question 3: Does a $10 pre-paid PayPal offer increase early response rates?

    o 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.

- Research question 4: 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, 723 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	10,202	723	9,479	0.916	0.966	0.925
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	10,716	723	9,993	0.917	0.967	0.926
C	8: Private for-profit less-than-2-year	1,450	723	727	0.715	0.816	0.723
D	9: Private for-profit 2-year	2,989	723	2,266	0.861	0.933	0.873
E	10: Private for-profit 4-year	8,401	723	7,678	0.912	0.964	0.921

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=2,972) will be randomly split into two equally sized groups of 1,486 sample members. 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[9]. 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=870) into two equally sized groups (n=435) where one group will receive a $30 baseline offer with an abbreviated interview, while the other 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

---

[9] Calculated using SAS Proc Power. https://support.sas.com/documentation/cdl/en/statugpower/61819/PDF/default/statugpower.pdf

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%