NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report   **July 2002**

Beginning Postsecondary Students Longitudinal Study: 1996-2001 (BPS:1996/2001) Methodology Report

Chapter 6
Weighting and Variance Estimation

Development of statistical analysis weights for the BPS:1996/2001 sample is discussed in section A below. Cross-sectional weights were constructed for analyzing the respondents to BPS:1996/2001. In addition, two longitudinal weights were constructed, one for analyzing the students who participated in all three interviews—NPSAS:96, BPS:96/98, and BPS:1996/2001—and another for analyzing the students who responded to NPSAS:96 and BPS:1996/2001. Analysis procedures that can be used to produce design-unbiased estimates of sampling variances are discussed in section B, including variances computed using Taylor series and balanced repeated replications (BRR) techniques. Section C discusses the accuracy of BPS:1996/2001 estimates in terms of both precision and potential for bias. This section includes survey design effect tables that illustrate the level of precision achieved by the BPS:1996/2001 survey for key analytic outcomes for several important analysis domains. Finally, section D gives weighted response rates.

A. Analysis Weights

The initial file for the BPS:1996/2001 sample contained approximately

- 10,300 BPS:96/98 respondents and
- 1,800 BPS:96/98 nonrespondents.

As noted in chapter 2, the final BPS:1996/2001 sample consisted of

- the eligible respondents to BPS:96/98 and
- a subsample of nonrespondents to BPS:96/98 who were NPSAS:96 respondents.

Among these, over 20 were identified as deceased either prior to data collection and after data collection began.

A statistical analysis weight was computed to be used for analyzing data from the BPS:1996/2001 respondents. In addition, two longitudinal weights were computed: a weight for analyzing those BPS:1996/2001 respondents who also responded to NPSAS:96 and BPS:96/98, and a weight for analyzing the BPS:1996/2001 respondents who only responded to NPSAS:96 and BPS:1996/2001.

The weights for the BPS:96/98 respondents were constructed by applying a series of adjustments for subsampling and nonresponse to the base weights for the 2001 follow-up of the BPS:96 cohort, namely B01IAWT[1]. Specifically, four adjustments were made:

---

[1] The rationale for the variable name "B01IAWT" is the following:

- to account for subsampling of the BPS:96/98 nonrespondents;
- to account for those not located;
- to account for refusals, among those who were located; and
- to account for types of nonresponse other than refusals among those who were located and did not refuse.

Construction of the longitudinal weight for those who responded to all three surveys consisted of an additional adjustment for nonresponse to either NPSAS:96 or BPS:96/98. Construction of the analysis weight for those who responded to both NPSAS:96 and BPS:1996/2001, but not to BPS:96/98, consisted of an additional adjustment for nonresponse to NPSAS:96.

1.   Base Weight for BPS:1996/2001—Adjustment for Subsampling of BPS:96/98 Nonrespondents

As discussed in chapter 2, a subsample of BPS:96/98 nonrespondents was included in BPS:1996/2001. The subsample, rather than all nonrespondents, was fielded in order to reduce data collection costs. The weight B01IAWT was adjusted for those students, j, in the subsample by multiplying by the inverse of their selection probabilities. These probabilities take into account the stratification and probability proportional to size (PPS) sampling that was used in selecting the subsample. The adjustment was

$$ADJ1_j = 1/\pi_{j.}$$

The weight was calculated as:

B01_100U   = B01IAWT * ADJ1, for students in the BPS:96/98 nonrespondent subsample
            = B01IAWT for all other students.

The weights B01_100U for the students in the subsample were then adjusted so that they summed to the weight sum of B01IAWT for the BPS:96/98 nonrespondents. This adjustment resulted in the initial sampling weight for the BPS:1996/2001 sample, which is denoted B01_100. B01_100 was further adjusted to produce the BPS:1996/2001 analysis weights, as described below.

The weight B01_100 is nonnegative for both the eligible and ineligible (i.e., deceased) students. Weighted response rate tables later in this chapter were computed using B01_100 and were based on the set of eligible students. The eligible students are those with B01ELIG=1 where B01ELIG is the eligibility indicator for BPS:1996/2001.

B   denotes the BPS survey
01   denotes the year 2001
I   stands for "initial"
A   stands for "analysis"
WT   stands for "weight"

2.   BPS:1996/2001 Cross-Sectional Weights

Analysis weights were constructed for the respondents to BPS:1996/2001. The weights were constructed by applying adjustments to the base weight B01_100. This section describes each of the adjustment steps, the variables considered for the adjustments, and the variables in the final weight adjustment models.

The adjustment for nonresponse was performed in three steps because the predictors of response propensity were potentially different for each of the following outcomes:

- inability to locate the student,
- refusal to be interviewed, and
- other noninterview.

Using these three steps of nonresponse adjustment achieved greater reduction in nonresponse bias to the extent that different variables were significant predictors of nonresponse propensity at each step.

All nonresponse adjustments were fit using RTI's proprietary generalized exponential modeling procedure (GEM[2]), which is similar to logistic modeling using bounds for adjustment factors. A key feature and advantage of the GEM software is that the nonresponse adjustment and weight trimming and smoothing are all accomplished in one step. Lower and upper bounds are set on the weight adjustment factors. The bounds can be varied, depending on whether the weight falls inside or outside a range, such as one defined by the bounds (median – 3 times the interquartile range, median + 3 times the interquartile range). This allows different bounds to be set for adjustments for weights that are considered high extreme, low extreme, or nonextreme. In this way, the extreme weights can be controlled and the design effect due to unequal weighting reduced.

Candidate predictor variables selected were those thought to be predictive of nonresponse and nonmissing for most of the sample (nonrespondents as well as respondents). The candidate predictor variables included

- age (categorical);
- typical age for a beginning student (yes or no);
- race/ethnicity;
- gender;
- citizenship status in the base year;
- attendance status in the base year;
- level of institution attended in the base year;
- control of institution attended in the base year;
- region of institution attended in the base year;
- size of institution attended in the base year (categorical);
- applied for financial aid in the base year (yes or no);

---

[2] Folsom, R.E. and. Singh, A.C. (2000). "The Generalized Exponential for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification." *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp. 598-603.

- receipt of federal aid in the base year (yes or no);
- receipt of Pell Grant in the base year (yes or no);
- receipt of Stafford Loan in the base year (yes or no);
- receipt of state aid in the base year (yes or no);
- receipt of institutional aid in the base year (yes or no);
- receipt of any aid in the base year (yes or no);
- previous response status (whether the student was a respondent to both NPSAS:96 and BPS:96/98 versus a nonrespondent to either NPSAS:96 or BPS:96/98);
- income of independent students and parents of dependent students (collapsed);
- parents' highest educational attainment;
- degree completion status in 1998;
- number of telephone numbers available;
- number of times an answering machine was encountered (for located students); and
- whether the student was in a field cluster.

To detect important interactions for the logistic models, a Chi-squared automatic interaction detection analysis (CHAID) was performed on the predictor variables. The CHAID analysis divided the data into segments that differed with respect to the response variable (located, did not refuse, or respondent, depending on the model). The segmentation process first divided the sample into groups based on categories of the most significant predictor of response. It then split each of these groups into smaller subgroups based on other predictor variables. It also merged categories of a variable that were found to be nonsignificant. This splitting and merging process continued until no more statistically significant predictors were found (or until some other stopping rule was met). The interactions from the final CHAID segments were then defined.

The nonresponse bias for these same variables was estimated, and then a statistical test of whether or not the bias was significant was performed. Tests were performed to identify significant differences between refusal conversions and other respondents; significant differences suggest a potential for nonresponse bias because of the refusal population being different from the other respondents. Additional tests were performed to detect significant differences between late respondents and other respondents; significant differences would suggest a potential for nonresponse bias because of the noncontacts/late-contact population being different from the other respondents. Results and further details of these analyses are given below in section C.

The interaction segments and all the main effects were then subjected to variable screening in the GEM logistic procedure. Variables with significant bias were included in each nonresponse model. The models initially included all of the potentially important variables. The interaction segments identified by CHAID were also retained in all of the models. The most nonsignificant variables were deleted sequentially until the deletion of additional variables did not appreciably improve the unequal weighting effect (UWE). Different bounds on the weight adjustments, depending on whether the weight is classified as high extreme, nonextreme, or low extreme, were applied within the NPSAS:96 institutional sampling strata to accomplish nonresponse adjustment, truncation, and smoothing in one step. A large number of predictor variables in each nonresponse model were kept. This allows the estimates to be calibrated based

on the respondents to as many totals as possible that are known for both respondents and nonrespondents.

a.   Weight Adjustment for Nonrespondents Who Were Not Located

Of the individuals eligible for the BPS:1996/2001 sample, 92 percent was contacted. An adjustment was performed to the weight B01_100 to adjust for the remaining sample members who did not respond because they were not located. As described above, a CHAID analysis was performed on all of the predictor variables to detect important interactions. All potentially important variables were included in the model. Highly nonsignificant variables were deleted from the model until there was little change in the unequal weighting effect.

Table 6.1 presents the final predictor variables used in the logistic model that adjusted the weights for those who were not located, and gives the weighted location rate and the average weight adjustment factors resulting from these variables. The weighting adjustment factor for student j was the reciprocal of the predicted response probability, or

$$ADJ2_j = 1/ p_{L,j}.$$

The weight, adjusted for those who were not located, was then computed as

$$LOCWT = B01\_100 * AJD2 \text{ for those who were located}$$
$$= 0 \text{ otherwise.}$$

b.   Weight Adjustment for Nonrespondents Who Refused

Of the sample members who were eligible and located for the BPS:1996/2001 sample, 3 percent refused. An additional adjustment was performed to the weight that had been adjusted for the not located, LOCWT, to adjust for those who refused. As in the case of the adjustment for the not located, a CHAID analysis was performed on all of the predictor variables to detect important interactions. All potentially important variables were included in the model. Highly insignificant variables were deleted from the model until there was little change in the unequal weighting effect.

Table 6.2 presents the final predictor variables used in the logistic model that adjusted the weights for those who refused and gives the weighted nonrefusal rate (for those who were located) and the average weight adjustment factors resulting from these variables. The weighting adjustment factor for student j was the reciprocal of the predicted response probability, or

$$ADJ3_j = 1/p_{NRef,j.}$$

The weight adjusted for those who refused was computed

$$NREFWT = LOCWT * ADJ3 \text{ for those who did not refuse}$$
$$= 0 \text{ otherwise.}$$

Table 6.1.—Average weight adjustment factors from the logistic model used to adjust for student location nonresponse

Predictor variables	Weighted response rate	Average weight adjustment
Total	89.6	1.45
Age		
19 or younger	91.7	1.40
20 to 23	80.6	1.77
24 to 29	92.5	1.69
30 to 39	78.7	1.49
40 or older	94.1	1.74
Gender		
Male	87.3	1.50
Female	91.5	1.42
Citizenship		
U.S. citizen	89.8	1.45
Non-citizen	86.2	1.55
Attendance status		
Full-time/full year 1 institution	92.7	1.42
Full-time/full year more than 1 institution	96.1	1.05
Full-time/part year	79.8	1.59
Part-time/full year 1 institution	88.6	1.57
Part-time/full year more than 1 institution	60.4	1.55
Part-time/part year	91.8	1.61
Institution level		
4-year	91.7	1.44
2-year	89.0	1.45
Less-than-2-year	83.3	1.57
Institution control		
Public	90.2	1.41
Private not-for-profit	91.0	1.49
Private for-profit	83.5	1.55
Institution region		
New England	90.1	1.65
Mid East	91.4	1.30
Great Lakes	91.7	1.37
Plains	93.5	1.57
Southeast	86.9	1.46
Southwest	88.2	1.56
Rocky Mountains	84.4	1.67
Far West	89.8	1.45
Outlying Area	92.3	1.19
Type of institution and enrollment category		
Public		
Fewer than 1,000	95.5	1.09
1,000–2,499	81.1	1.79
2,500–4,999	91.6	1.42
5,000–9,999	81.6	1.54
10,000–19,999	92.6	1.43
20,000 or more	94.4	1.32
Private not-for-profit		
Fewer than 1,000	84.5	1.32
1,000–2,499	93.7	1.54
2,500–4,999	93.6	1.42
5,000–9,999	84.6	1.45
10,000 or more	92.3	1.58
Private for-profit		
Fewer than 300	82.0	1.63
300–999	86.6	1.49
1,000 or more	79.7	1.59

See footnotes at end of table.

Table 6.1.—Average weight adjustment factors from the logistic model used to adjust for student location nonresponse—Continued

Predictor variables	Weighted response rate	Average weight adjustment
Applied for aid		
Yes	88.6	1.42
No	91.9	1.62
Receipt of federal aid		
Yes	88.8	1.43
No	90.3	1.49
Receipt of Pell Grant		
Yes	86.9	1.47
No	90.7	1.44
Receipt of Stafford Loan		
Yes	88.8	1.45
No	89.9	1.46
Receipt of state aid		
Yes	93.4	1.32
No	88.9	1.49
Receipt of institution aid		
Yes	92.4	1.42
No	88.8	1.47
Receipt of any aid		
Yes	89.2	1.42
No	90.1	1.54
Parents' income (for dependent students)		
$29,999 or less	88.7	1.45
$30,000–$59,999	94.1	1.35
$60,000–$99,999	92.8	1.41
$100,000 or more	85.8	1.56
Student's income (for independent students)		
$5,000 or less	85.2	1.55
$5,000–$9,999	82.6	1.66
$10,000–$19,999	75.0	1.80
$20,000 or more	94.6	1.53
Parents' highest education		
Less than high school	86.1	1.44
High school diploma	88.1	1.41
Some college	90.9	1.47
Degree status in 1998		
Certificate or associate degree received	90.5	1.44
Vocational or associate degree program, degree not received	87.2	1.49
Bachelor's degree or program	92.3	1.44
Telephone numbers available		
0 or 1 number	88.2	1.71
2 numbers	88.2	1.59
3 numbers	91.4	1.37
4 numbers	90.3	1.48
5 numbers	93.7	1.39
6 numbers	83.8	1.46
7 or more numbers	83.2	1.40
Number of times answering machine was encountered		
None	91.4	1.38
Once	92.8	1.36
More than once	88.1	1.50

See footnotes at end of table.

Table 6.1.—Average weight adjustment factors from the logistic model used to adjust for student location nonresponse—Continued

Predictor variables	Weighted response rate	Average weight adjustment
Interaction segments		
1 = Not in field cluster, White, non-Hispanic, certificate, associate's, or bachelor's received, or in bachelor's program	91.2	1.41
2 = Not in field cluster, White, non-Hispanic, vocational or associate's degree program	82.0	1.66
3 = Not in field cluster, other than White non-Hispanic, did not encounter answering machine	70.5	1.71
4 = Not in field cluster, other than White non-Hispanic, encountered answering machine 1 or more times	84.8	1.72
5 = In field cluster, White, non-Hispanic, attended full-time/full year	96.0	1.42
6 = In field cluster, White, non-Hispanic, attended other than full-time/full year	93.9	1.46
7 = In field cluster, Black, non-Hispanic, male	84.2	1.46
8 = In field cluster, Black, non-Hispanic, female	93.9	1.42
9 = In field cluster, Hispanic, 0 or 1 telephone numbers available	78.2	1.40
10 = In field cluster, Hispanic, 2–4 telephone numbers available	94.4	1.18
11 = In field cluster, Hispanic, 5 or more telephone numbers available	90.1	1.52
12 = In field cluster, other race, not typical age for level	93.9	1.76
13 = In field cluster, other race, typical age for level	95.1	1.32

NOTE: Predictor variables are from base year data (NPSAS:96) with the exception of degree status in 1998, telephone numbers available, number of times answering machine was encountered, and certain interaction variables.

The weight used is B01_100.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.2.—Average weight adjustment factors from the logistic model used to adjust for student refusal nonresponse

Predictor variables	Weighted response rate	Average weight adjustment
Total	96.1	1.04
Age		
19 or younger	96.6	1.03
20 to 23	95.2	1.05
24 to 29	94.5	1.06
30 to 39	94.9	1.06
40 or older	97.0	1.03
Race/ethnicity		
White, non-Hispanic	95.8	1.04
Black, non-Hispanic	97.4	1.03
Hispanic	96.9	1.03
Asian/Pacific Islander	96.5	1.03
American Indian/Alaska Native	97.8	1.02
Other	93.5	1.08
Gender		
Male	96.8	1.03
Female	95.6	1.04
Citizenship		
U.S. citizen	96.1	1.04
Non-citizen, eligible	96.7	1.03
Non-citizen, ineligible	98.4	1.00
Attendance status		
Full-time/full year 1 institution	96.2	1.03
Full-time/full year more than 1 institution	97.0	1.03
Full-time/part year	95.8	1.05
Part-time/full year 1 institution	97.0	1.03
Part-time/full year more than 1 institution	96.7	1.01
Part-time/part year	95.3	1.05
Institution level		
4-year	97.0	1.03
2-year	95.8	1.05
Less-than-2-year	94.2	1.06
Institution control		
Public	96.4	1.03
Private not-for-profit	96.9	1.03
Private for-profit	93.6	1.06
Institution region		
New England	93.8	1.06
Mid East	94.4	1.05
Great Lakes	96.1	1.03
Plains	95.1	1.05
Southeast	97.2	1.03
Southwest	96.2	1.03
Rocky Mountains	98.6	1.02
Far West	96.5	1.03
Outlying Area	97.9	1.02
Type of institution and enrollment category		
Public		
Fewer than 1,000	95.6	1.06
1,000–2,499	96.2	1.05
2,500–4,999	99.5	1.00
5,000–9,999	95.5	1.05
10,000–19,999	96.0	1.03
20,000 or more	96.3	1.03

See footnotes at end of table.

Table 6.2.—Average weight adjustment factors from the logistic model used to adjust for student refusal nonresponse—Continued

Predictor variables	Weighted response rate	Average weight adjustment
Private not-for-profit		
Fewer than 1,000	96.0	1.05
1,000–2,499	97.3	1.03
2,500–4,999	95.9	1.04
5,000–9,999	97.7	1.02
10,000 or more	97.1	1.03
Private for-profit		
Fewer than 300	94.6	1.06
300–999	93.4	1.07
1,000 or more	92.3	1.07
Applied for aid		
Yes	96.2	1.04
No	96.1	1.03
Receipt of federal aid		
Yes	95.9	1.04
No	96.3	1.03
Receipt of Pell Grant		
Yes	96.0	1.04
No	96.2	1.03
Receipt of Stafford Loan		
Yes	95.6	1.04
No	96.4	1.03
Receipt of state aid		
Yes	96.8	1.03
No	96.0	1.04
Receipt of institution aid		
Yes	97.2	1.03
No	95.9	1.04
Receipt of any aid		
Yes	96.3	1.04
No	95.9	1.04
Parents' income (for dependent students)		
$29,999 or less	95.8	1.04
$30,000–$59,999	96.6	1.03
$60,000–$99,999	96.0	1.04
$100,000 or more	97.6	1.02
Student's income (for independent students)		
$5,000 or less	96.6	1.04
$5,000–$9,999	93.3	1.07
$10,000–$19,999	96.1	1.05
$20,000 or more	95.9	1.04
Parents' highest education		
Less than high school	96.9	1.03
High school diploma	95.7	1.04
Some college	96.3	1.03
Degree status in 1998		
Certificate or associate degree received	93.1	1.07
Vocational or associate degree program, degree not received	96.4	1.03
Bachelor's degree or program	97.0	1.03
Telephone numbers available		
0 or 1 number	93.6	1.07
2 numbers	94.9	1.06
3 numbers	96.4	1.04
4 numbers	96.2	1.04
5 numbers	97.6	1.02
6 numbers	95.7	1.04
7 or more numbers	98.1	1.02

See footnotes at end of table.

Table 6.2.—Average weight adjustment factors from the logistic model used to adjust for student refusal nonresponse—Continued

Predictor variables	Weighted response rate	Average weight adjustment
Number of times answering machine was encountered		
None	96.6	1.03
Once	97.8	1.02
More than once	95.6	1.04
In field cluster		
Yes	95.8	1.04
No	96.5	1.03
Interaction segment		
1 = Not prior respondent, not typical age for level	85.4	1.20
2 = Not prior respondent, typical age for level	92.3	1.09
3 = Prior respondent, 0 or 1 telephone numbers available	94.1	1.07
4 = Prior respondent, 2 telephone numbers available, encountered answering machine none or once	98.1	1.03
5 = Prior respondent, 2 telephone numbers available, encountered answering machine more than once	91.4	1.10
6 = Prior respondent, 3 telephone numbers available, encountered answering machine none or once	97.7	1.02
7 = Prior respondent, 3 telephone numbers available, encountered answering machine more than once	96.0	1.05
8 = Prior respondent, 4 or more telephone numbers available, 4-year institution	97.9	1.02
9 = Prior respondent, 4 or more telephone numbers available, 2-year or less-than-2-year institution	96.7	1.04

NOTE: Predictor variables are from base year data (NPSAS:96) with the exception of degree status in 1998, telephone numbers available, number of times answering machine was encountered, in field cluster, and certain interaction variables.

The weight used is LOCWT.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

c.   Weight Adjustments for Located Nonrespondents Who Were Not Refusals

Of the 9,259 who were eligible, located, and did not refuse, 9,132 responded to the BPS:1996/2001 survey and the remaining 127 did not respond for reasons other than refusal. Next, an adjustment was made to NREFWT to adjust for these 127. As in the case of the other adjustments, a CHAID analysis was performed on all of the predictor variables to detect important interactions. All potentially important variables were included in the model. Highly insignificant variables were deleted from the model until there was little change in the unequal weighting effect.

Table 6.3 presents the final predictor variables used in the logistic model that adjusted the weights for those who were interviewed, and gives the weighted interview rate (for those who were located and did not refuse) and the average weight adjustment factors resulting from these variables. The weighting adjustment factor for student j was the reciprocal of the predicted response probability, or

$$ADJ4_j = 1/p_{R,j}$$

and the weight was computed as:

B01AWT = NREFWT * ADJ4 for the 9,132 who responded, and
= 0 otherwise.

This final weight was rounded to the nearest integer and is denoted by B01AWT. This weight is to be used for analyzing the data collected from the 9,132 responses to BPS:1996/2001.

3.   Longitudinal Analysis Weights

Two longitudinal weights were constructed:

- one weight (B01LWT1) was computed for the 8,934 eligible NPSAS:96 sample members who responded to all three rounds of the survey (i.e., responded to NPSAS:96, BPS:96/98, and BPS:1996/2001); and

- the second weight (B01LWT2) was computed for the 8,999 eligible NPSAS:96 sample members who responded to both BPS:1996/2001 and NPSAS:96.

These two weights were each constructed by applying additional nonresponse adjustments to the final BPS:1996/2001 cross-sectional weight (i.e., B01AWT).

Table 6.3.—Average weight adjustment factors from the logistic model used to adjust for nonresponse other than refusal

Predictor variables	Weighted response rate	Average weight adjustment
Total	98.8	1.01
Age		
19 or younger	98.8	1.01
20 to 23	99.5	1.01
24 to 29	98.3	1.02
30 to 39	97.8	1.03
40 or older	99.6	1.01
Race/ethnicity		
White, non-Hispanic	99.0	1.01
Black, non-Hispanic	98.8	1.01
Hispanic	98.1	1.02
Asian/Pacific Islander	97.9	1.02
American Indian/Alaska Native	98.4	1.01
Other	98.7	1.02
Gender		
Male	98.5	1.02
Female	99.1	1.01
Attendance status		
Full-time/full year 1 institution	99.0	1.01
Full-time/full year more than 1 institution	99.0	1.01
Full-time/part year	98.3	1.02
Part-time/full year 1 institution	98.6	1.02
Part-time/full year more than 1 institution	100.0	1.00
Part-time/part year	99.0	1.02
Institution level		
4-year	98.6	1.01
2-year	99.2	1.01
Less-than-2-year	97.6	1.02
Institution control		
Public	99.1	1.01
Private not-for-profit	98.3	1.02
Private for-profit	97.8	1.02
Institution region		
New England	96.7	1.04
Mid East	98.5	1.02
Great Lakes	99.3	1.01
Plains	99.3	1.01
Southeast	99.3	1.01
Southwest	99.3	1.01
Rocky Mountains	99.5	1.01
Far West	98.0	1.02
Outlying Area	99.2	1.01
Type of institution and enrollment category		
Public		
Fewer than 1,000	98.9	1.01
1,000–2,499	99.7	1.01
2,500–4,999	99.8	1.00
5,000–9,999	99.1	1.01
10,000–19,999	98.6	1.02
20,000 or more	99.2	1.01
Private not-for-profit		
Fewer than 1,000	99.6	1.00
1,000–2,499	98.0	1.02
2,500–4,999	99.3	1.01
5,000–9,999	97.5	1.03
10,000 or more	98.1	1.02

See footnotes at end of table.

Table 6.3.—Average weight adjustment factors from the logistic model used to adjust for nonresponse other than refusal—Continued

Predictor variables	Weighted response rate	Average weight adjustment
Private for-profit		
Fewer than 300	98.9	1.01
300–999	98.0	1.02
1,000 or more	95.7	1.04
Applied for aid		
Yes	98.8	1.01
No	98.8	1.02
Receipt of federal aid		
Yes	98.9	1.01
No	98.7	1.02
Receipt of Pell Grant		
Yes	99.0	1.01
No	98.8	1.01
Receipt of Stafford Loan		
Yes	98.7	1.01
No	98.9	1.02
Receipt of state aid		
Yes	99.2	1.01
No	98.7	1.02
Receipt of institution aid		
Yes	98.6	1.02
No	98.9	1.01
Receipt of any aid		
Yes	98.9	1.01
No	98.7	1.02
Prior respondent		
Yes	98.8	1.01
No	99.7	1.00
Parents' income (for dependent students)		
$29,999 or less	98.5	1.02
$30,000–$59,999	98.9	1.01
$60,000–$99,999	99.1	1.01
$100,000 or more	98.6	1.02
Student's income (for independent students)		
$5,000 or less	98.5	1.02
$5,000–$9,999	98.7	1.02
$10,000–$19,999	99.3	1.01
$20,000 or more	98.9	1.02
Parents' highest education		
Less than high school	97.7	1.03
High school diploma	99.1	1.01
Some college	98.8	1.01
Degree status in 1998		
Certificate or associate degree received	99.2	1.01
Vocational or associate degree program, degree not received	98.9	1.01
Bachelor's degree or program	98.5	1.02
Telephone numbers available		
0 or 1 number	98.7	1.02
2 numbers	98.7	1.02
3 numbers	99.4	1.01
4 numbers	98.0	1.02
5 numbers	99.1	1.01
6 numbers	99.0	1.01
7 or more numbers	99.1	1.01
Number of times answering machine was encountered		
None	98.7	1.02
Once	99.5	1.01
More than once	98.7	1.02

See footnotes at end of table.

Table 6.3.—Average weight adjustment factors from the logistic model used to adjust for nonresponse other than refusal—Continued

Predictor variables	Weighted response rate	Average weight adjustment
In field cluster		
Yes	98.4	1.02
No	99.2	1.01
Interaction segment		
1 = U.S. citizen, male	98.8	1.01
2 = U.S. citizen, female	99.1	1.01
3 = Non-citizen, public institution	97.9	1.03
4 = Non-citizen, private institution, male	86.4	1.16
5 = Non-citizen, private institution, female	97.6	1.03

NOTE: Predictor variables are from base year data (NPSAS:96) with the exception of prior respondent, degree status in 1998, telephone numbers available, number of times answering machine was encountered, and in field cluster.

The weight used is NREFWT.

SOURCE: U.S. Department of Education, national Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

The weight for analyzing respondents to all three surveys, NPSAS:96, BPS:96/98, and BPS:1996/2001, was constructed by applying an additional nonresponse adjustment to the final unrounded BPS:1996/2001 cross-sectional weight (unrounded version of B01AWT). As for the other models, CHAID was used to determine the interaction segments, then the GEM modeling procedure was used to determine the adjustment factor. **Table 6.4** presents the final predictor variables used in the logistic model that adjusted the weights for those who were not also interviewed in both NPSAS:96 and BPS:96/98 and gives the weighted interview rate (for those who were interviewed in BPS:1996/2001) and the average weight adjustment factors resulting from these variables. The final weight was rounded to integer values, and is denoted as B01LWT1.

Specifically,

B01LWT1 = B01AWT * ADJ5 for those who responded to all three surveys, and

= 0 otherwise,

where

$ADJ5 = 1/p_{96\text{-}98\text{-}01,j}$

is the reciprocal of the predicted response probability.

The weight for analyzing respondents to both NPSAS:96 and BPS:1996/2001 was also constructed by applying an additional nonresponse adjustment to the final BPS:1996/2001 cross-sectional weight, following the same steps as for the other adjustments. **Table 6.5** presents the final predictor variables used in the model and the weighted response rates and adjustment factors. The final weight was rounded to integer values and is denoted as B01LWT2.

Specifically,

B01LWT2 = B01AWT * ADJ6 for those who responded to both NPSAS:96

and BPS:1996/2001, and

= 0 otherwise,

where

$ADJ6 = 1/p_{96\text{-}01,j}$

is the reciprocal of the predicted response probability.

The distributions of the weight adjustment factors for the BPS:1996/2001 analysis weights and the two longitudinal weights are presented in **table 6.6**. **Table 6.7** presents the distributions of the initial, intermediate, and final weights along with their unequal weighting design effects.

Table 6.4.—Average weight adjustment factors from the logistic model used to adjust for nonresponse to either NPSAS:96 or BPS:96/98, among the respondents to BPS:1996/2001

Predictor variables	Weighted response rate	Average weight adjustment
Total	93.6	1.05
Age		
19 or younger	93.7	1.05
20 to 23	94.9	1.05
24 to 29	91.7	1.10
30 to 39	96.5	1.03
40 or older	89.9	1.09
Race/ethnicity		
White, non-Hispanic	93.4	1.05
Black, non-Hispanic	94.0	1.06
Hispanic	94.5	1.05
Asian/Pacific Islander	93.7	1.05
American Indian/Alaska Native	93.9	1.05
Other	94.7	1.05
Gender		
Male	93.5	1.05
Female	93.7	1.05
Citizenship		
U.S. citizen	93.7	1.05
Non-citizen, eligible	94.7	1.04
Non-citizen, ineligible	81.3	1.22
Attendance status		
Full-time/full year 1 institution	94.4	1.05
Full-time/full year more than 1 institution	100.0	1.00
Full-time/part year	95.4	1.05
Part-time/full year 1 institution	91.5	1.08
Part-time/full year more than 1 institution	100.0	1.00
Part-time/part year	90.4	1.10
Institution level		
4-year	95.4	1.05
2-year	92.2	1.08
Less-than-2-year	93.7	1.06
Institution control		
Public	93.0	1.06
Private not-for-profit	97.6	1.02
Private for-profit	92.2	1.09
Institution region		
New England	93.6	1.06
Mid East	96.3	1.03
Great Lakes	93.6	1.05
Plains	91.9	1.07
Southeast	94.9	1.05
Southwest	90.6	1.09
Rocky Mountains	93.4	1.06
Far West	92.3	1.07
Outlying Area	97.9	1.02
Type of institution and enrollment category		
Public		
Fewer than 1,000	100.0	1.00
1,000–2,499	94.7	1.05
2,500–4,999	93.6	1.05
5,000–9,999	95.2	1.04
10,000–19,999	90.6	1.09
20,000 or more	92.9	1.07

See footnotes at end of table.

Table 6.4.—Average weight adjustment factors from the logistic model used to adjust for nonresponse to either NPSAS:96 or BPS:96/98, among the respondents to BPS:1996/2001—Continued

Predictor variables	Weighted response rate	Average weight adjustment
Private not-for-profit		
Fewer than 1,000	98.5	1.02
1,000–2,499	97.2	1.03
2,500–4,999	98.1	1.02
5,000–9,999	98.4	1.01
10,000 or more	97.0	1.03
Private for-profit		
Fewer than 300	93.1	1.07
300–999	90.5	1.11
1,000 or more	93.9	1.06
Applied for aid		
Yes	94.7	1.05
No	91.2	1.08
Receipt of federal aid		
Yes	94.2	1.05
No	93.2	1.06
Receipt of Pell Grant		
Yes	93.7	1.06
No	93.6	1.05
Receipt of Stafford Loan		
Yes	93.9	1.05
No	93.5	1.05
Receipt of state aid		
Yes	95.5	1.04
No	93.3	1.06
Receipt of institution aid		
Yes	95.3	1.04
No	93.2	1.06
Receipt of any aid		
Yes	94.6	1.05
No	92.3	1.07
Parents' income (for dependent students)		
$29,999 or less	93.4	1.06
$30,000–$59,999	94.2	1.05
$60,000–$99,999	93.1	1.06
$100,000 or more	94.2	1.05
Student's income (for independent students)		
$5,000 or less	93.6	1.06
$5,000–$9,999	93.4	1.07
$10,000–$19,999	97.5	1.02
$20,000 or more	90.8	1.09
Parents' highest education		
Less than high school	93.0	1.06
High school diploma	94.1	1.05
Some college	93.4	1.06
Degree status in 1998		
Certificate or associate degree received	93.2	1.07
Vocational or associate degree program, degree not received	92.5	1.07
Bachelor's degree or program	95.3	1.05
Interaction segment		
1 = Did not apply for aid, public	94.4	1.06
2 = Did not apply for aid, private not-for-profit	97.8	1.02
3 = Did not apply for aid, private for-profit	91.6	1.09
4 = Applied for aid, public	90.6	1.09
5 = Applied for aid, private not-for-profit	96.3	1.04
6 = Applied for aid, private for-profit	97.2	1.04

NOTE: Predictor variables are from base year data (NPSAS:96) with exception of degree status in 1998. Weight used is B01AWT.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.5.—Average weight adjustment factors from the logistic model used to adjust for nonresponse to NPSAS:96, among the respondents to BPS:1996/2001

Predictor variables	Weighted response rate	Average weight adjustment
Total	96.3	1.04
Age		
19 or younger	96.1	1.04
20 to 23	96.7	1.03
24 to 29	95.2	1.05
30 to 39	97.6	1.02
40 or older	97.9	1.02
Race/ethnicity		
White, non-Hispanic	96.3	1.04
Black, non-Hispanic	96.4	1.04
Hispanic	96.0	1.04
Asian/Pacific Islander	97.2	1.03
American Indian/Alaska Native	93.9	1.05
Other	94.7	1.04
Gender		
Male	95.9	1.04
Female	96.7	1.03
Citizenship		
U.S. citizen	96.3	1.04
Non-citizen, eligible	95.8	1.05
Non-citizen, ineligible	100.0	1.00
Attendance status		
Full-time/full year 1 institution	96.7	1.04
Full-time/full year more than 1 institution	100.0	1.00
Full-time/part year	97.0	1.03
Part-time/full year 1 institution	93.9	1.07
Part-time/full year more than 1 institution	100.0	1.00
Part-time/part year	96.0	1.04
Institution level		
4-year	96.3	1.04
2-year	96.3	1.03
Less-than-2-year	96.4	1.03
Institution control		
Public	95.8	1.05
Private not-for-profit	98.4	1.02
Private for-profit	96.5	1.03
Institution region		
New England	97.4	1.03
Mid East	98.3	1.02
Great Lakes	96.2	1.04
Plains	95.7	1.04
Southeast	97.5	1.03
Southwest	93.2	1.07
Rocky Mountains	96.4	1.04
Far West	94.9	1.05
Outlying Area	97.9	1.02
Type of institution and enrollment category		
Public		
Fewer than 1,000	100.0	1.00
1,000–2,499	97.3	1.03
2,500–4,999	96.6	1.03
5,000–9,999	97.0	1.03
10,000–19,999	94.8	1.06
20,000 or more	95.3	1.05
Private not-for-profit		
Fewer than 1,000	99.0	1.01
1,000–2,499	98.4	1.02
2,500–4,999	98.6	1.01
5,000–9,999	98.7	1.01
10,000 or more	97.7	1.02
Private for-profit		
Fewer than 300	95.3	1.04
300–999	97.0	1.03
1,000 or more	97.5	1.03

See footnotes at end of table.

Table 6.5.—Average weight adjustment factors from the logistic model used to adjust for nonresponse to NPSAS:96, among the respondents to BPS:1996/2001—Continued

Predictor variables	Weighted response rate	Average weight adjustment
Receipt of federal aid		
Yes	96.3	1.04
No	96.3	1.03
Receipt of Pell Grant		
Yes	95.9	1.04
No	96.5	1.03
Receipt of Stafford Loan		
Yes	96.2	1.04
No	96.3	1.04
Receipt of state aid		
Yes	96.9	1.03
No	96.2	1.04
Receipt of institution aid		
Yes	96.3	1.04
No	96.3	1.04
Receipt of any aid*		
Yes	96.5	1.04
No	96.0	1.04
Parents' income (for dependent students)		
$29,999 or less	95.1	1.05
$30,000–$59,999	96.5	1.03
$60,000–$99,999	95.8	1.04
$100,000 or more	97.9	1.02
Student's income (for independent students)		
$5,000 or less	97.1	1.03
$5,000–$9,999	95.9	1.04
$10,000–$19,999	98.2	1.02
$20,000 or more	96.4	1.03
Parents' highest education		
Less than high school	94.2	1.07
High school diploma	96.9	1.03
Some college	96.2	1.04
Degree status in 1998		
Certificate or associate degree received	97.3	1.03
Vocational or associate degree program, degree not received	96.0	1.04
Bachelor's degree or program	96.3	1.04
Interaction segment		
1 = Did not apply for aid	95.0	1.05
2 = Applied for aid, NPSAS:96 school greater than 13,184 enrolled	95.2	1.05
3 = Applied for aid, NPSAS:96 school 13,184 or fewer enrolled, parents' highest education less than high school	93.0	1.07
4 = Applied for aid, NPSAS:96 school 13,184 or fewer enrolled, parents' highest education high school or more	98.2	1.02

NOTE: Predictor variables are from base year data (NPSAS:96) with the exception of degree status in 1998.

The weight used is B01AWT.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.6.—Distribution of values for the BPS:1996/2001 weight adjustment factors

Quantile	Location ADJ2	Refusal ADJ3	Nonresponse ADJ4	Longitudinal ADJ5	Longitudinal AJD6
Minimum	0.05	0.56	0.91	0.89	0.87
1%	0.26	1.00	1.00	1.00	1.00
5%	0.94	1.00	1.00	1.00	1.00
10%	1.04	1.00	1.00	1.01	1.00
25%	1.18	1.01	1.00	1.02	1.01
Median	1.38	1.02	1.01	1.04	1.02
75%	1.68	1.05	1.02	1.07	1.05
90%	2.05	1.09	1.04	1.12	1.09
95%	2.29	1.12	1.06	1.15	1.11
99%	2.65	1.19	1.15	1.24	1.17
Maximum	3.49	1.30	1.20	1.42	1.35

NOTE: Some adjustment factors are less than one because the GEM weight adjustment procedure adjusts for nonresponse, and truncates and smooths the weights in one step.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.7.—Distribution of initial, intermediate, and final weights for BPS:1996/2001

Quantile	Initial weight B01_100	Intermediate weights Location LOCWT	Intermediate weights Refusal NREFWT	Cross-sectional analysis weight B01AWT	Longitudinal analysis weights B01LWT1	Longitudinal analysis weights B01LWT2
Minimum	7	8	8	8	8	8
1%	22	27	27	28	28	28
5%	38	50	51	52	54	53
10%	47	67	68	69	72	71
25%	76	113	116	118	120	120
Median	144	196	202	205	211	209
75%	225	324	334	340	347	344
90%	680	782	795	800	824	823
95%	1,151	1,278	1,333	1,354	1,392	1,384
99%	2,964	3,125	3,185	3,197	2,987	3,140
Maximum	12,897	5,588	4,453	4,340	4,357	4,371
Unequal weighting design effect	7.245	3.020	3.059	3.047	3.024	3.026

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

B. Variance Estimation

For probability-based sample surveys, most estimates are nonlinear statistics. For example, a mean or proportion is calculated as $\Sigma wy/\Sigma w$, which is nonlinear because the denominator is a survey estimate of the unknown population total. In this situation, the variances of the estimates cannot be expressed in closed form. Two common procedures for estimating the variances of nonlinear survey statistics are Taylor series linearization procedures and replication methods. The replication method used in BPS:1996/2001 is balanced repeated replication, or BRR. BRR is used because of its superiority for the estimation of the variances of quantiles, such as medians. The subsections below discuss the Taylor series and BRR methods of variance estimation for BPS:1996/2001.

1. Taylor Series

The Taylor series variance estimation procedure is a well-known technique for estimating variances of nonlinear statistics. The procedure substitutes the first-order Taylor series approximation of the nonlinear statistic into the variance formula based on the sampling design. Woodruff (1971)[3] presents the mathematical formulation of this procedure.

For stratified, multistage sampling designs, the Taylor series procedure requires analysis strata and analysis replicates based on the first-stage sampling design. Since the BPS:96 cohort is a subset of the NPSAS:96 sample, the first stage of the sampling design was the first stage of the NPSAS:96 sample. Hence, the analysis strata and analysis replicates for BPS:1996/2001 were defined from those computed for the NPSAS:96 undergraduate student sample. In fact, the BPS:1996/2001 analysis strata, B01ASTR, are identical to the 51 NPSAS:96 undergraduate analysis strata, UANALSTR. Within analysis strata, adjacent NPSAS:96 analysis replicates were collapsed to form BPS:1996/2001 analysis replicates, B01AREP, so that each contained at least four BPS:1996/2001 respondents. Thus, the variables that are to be used to denote the analysis strata and analysis replicates in software packages that use Taylor series variance estimation are B01ASTR and B01AREP.

The following summarizes the variable names for the weights, analysis strata, and analysis replicates for use with the Taylor series variance estimation on the BPS:1996/2001 data file:

B01ASTR   BPS:1996/2001 analysis strata

B01AREP   BPS:1996/2001 analysis replicates

B01AWT   BPS:1996/2001 analysis weight for 2001 respondents, for cross-sectional analyses

B01LWT1   longitudinal analysis weight for 1996, 1998, and 2001 respondents

B01LWT2   longitudinal analysis weight for 1996 and 2001 respondents

---

[3] Woodruff, R.S. (1971). "A simple method for approximating the variance of a complicated estimate." *Journal of the American Statistical Association*, 66, pp. 411–4.

Table 6.8 summarizes the variables and how they are used in selected software packages that allow for Taylor series for variance estimation (SUDAAN, STATA, and the SAS procedures SURVEYMEANS and SURVEYREG).

Table 6.8.—Analysis weight, strata, and replicate variables that are available from BPS:1996/2001

	Type of analysis		
	Cross-sectional	Longitudinal	
	BPS:1996/2001 respondents	Respondents to all three of NPSAS:96, BPS:96/98, and BPS:1996/ 2001	Respondents to both NPSAS:96 and BPS:1996/2001
Weight variables for estimates	B01AWT	B01LWT1	B01LWT2
Taylor series variance estimation			
Variables denoting analysis strata and replicates	B01ASTR B01AREP	B01ASTR B01AREP	B01ASTR B01AREP
SUDAAN	DESIGN=WR WEIGHT B01AWT; NEST B01ASTR B01AREP;	DESIGN=WR WEIGHT B01LWT1; NEST B01ASTR B01AREP;	DESIGN=WR WEIGHT B01LWT2; NEST B01ASTR B01AREP;
STATA	PWEIGHT B01AWT STRATA B01ASTR PSU B01AREP	PWEIGHT B01LWT1 STRATA B01ASTR PSU B01AREP	PWEIGHT B01LWT2 STRATA B01ASTR PSU B01AREP
SAS SURVEYMEANS and SURVEYREG	WEIGHT B01AWT; STRATA B01ASTR; CLUSTER B01AREP;	WEIGHT B01LWT1; STRATA B01ASTR; CLUSTER B01AREP;	WEIGHT B01LWT2; STRATA B01ASTR; CLUSTER B01AREP;
BRR variance estimation			
Replicate weight variables	B01BRR01 – B01BRR51	B1LBRR01 – B1LBRR51	B2LBRR01 – B2LBRR51
SUDAAN	DESIGN=BRR WEIGHT B01AWT; REPWGT B01BRR01 – B01BRR51;	DESIGN=BRR WEIGHT B01LWT1; REPWGT B1LBRR01 – B1LBRR51;	DESIGN=BRR WEIGHT B01LWT2; REPWGT B2LBRR01 – B2LBRR51;
WESVAR	Method BRR Full sample weight B01AWT Replicates B01BRR01 – B01BRR51	Method BRR Full sample weight B01LWT1 Replicates B1LBRR01 – B1LBRR51	Method BRR Full sample weight B01LWT2 Replicates B2LBRR01 – B2LBRR51

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

2. Balanced Repeated Replication

BRR is one of two replication techniques commonly used to estimate the variances of survey statistics computed from complex sample surveys. (The other commonly used replication technique is the jackknife replication technique.) Wölter (1985) reviews both the Taylor series and replication techniques[4].

---

[4] Wölter, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.

The BRR method is designed for a survey with L primary sampling strata and two primary sampling units (PSUs) selected per stratum. A half-sample replicate is formed by selecting one PSU from each stratum. For any given sample, there are 2^L such half-samples. If $\bar{y}_{st,\alpha}$ represents the estimate of the population mean calculated from the α-th replicate and $\bar{y}_{st}$ represents the stratified mean from the full sample, then the mean of $(\bar{y}_{st,\alpha} - \bar{y}_{st})^2$ over all 2^L half samples is identical to the textbook stratified variance estimator. BRR is essentially a method for selecting a set of k "balanced" replicates where k is much smaller than 2^L so that this same property holds for the set of k replicates (see chapter 3 of Wölter, 1985). The BRR variance estimate is then computed as:

$$\text{Var}_{\text{BRR}}\ (\bar{y}_{st}) = \sum_{\alpha=1}^{k} \frac{(\bar{y}_{st,\alpha} - \bar{y}_{st})^2}{k} \ .$$

BRR weights were computed for BPS:1996/2001 because of concern that the variances for medians and other quartiles may not be appropriate when computed using either Taylor series or jackknife methods. The Taylor series approach estimates the cumulative distribution function at several points and then estimates variances for quartiles through inverse interpolation (see Francisco and Fuller, 1991)[5]. Because these results depend on the points at which the cumulative distribution function and its variances are evaluated, they are subjective and require considerable care by the user. Likewise jackknife methods are inconsistent for estimating the variances of nonsmooth functions, such as quartiles (see chapter 3 of Efron, 1982)[6]; as the sample size increases, the estimates do not converge to the true value. Moreover, the resulting jackknife variance estimator has only two degrees of freedom, irrespective of the sample size.

Computation of BRR weights. As mentioned above, the BRR method is designed for surveys with two PSUs per stratum. Because the NPSAS:96 was not a two-PSU-per-stratum design, the first task was to approximate the design for variance estimation purposes as one with two analysis PSUs per stratum. Fortunately, that problem had already been solved when the NPSAS:96 jackknife weights were computed. As explained in section 6.4.2 of the NPSAS:96 Methodology Report, when computing the jackknife weights, two such sets of pseudo-strata were developed:

- 51 strata for all-student and undergraduate student analyses, and
- 42 strata for graduate/first-professional analyses.

Instead of continuing with jackknife weights, BRR weights were computed because of the superiority of BRR variance estimation for medians and other quartiles, and estimates of quartiles and medians for amounts of student aid received are important survey estimates.

The $L = 51$ pseudo-strata defined for undergraduate students were used to compute BRR weights based on the initial weights for the 2001 follow-up of the BPS:96 cohort, namely B01_100. Wölter (1985) explains that to achieve "full orthogonal balance," k half-sample

---

[5] Francisco, C. A. and Fuller, W. A. (1991). "Quantile estimation with a complex survey design." *Annals of Statistics*, 19, 454–69.

[6] Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans.* Philadelphia: Society for Industrial and Applied Mathematics.

replicates should be used where $k > L$ and k is a multiple of 4. Since 13*4=52, $k = 52$ was used. As Wölter further explains, any 52x52 Hadamard matrix can be used to define the 52 balanced half-samples. In particular, any 52 rows (or columns) can be used to represent the 52 BRR replicates and any 51 columns (or rows) can be used to represent the 51 NPSAS:96 pseudo-strata. The rationale for 51 pseudo-strata (instead of 52) is explained in the following paragraph.

Although all $k = 52$ balanced replicates are needed to achieve "full orthogonal balance," using the full set of 52 replicates results in 52 degrees of freedom for the error variance. Since a two-PSU-per-stratum design with 51 strata only has 51 degrees of freedom for error, using 52 replicates could result in spurious indications of statistical significance. Therefore, $L = 51$ replicates were used, instead of 52 replicates. This results in a small positive bias in the variance estimate and, hence, conservative hypothesis test results.

The same Hadamard matrix that had been used to compute the BRR weights for NPSAS:96 and BPS:96/98 was used for BPS:1996/2001. The initial matrix was shown to be a 52x52 Hadamard matrix by verifying that $\mathbf{H^T H} = 52\mathbf{I}$. The same 51 columns that were used for NPSAS:96 (deleting an identity column) were used to identify 51 BRR replicate samples, as discussed below.

Using Wölter's notation (with rows and columns reversed), let $\delta_h^{(\alpha)}$ denote the element of the 52x52 Hadamard matrix in row h and column α. The '+1' and '−1' elements of the matrix were used to define 51 initial balanced replicate weights from WBPSBASE, and the NPSAS:96 jackknife replicate and stratum variables, JACKREP and JACKSTR, as follows:

+1 ==> the α-th BRR replicate contains the pseudo-replicate 1 observation from pseudo-stratum h (BRRWTα = 2 times WBPSBASE if JACKREP=1; BRRWTα = 0 if JACKREP=2); and

−1 ==> the α-th BRR replicate contains the pseudo-replicate 2 observations from pseudo-stratum h (BRRWTα = 2 times WBPSBASE if JACKREP=2; BRRWTα = 0 if JACKREP=1).

From each of the 51 BRR initial replicate weights defined in this manner, the final BRR replicate weight was computed using exactly the same weight adjustment procedures that had been implemented for the full BPS sample, except that the bounds were increased when necessary in order for the models to converge. Three sets of BRR weights were computed. The final BRR weights, rounded to integer values, are as follows:

B01BRR01–B01BRR51 are the BRR weights for the 2001 respondents, to be used for cross-sectional analyses;

B1LBRR01–B1LBRR51 are the BRR weights for respondents to the 1996, 1998, and 2001 surveys, to be used for longitudinal analyses; and

B2LBRR01–B2LBRR51 are the BRR weights for respondents to the 1996 and 2001 surveys, to be used for longitudinal analyses.

Table 6.8 summarizes the variables and how they are used in selected software packages that allow for BRR variance estimation (SUDAAN and WESVAR).

C.   Accuracy of Estimates

The accuracy of survey statistics is affected by both random and nonrandom errors. Random errors reduce the precision of survey statistics, while nonrandom errors result in bias (i.e., estimates that do not converge to the true population parameter as the sample size increases without limit).

The sources of error in a survey are often dichotomized as sampling and nonsampling errors. Sampling error refers to the error that occurs simply because the survey is based on a sample of population members, rather than the entire population. All other types of errors are nonsampling errors, including survey nonresponse (because of inability to contact sampling members, their refusal to participate in the study, etc.) and measurement errors, such as the errors that occur because the intent of survey questions was not clear to the respondent, because the respondent had insufficient knowledge to answer correctly, or because the data were not captured correctly (e.g., because of recording, editing, or data entry errors).

Sampling errors are primarily random errors for well-designed surveys like NPSAS:96, BPS:96/98, and BPS:1996/2001. However, nonrandom errors can occur also if the sampling frame does not provide complete coverage of the target population. The BPS survey instrument and data collection procedures were subjected to thorough development and testing to minimize nonsampling errors because these errors are difficult to quantify and are likely to be nonrandom errors.

In this section sampling errors and design effects for some BPS:1996/2001 estimates are presented for a variety of domains. Next the results of analyses comparing BPS:1996/2001 nonrespondents and respondents using characteristics known for nonrespondents as well as respondents are presented. Finally, the pattern of response by date of response is modeled to see if late respondents tend to be different from early respondents.

1.   Measures of Precision: Standard Errors and Design Effects

The cumulative effect of random errors on the precision of a survey statistic is measured by the standard error of that statistic. The standard error of a statistic is the estimated standard deviation of the sampling distribution of the statistic over repeated samples of the same size using the same sampling design. Hence, the standard error of a survey statistic depends not only on the natural variability of the observations in the population and on the sample size but also on the characteristics of the sampling design. Features of the sampling design that affect the sampling variance of a survey statistic (the square of the standard error) include stratification, multistage or cluster sampling, and unequal sampling rates. Stratification can increase precision if outcomes are more homogeneous within strata than between strata, but the other survey design features usually decrease precision. Moreover, statistical adjustment of the analysis weights to reduce the potential for bias due to nonresponse often decreases precision.

The cumulative effect of the various factors affecting the precision of a survey statistic is often modeled as the survey design effect. The design effect, designated as DEFF, is defined as the ratio of the sampling variance of the statistic under the actual sampling design divided by the variance that would be expected for a simple random sample of the same size. The square root of the design effect (also called the root design effect, and designated as DEFT) is also useful. The following formulas define the design effects and root design effects for this section:

$$\text{DEFF}(\hat{\theta}) = \frac{\text{Var}(\hat{\theta})}{\text{Var}_{\text{SRS}}(\hat{\theta})}$$

$$\text{DEFT}(\hat{\theta}) = \frac{\text{SE}(\hat{\theta})}{\text{SE}_{\text{SRS}}(\hat{\theta})}$$

In these formulas, $\hat{\theta}$ represents the survey statistic of interest (*e.g.*, estimated proportion of the population still enrolled in an undergraduate program). Hence, the design effect is unity (1.00), by definition, for simple random samples. For most practical sampling designs, the survey design effect is greater than unity, reflecting that the precision is less than could be achieved with a simple random sampling of the same size (if such a design were practical). The size of the survey design effect depends largely on the sample size and intracluster correlation within the primary sampling units (e.g., number of students per institution and within-institution correlations). Hence, statistics that are based on observations that are highly correlated within institutions will have higher design effects for BPS.

The simple random sample variance was computed conditional on the sample size of the analysis domain. Specifically, if n_d is the respondent sample size in the domain and $\hat{\theta}_d$ is the weighted estimate of the proportion for the domain, then the simple random sample variance was

computed as $\text{Var}_{\text{SRS}}(\hat{\theta}_d) = \dfrac{\hat{\theta}_d(1-\hat{\theta}_d)}{n_d}$.

In order to provide an approximate characterization of the precision with which BPS:1996/2001 survey statistics can be estimated, a series of tables was prepared that provide estimates of key statistics, their standard errors, and the estimated survey design effects. **Appendix G** presents a variety of survey estimates for domains defined by

- all respondents,
- age,
- race/ethnicity,
- gender,
- level of institution in the base year,
- control of institution in the base year,
- whether the respondent had received a degree by June 2001,
- employment status,
- highest degree, and
- whether the respondent is the first generation in postsecondary education.

The tables give the percentage estimates, the design based standard errors (produced using Taylor series and SUDAAN[7] [Release 8.0]), the denominator sample size, and DEFF and DEFT. The tables also give the mean, minimum, and maximum values of DEFF and DEFT for each

---

[7] Research Triangle Institute. (2001). *SUDAAN User's Manual, Release 8.0.* Research Triangle Park, NC: Research Triangle Institute.

domain. Variables with fewer than 30 respondents in the denominator for a particular domain were not included in the tables.

2. Measures of Bias

a. Nonresponse Bias Analysis

Unit nonresponse causes bias in survey estimates when the outcomes of respondents and nonrespondents are different. A bias analysis was conducted to determine if any variables were significantly biased due to nonresponse. Three types of nonresponse bias analysis were considered:

- nonrespondents versus respondents;
- early refusals who were later converted to respondents versus other respondents; and
- late respondents (those who responded between July and September 2001) versus earlier respondents.

For the first of these, respondents and nonrespondents were characterized by comparing the weighted[8] percentage of respondents with the weighted percentage of nonrespondents for each category of important characteristics known for both respondents and nonrespondents. T-tests were performed to determine if the difference between respondents and nonrespondents was significant at the 5 percent level. **Table 6.9** compares the demographic characteristics of respondents and nonrespondents. This table shows that the distributions of demographic characteristics, such as typical age for level, attendance status, institution level, control, and receipt of state aid are significantly different for respondents and nonrespondents.

Table 6.10 performs a similar analysis, but compares demographic characteristics of those respondents who initially refused but were later converted to respondents with other respondents. The refusals who were converted are likely similar to the refusal nonrespondents who were not converted. This analysis shows that the distribution of demographic characteristics such as race/ethnicity, gender, attendance status, institution level, and receipt of various types of aid are significantly different for the converted refusals versus other respondents.

Table 6.11 compares the distributions of those who responded early (June 30, 2001, or earlier) with those who responded later (July through September 2001). This analysis shows that the distribution of demographic characteristics such as institution level, institution control, receipt of various types of aid, and whether the student was a prior respondent are significantly different for the early versus late respondents.

The nonresponse bias was estimated for variables known for both respondents and nonrespondents. The bias in an estimated mean based on respondents, $\bar{y}_R$, was also estimated as the difference between this mean and the target parameter, π, being estimated, i.e., the mean that would be estimated if a complete census of the target population were conducted. This bias can be expressed as follows:

$$B(\bar{y}_R) = \bar{y}_r - \pi.$$

---

[8] The base weights, B01_100, were used. Missing values were excluded for most of the demographic variables; for this reason, some percentages may not sum to 100. The exception is that imputed values were used for the income valuables.

Table 6.9.—Comparison of BPS:1996/2001 respondents and nonrespondents

Demographic characteristics	Respondents Percent estimate[1]	Nonrespondents Percent estimate[1]	Full sample Percent estimate[1]
Age			
19 or younger	68.95	57.37 *	67.05
20 to 23	11.54	21.78	13.22
24 to 29	8.21	6.53	7.94
30 to 39	5.39	10.24	6.19
40 or older	5.90	4.08	5.60
Typical age for level			
No, not typical age for level	30.63	42.29 *	32.54
Yes, typical age for level	68.62	56.90 *	66.70
Race/ethnicity			
White, non-Hispanic	72.12	69.47	71.69
Black, non-Hispanic	11.58	11.94	11.64
Hispanic	9.95	10.66	10.07
Asian/Pacific Islander	5.03	6.35	5.25
American Indian/Alaska Native	0.71	0.99	0.76
Other	0.60	0.59	0.60
Gender			
Male	44.57	48.21	45.17
Female	55.43	51.79	54.83
Citizenship			
U.S. citizen	93.72	91.79	93.40
Non-citizen, eligible	4.92	4.79	4.90
Non-citizen, ineligible	0.61	2.61	0.94
Attendance status			
Full-time/full year 1 institution	50.79	42.48	49.42
Full-time/full year more than 1 institution	1.73	0.74 *	1.57
Full-time/part year	14.62	28.20 *	16.85
Part-time/full year 1 institution	14.44	12.31	14.09
Part-time/full year more than 1 institution	0.53	1.95	0.76
Part-time/part year	15.66	13.13	15.24
Institution level			
4-year	41.99	37.27	41.21
2-year	50.15	48.28	49.84
Less-than-2-year	7.86	14.45 *	8.94
Institution control			
Public	74.32	67.10	73.13
Private not-for-profit	15.89	15.54	15.84
Private for-profit	9.79	17.36 *	11.03
Institution region			
New England	5.64	7.50	5.94
Mid East	13.68	11.98	13.40
Great Lakes	15.91	13.06	15.44
Plains	7.9	9.67	8.19
Southeast	23.87	24.52	23.98
Southwest	11.30	11.83	11.39
Rocky Mountains	3.41	4.20	3.54
Far West	15.94	15.51	15.87
Outlying Area	1.62	0.93	1.50
Type of institution and enrollment category			
Public			
Fewer than 1,000	1.97	1.08	1.83
1,000–2,499	3.63	7.49	4.26
2,500–4,999	7.85	4.57	7.31
5,000–9,999	13.08	23.44	14.78
10,000–19,999	23.75	16.23	22.52
20,000 or more	24.04	14.29 *	22.44
Private not-for-profit			
Fewer than 1,000	1.82	2.20	1.88
1,000–2,499	5.39	4.51	5.25
2,500–4,999	2.81	3.13	2.86
5,000–9,999	2.17	2.49	2.22
10,000 or more	3.71	3.23	3.63

See footnotes at end of table.

Table 6.9.—Comparison of BPS:1996/2001 respondents and nonrespondents—Continued

Demographic characteristics	Respondents Percent estimate[1]	Nonrespondents Percent estimate[1]	Full sample Percent estimate[1]
Private for-profit			
Fewer than 300	3.40	6.47	3.90
300–999	4.53	6.03	4.78
1,000 or more	1.85	4.85	2.34
Applied for aid			
Yes	69.57	72.01	69.97
No	29.68	27.19	29.27
Receipt of federal aid			
Yes	44.52	47.79	45.06
No	54.73	51.40	54.18
Receipt of Pell Grant			
Yes	27.28	32.29	28.11
No	71.97	66.90	71.13
Receipt of Stafford Loan			
Yes	28.79	31.67	29.26
No	70.46	67.52	69.98
Receipt of state aid			
Yes	15.20	11.29 *	14.56
No	84.05	87.90 *	84.68
Receipt of institution aid			
Yes	22.32	18.43	21.68
No	76.93	80.77	77.56
Receipt of any aid			
Yes	59.66	59.48	59.63
No	39.59	39.71	39.61
Prior respondent			
Yes	77.87	56.90 *	74.42
No	22.13	43.10 *	25.58
Parents' income (for dependent students)			
$29,999 or less	22.13	23.90	22.42
$30,000–$59,999	26.79	19.59	25.60
$60,000–$99,999	16.34	13.75	15.92
$100,000 or more	7.66	8.51	7.80
Student's income (for independent students)			
$5,000 or less	5.40	6.52	5.58
$5,000–$9,999	4.13	5.76	4.40
$10,000–$19,999	6.21	14.57	7.59
$20,000 or more	10.58	6.60	9.93

*Difference between respondents and nonrespondents is significant at the $p<0.05/(c-1)$ level, where the (c-1) divisor corrects for the (c-1) multiple comparisons among the c levels of the variable.

[1] The weight used is B01_100.

NOTE: Characteristics are from base year data (NPSAS:96) with the exception of the prior respondent category. Missing values are not shown in the table, and therefore some percentages may not sum to 100% for a variable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.10.—Comparison of BPS:1996/2001 converted refusals and other respondents

Demographic characteristics	Converted refusals Percent estimate[1]	Other respondents Percent estimate[1]	Total respondents Percent estimate[1]
Age			
19 or younger	72.14	68.37	68.95
20 to 23	10.68	11.70	11.54
24 to 29	5.85	8.64	8.21
30 to 39	5.84	5.31	5.39
40 or older	5.49	5.98	5.90
Typical age for level			
No, not typical age for level	27.86	31.13	30.63
Yes, typical age for level	72.14	67.98	68.62
Race/ethnicity			
White, non-Hispanic	79.44	70.79 *	72.12
Black, non-Hispanic	11.42	11.61	11.58
Hispanic	5.08	10.84 *	9.95
Asian/Pacific Islander	2.56	5.48 *	5.03
American Indian/Alaska Native	1.19	0.62	0.71
Other	0.31	0.65	0.60
Gender			
Male	52.65	43.10 *	44.57
Female	47.35	56.90 *	55.43
Citizenship			
U.S. citizen	95.08	93.47	93.72
Non-citizen, eligible	3.15	5.24	4.92
Non-citizen, ineligible	1.77	0.40	0.61
Attendance status			
Full-time/full year 1 institution	43.24	52.16	50.79
Full-time/full year more than 1 institution	1.26	1.82	1.73
Full-time/part year	15.06	14.54	14.62
Part-time/full year 1 institution	13.92	14.53	14.44
Part-time/full year more than 1 institution	0.49	0.54	0.53
Part-time/part year	25.39	13.89 *	15.66
Institution level			
4-year	33.78	43.48 *	41.99
2-year	57.38	48.83 *	50.15
Less-than-2-year	8.83	7.69	7.86
Institution control			
Public	76.28	73.96	74.32
Private not-for-profit	13.82	16.27	15.89
Private for-profit	9.9	9.77	9.79
Institution region			
New England	4.32	5.88	5.64
Mid East	16.73	13.12	13.68
Great Lakes	17.83	15.56	15.91
Plains	8.02	7.87	7.90
Southeast	26.79	23.34	23.87
Southwest	8.95	11.73	11.30
Rocky Mountains	3.99	3.30	3.41
Far West	12.99	16.48	15.94
Outlying Area	0.38	1.84 *	1.62
Type of institution and enrollment category			
Public			
Fewer than 1,000	1.44	2.07	1.97
1,000–2,499	4.32	3.50	3.63
2,500–4,999	7.02	7.99	7.85
5,000–9,999	17.17	12.34	13.08
10,000–19,999	22.68	23.95	23.75
20,000 or more	23.64	24.11	24.04

See footnotes at end of table.

Table 6.10.—Comparison of BPS:1996/2001 converted refusals and other respondents— Continued

Demographic characteristics	Converted refusals Percent estimate[1]	Other respondents Percent estimate[1]	Total respondents Percent estimate[1]
Private not-for-profit			
Fewer than 1,000	1.16	1.94	1.82
1,000–2,499	4.17	5.62	5.39
2,500–4,999	2.27	2.91	2.81
5,000–9,999	2.3	2.15	2.17
10,000 or more	3.93	3.67	3.71
Private for-profit			
Fewer than 300	2.82	3.50	3.40
300–999	4.96	4.46	4.53
1,000 or more	2.13	1.80	1.85
Applied for aid			
Yes	63.08	70.75 *	69.57
No	36.92	28.36 *	29.68
Receipt of federal aid			
Yes	35.85	46.10 *	44.52
No	64.15	53.02 *	54.73
Receipt of Pell Grant			
Yes	17.46	29.07 *	27.28
No	82.54	70.05 *	71.97
Receipt of Stafford Loan			
Yes	25.83	29.33	28.79
No	74.17	69.79	70.46
Receipt of state aid			
Yes	9.96	16.16 *	15.20
No	90.04	82.96 *	84.05
Receipt of institution aid			
Yes	14.43	23.75 *	22.32
No	85.57	75.36 *	76.93
Receipt of any aid			
Yes	48.73	61.64 *	59.66
No	51.27	37.47 *	39.59
Prior respondent			
Yes	73.81	78.60	77.87
No	26.19	21.40	22.13
Parents' income (for dependent students)			
$29,999 or less	21.25	22.30	22.13
$30,000–$59,999	32.23	25.80	26.79
$60,000–$99,999	14.43	16.69	16.34
$100,000 or more	10.02	7.24	7.66
Student's income (for independent students)			
$5,000 or less	3.85	5.68	5.40
$5,000–$9,999	3.17	4.30	4.13
$10,000–$19,999	5.89	6.27	6.21
$20,000 or more	9.15	10.84	10.58

*Difference between respondents and nonrespondents is significant at the $p<0.05/(c-1)$ level, where the (c-1) divisor corrects for the (c-1) multiple comparisons among the c levels of the variable.

[1] The weight used is B01_100.

NOTE: Characteristics are from base year data (NPSAS:96) with the exception of the prior respondent category. Missing values are not shown in the table, and therefore some percentages may not sum to 100% for a variable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.11.—Comparison of BPS:1996/2001 late respondents and early respondents

Demographic characteristics	Late respondents Percent estimate[1]	Early respondents Percent estimate[1]	Total respondents Percent estimate[1]
Age			
19 or younger	69.33	68.87	68.95
20 to 23	12.83	11.26	11.54
24 to 29	11.27	7.55	8.21
30 to 39	5.94	5.27	5.39
40 or older	0.63	7.05 *	5.90
Typical age for level			
No, not typical age for level	30.67	30.62	30.63
Yes, typical age for level	69.33	68.47	68.62
Race/ethnicity			
White, non-Hispanic	64.26	73.83	72.12
Black, non-Hispanic	15.69	10.69	11.58
Hispanic	11.87	9.54	9.95
Asian/Pacific Islander	6.99	4.60	5.03
American Indian/Alaska Native	0.66	0.72	0.71
Other	0.53	0.61	0.60
Gender			
Male	51.67	43.03	44.57
Female	48.33	56.97	55.43
Citizenship			
U.S. citizen	95.76	93.28	93.72
Non-citizen, eligible	3.99	5.12	4.92
Non-citizen, ineligible	0.25	0.69	0.61
Attendance status			
Full-time/full year 1 institution	43.77	52.31	50.79
Full-time/full year more than 1 institution	1.38	1.81	1.73
Full-time/part year	14.22	14.71	14.62
Part-time/full year 1 institution	18.20	13.62	14.44
Part-time/full year more than 1 institution	1.05	0.42	0.53
Part-time/part year	20.39	14.63	15.66
Institution level			
4-year	27.80	45.06 *	41.99
2-year	64.48	47.05 *	50.15
Less-than-2-year	7.73	7.89	7.86
Institution control			
Public	76.83	73.77	74.32
Private not-for-profit	11.33	16.88 *	15.89
Private for-profit	11.84	9.34	9.79
Institution region			
New England	6.28	5.50	5.64
Mid East	12.40	13.95	13.68
Great Lakes	15.48	16.00	15.91
Plains	5.26	8.47	7.90
Southeast	21.31	24.43	23.87
Southwest	10.71	11.43	11.30
Rocky Mountains	3.74	3.34	3.41
Far West	23.60	14.28	15.94
Outlying Area	1.23	1.70	1.62
Type of institution and enrollment category			
Public			
Fewer than 1,000	0.79	2.23	1.97
1,000–2,499	4.04	3.54	3.63
2,500–4,999	3.98	8.68 *	7.85
5,000–9,999	12.42	13.22	13.08
10,000–19,999	28.62	22.70	23.75
20,000 or more	26.98	23.40	24.04

See footnotes at end of table.

Table 6.11.—Comparison of BPS:1996/2001 late respondents and early respondents— Continued

Demographic characteristics	Late respondents Percent estimate[1]	Early respondents Percent estimate[1]	Total respondents Percent estimate[1]
Private not-for-profit			
Fewer than 1,000	1.31	1.93	1.82
1,000–2,499	4.32	5.63	5.39
2,500–4,999	2.23	2.93	2.81
5,000–9,999	1.30	2.36	2.17
10,000 or more	2.18	4.04 *	3.71
Private for-profit			
Fewer than 300	2.72	3.54	3.40
300–999	6.53	4.10	4.53
1,000 or more	2.59	1.69	1.85
Applied for aid			
Yes	62.19	71.17 *	69.57
No	37.81	27.92 *	29.68
Receipt of federal aid			
Yes	41.01	45.28	44.52
No	58.99	53.81	54.73
Receipt of Pell Grant			
Yes	27.34	27.27	27.28
No	72.66	71.82	71.97
Receipt of Stafford Loan			
Yes	26.82	29.22	28.79
No	73.18	69.87	70.46
Receipt of state aid			
Yes	10.10	16.31 *	15.20
No	89.9	82.78 *	84.05
Receipt of institution aid			
Yes	16.97	23.48 *	22.32
No	83.03	75.61 *	76.93
Receipt of any aid			
Yes	52.44	61.22 *	59.66
No	47.56	37.87 *	39.59
Prior respondent			
Yes	59.37	81.87 *	77.87
No	40.63	18.13 *	22.13
Parents' income (for dependent students)			
$29,999 or less	24.87	21.54	22.13
$30,000–$59,999	28.88	26.33	26.79
$60,000–$99,999	14.74	16.69	16.34
$100,000 or more	6.67	7.88	7.66
Student's income (for independent students)			
$5,000 or less	8.43	4.74	5.40
$5,000–$9,999	4.58	4.03	4.13
$10,000–$19,999	4.5	6.59	6.21
$20,000 or more	7.34	11.28	10.58

*Difference between respondents and nonrespondents is significant at the $p<0.05/(c-1)$ level, where the (c-1) divisor corrects for the (c-1) multiple comparisons among the c levels of the variable.

[1] The weight used is B01_100.

NOTE: Characteristics are from base year data (NPSAS:96) with the exception of the prior respondent category. Missing values are not shown in the table, and therefore some percentages may not sum to 100% for a variable.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

The estimated mean based on nonrespondents, $\overline{y}_{NR}$, can be computed using data for the particular variable for which the data for most of the nonrespondents were available. π can be estimated as follows:

$$\hat{\pi} = (1-\eta)\,\overline{y}_R + \eta\overline{y}_{NR}$$

where η is the weighted unit nonresponse rate. Therefore, the bias can be estimated as follows:

$$\hat{B}(\overline{y}_R) = \overline{y}_R - \hat{\pi}$$

or equivalently

$$\hat{B}(\overline{y}_R) = \eta(\overline{y}_R - \overline{y}_{NR}).$$

This formula shows that the estimate of the nonresponse bias is the difference between the mean for respondents and nonrespondents multiplied by the weighted nonresponse rate.

The variance of the bias was then computed as follows:

$$\mathrm{var}(\hat{B}) = \eta^2\,\mathrm{var}(\overline{y}_R - \overline{y}_{NR})$$

where $\overline{y}_R$ and $\overline{y}_{NR}$ are the estimates using the original weights and $\mathrm{var}(\overline{y}_R - \overline{y}_{NR})$ was estimated using Taylor series linearization (taking into account the covariance between $\overline{y}_R$ and $\overline{y}_{NR}$). A t-test was used to determine which variables had significant nonresponse bias at the 5 percent level.

The first set of columns in **table 6.12** shows the estimated bias, before weighting adjustments, for variables available for most responding and nonresponding students. The bias of several variables, such as typical age for level, attendance status, institution level and control, receipt of state aid, parents' high school education status, and prior response status, is significant, although the bias is small for some of these variables.

Weight adjustments are typically used to reduce bias due to unit nonresponse, and the results in **tables 6.9–6.12** show that these adjustments are definitely important for reducing the potential for nonresponse bias due to the differences between respondents and nonrespondents. The initial nonresponse models incorporated the survey stratification variables, variables identified during the CHAID analysis, and other variables that were thought to be predictive of nonresponse (which included the variables identified in **tables 6.9–6.12**) in the nonresponse models. The three steps of nonresponse adjustment

- inability to locate the student,
- refusal to be interviewed, and
- other noninterview,

were used to adjust for the potential bias resulting from the three different types of nonresponse. All nonresponse models were fit using RTI's proprietary generalized exponential models (GEMs)[9], which are similar to logistic models using bounds for adjustment factors. Section A gives the weighting details.

---

[9] Folsom, R.E. and Singh, A.C. (2000). "The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification." *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp. 598–603.

Table 6.12—Nonresponse bias before and after weight adjustment for selected variables

	Before nonresponse adjustment			After weight adjustment	
	Respondent percentage, original weights	Nonrespondent percentage, original weights	Estimated bias	Percentage, final adjusted weights	Estimated bias
Age					
19 or younger	68.95	57.37	1.9020 *	67.05	0.0000
20 to 23	11.54	21.78	-1.6816	13.22	0.0000
24 to 29	8.21	6.53	0.2772	7.94	0.0000
30 to 39	5.39	10.24	-0.7973	6.19	0.0000
40 or older	5.90	4.08	0.2998	5.60	0.0000
Typical age for level					
No, not typical age for level	30.63	42.29	-1.9329 *	32.83	-0.2819
Yes, typical age for level	68.62	56.90	1.9329 *	66.95	-0.2558
Race/ethnicity					
White, non-Hispanic	72.12	69.47	0.4366	71.69	0.0000
Black, non-Hispanic	11.58	11.94	-0.0592	11.98	-0.3422
Hispanic	9.95	10.66	-0.1159	10.19	-0.1250
Asian/Pacific Islander	5.03	6.35	-0.2173	4.73	0.5183
American Indian/Alaska Native	0.71	0.99	-0.0457	0.78	-0.0211
Other	0.60	0.59	0.0015	0.63	-0.0301
Gender					
Male	44.57	48.21	-0.5974	45.17	0.0000
Female	55.43	51.79	0.5974	54.83	0.0000
Citizenship					
U.S. citizen	93.72	91.79	0.3105	93.94	-0.5377
Non-citizen, eligible	4.92	4.79	0.0210	5.27	-0.3717
Non-citizen, ineligible	0.61	2.61	-0.3314	0.57	0.3717
Attendance status					
Full-time/full year 1 inst.	50.79	42.48	1.4851	50.04	-0.6194
Full-time/full year >1 inst.	1.73	0.74	0.1694 *	1.57	0.0014
Full-time/part year	14.62	28.20	-2.2515 *	16.75	0.0980
Part-time/full year 1 inst.	14.44	12.31	0.3828	14.07	0.0141
Part-time/full year >1 inst.	0.53	1.95	-0.2382	0.76	0.0000
Part-time/part year	15.66	13.13	0.4524	15.66	-0.4179
Institution level					
4-year	41.99	37.27	0.7752	41.21	0.0000
2-year	50.15	48.28	0.3069	49.84	0.0000
Less-than-2-year	7.86	14.45	-1.0821 *	8.94	0.0000
Institution control					
Public	74.32	67.10	1.1863	73.13	0.0000
Private not-for-profit	15.89	15.54	0.0573	15.84	0.0000
Private for-profit	9.79	17.36	-1.2436 *	11.03	0.0000
Institution region					
New England	5.64	7.50	-0.3089	5.94	0.0000
Mid East	13.68	11.98	0.2795	13.40	0.0000
Great Lakes	15.91	13.06	0.4698	15.61	-0.1726
Plains	7.90	9.67	-0.2941	8.26	-0.0769
Southeast	23.87	24.52	-0.1092	24.22	-0.2348
Southwest	11.30	11.83	-0.0883	11.44	-0.0534
Rocky Mountains	3.41	4.20	-0.1312	3.54	0.0000
Far West	15.94	15.51	0.0692	15.87	0.0000
Outlying Area	1.62	0.93	0.1131	1.50	0.0000
Type of institution and enrollment category					
Public					
Fewer than 1,000	1.97	1.08	0.1469	1.83	0.0000
1,000–2,499	3.63	7.49	-0.6342	4.26	0.0000
2,500–4,999	7.85	4.57	0.5381	7.31	0.0000
5,000–9,999	13.08	23.44	-1.7020	14.78	0.0000
10,000–19,999	23.75	16.23	1.2352	22.52	0.0000
20,000 or more	24.04	14.29	1.6024 *	22.44	0.0000

See footnotes at end of table.

Table 6.12.—Nonresponse bias before and after weight adjustment for selected variables—Continued

Demographic characteristics	Before nonresponse adjustment			After weight adjustment	
	Respondent percentage, original weights	Nonrespondent percentage, original weights	Estimated bias	Percentage, final adjusted weights	Estimated bias
Private not-for-profit					
Fewer than 1,000	1.82	2.20	-0.0618	1.88	0.0000
1,000–2,499	5.39	4.51	0.1454	5.25	0.0000
2,500–4,999	2.81	3.13	-0.0523	2.86	0.0000
5,000–9,999	2.17	2.49	-0.0523	2.22	0.0000
10,000 or more	3.71	3.23	0.0786	3.63	0.0000
Private for-profit					
Fewer than 300	3.40	6.47	-0.5057	3.90	0.0000
300–999	4.53	6.03	-0.2457	4.78	0.0000
1,000 or more	1.85	4.85	-0.4926	2.34	0.0000
Applied for aid					
Yes	69.57	72.01	-0.4100	70.33	-0.3588
No	29.68	27.19	0.4100	29.45	-0.1789
Receipt of federal aid					
Yes	44.52	47.79	-0.5454	45.36	-0.3054
No	54.73	51.40	0.5454	54.42	-0.2323
Receipt of Pell Grant					
Yes	27.28	32.29	-0.8321	28.36	-0.2495
No	71.97	66.90	0.8321	71.42	-0.2882
Receipt of Stafford Loan					
Yes	28.79	31.67	-0.4801	29.57	-0.3054
No	70.46	67.52	0.4801	70.21	-0.2323
Receipt of state aid					
Yes	15.20	11.29	0.6454 *	14.73	-0.1726
No	84.05	87.90	-0.6454 *	85.05	-0.3651
Receipt of institution aid					
Yes	22.32	18.43	0.6424	21.68	0.0000
No	76.93	80.77	-0.6424	78.10	-0.5377
Receipt of any aid					
Yes	59.66	59.48	0.0237	59.93	-0.3054
No	39.59	39.71	-0.0237	39.84	-0.2323
Prior respondent					
Yes	77.87	56.90	3.4449 *	93.64	-19.2170 *
No	22.13	43.10	-3.4449 *	6.36	19.2175 *
Parents' income (for dependent students)					
$29,999 or less	22.13	23.90	-0.2936	22.73	-0.3092
$30,000–$59,999	26.79	19.59	1.1889	25.66	-0.0559
$60,000–$99,999	16.34	13.75	0.4273	15.92	0.0000
$100,000 or more	7.66	8.51	-0.1408	7.80	0.0000
Student's income (for independent students)					
$5,000 or less	5.40	6.52	-0.1858	5.58	0.0000
$5,000–$9,999	4.13	5.76	-0.2699	4.40	0.0000
$10,000–$19,999	6.21	14.57	-1.3837	7.59	0.0000
$20,000 or more	10.58	6.60	0.6576	10.10	-0.1726

* Bias is significant at the $p < 0.05/(c-1)$ level, where c is the number of categories within the primary variable. The (c-1) divisor was used to correct for the (c-1) multiple comparisons among the c levels of the variable.

NOTE: Characteristics are from base year data (NPSAS:96) with the exception of the prior respondent category. Missing values are not shown in the table, and therefore some percentages may not sum to 100% for a variable.

Original weight is B01_100. Final adjusted weight is B01AWT.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

The second set of columns in **table 6.12** shows the estimated bias after weighting for the variables available for most responding and nonresponding students. Some variables have zero bias after weighting. The bias is not significantly different from zero for the remaining variables, except for whether the student was a prior respondent. This variable was not included in the located model because of convergence problems. It was included in the model for refusal to be interviewed (as a part of the interaction segments identified by CHAID), and in the other noninterview model.

b.   Mean Response by Date of Response

The bias of a survey estimate is the difference between the estimate and the true value of the corresponding population parameter. The bias is necessarily unknown for most estimates because the true value of the population parameter is unknown. If it were known, the difference between the values of the survey statistic and the population parameter could be used to construct a confidence interval estimate of the bias. If the confidence interval included zero (0), one could conclude that the estimate appeared to be unbiased. Since the true values of the population parameters usually are not known, an alternative approach can be used to investigate the potential for bias in the BPS:1996/2001 survey estimates.

Although there are many other potential sources of bias, one of the most important sources of bias in sample surveys is survey nonresponse. Survey nonresponse results in bias when the unobserved outcomes for the nonrespondents are systematically different from the observed outcomes for the respondents.

Hence, the potential for nonresponse bias can be modeled by using the pattern of mean response by date of response. The survey respondents were subdivided into 10 groups based on date of interview. Then, within each institution level (less-than-2-year, 2-year, and 4-year), all respondents were again subdivided into 10 groups of approximately equal numbers of respondents based on date of interview. This strategy was adopted so that the mean response in each group would have approximately the same precision. Some of the resultant respondent groups had shorter ranges of dates at the beginning of data collection because relatively larger numbers of interviews were completed during the first few months of data collection.

The pattern of cumulative mean response (using unweighted means, or averages) by date of interview (both overall and within level of institution) was examined for the following:

- mean age in the base year (1995-96),
- percent non-White,
- percent enrolled in an undergraduate program in the spring of 2001,
- percent who had attained a degree by June 2001, and
- percent employed.

In addition, the mean of the institution level attended in the base year was examined for all students combined, where level was coded as follows: (1) less-than-2-year institution, (2) 2-year institution, and (3) 4-year institution.

If the mean responses from the later groups of respondents are reasonably consistent, then obtaining additional responses probably will have little effect on survey estimates and nonresponse bias probably is negligible. In this case, the plot of the cumulative mean response

will approach an asymptote toward the end of data collection. If the cumulative mean is either rising or falling sharply at the end of data collection, it suggests that the later respondents tended to have a mean response that was either higher or lower, respectively, than the overall mean. In this case, there is some evidence of potential for nonresponse bias.

The plots of cumulative mean by date of last interview are presented in **figures 6.1** through **6.6** for all students combined; **figures 6.7** through **6.11** for students who were enrolled in 4-year institutions in the base year (1995–96); **figures 6.12** through **6.16** for 2-year institutions; and **figures 6.17** through **6.21** for less-than-2-year institutions.[10] **Figure 6.1** shows some potential for bias by institutional level for overall population estimates because it appears that additional respondents would be more likely to have attended less-than-4-year institutions. Other evidence of potential bias was that for the sample as a whole, and for each of the three institution level samples, additional respondents were more likely to be non-White (see **figures 6.3**, **6.8**, **6.13**, and **6.18)**. For the sample as a whole, and for the 4-year institution and 2-year institution samples, additional respondents were less likely to have attained a degree by spring 2001 (see **figures 6.5**, **6.10**, and **6.15**).

---

[10] The date of last interview depends on the particular subpopulation of students included in each plot.

Figure 6.1.—Cumulative mean institutional level for all students: 2001
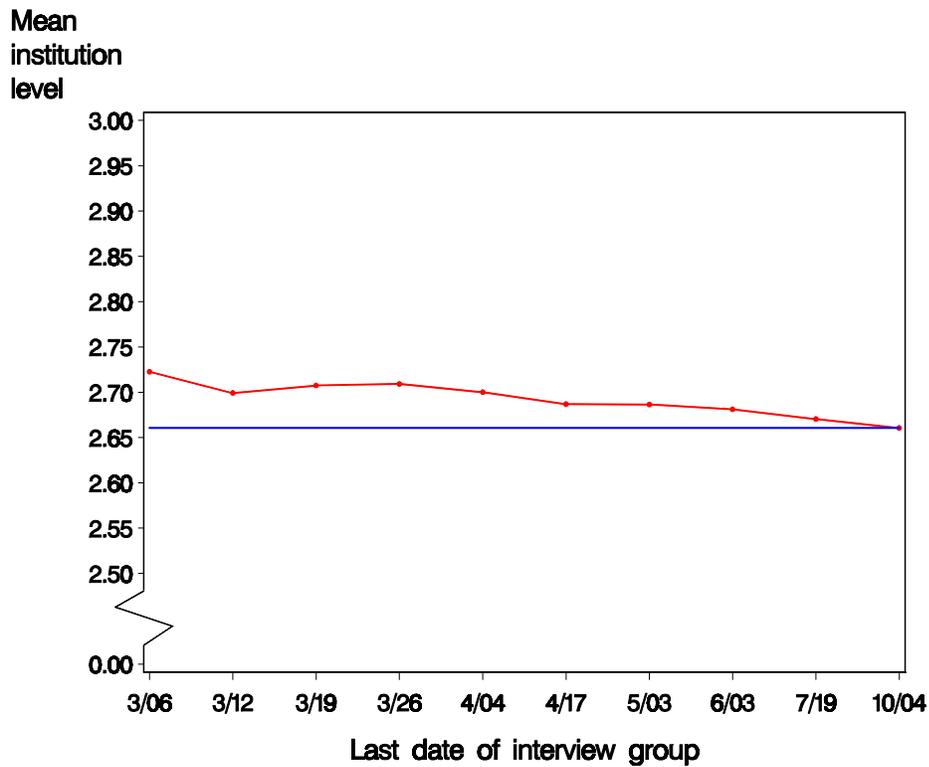
NOTE: The horizontal line gives the average institution level for all of the respondents.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

The cumulative mean institutional level appears to be decreasing for the last students interviewed. This result suggests some potential bias by level of institution for overall population estimates because additional respondents would be more likely to have attended 2-year or less-than-2-year institutions in the base year.

Figure 6.2.—Cumulative mean age in the base year of all students: 2001
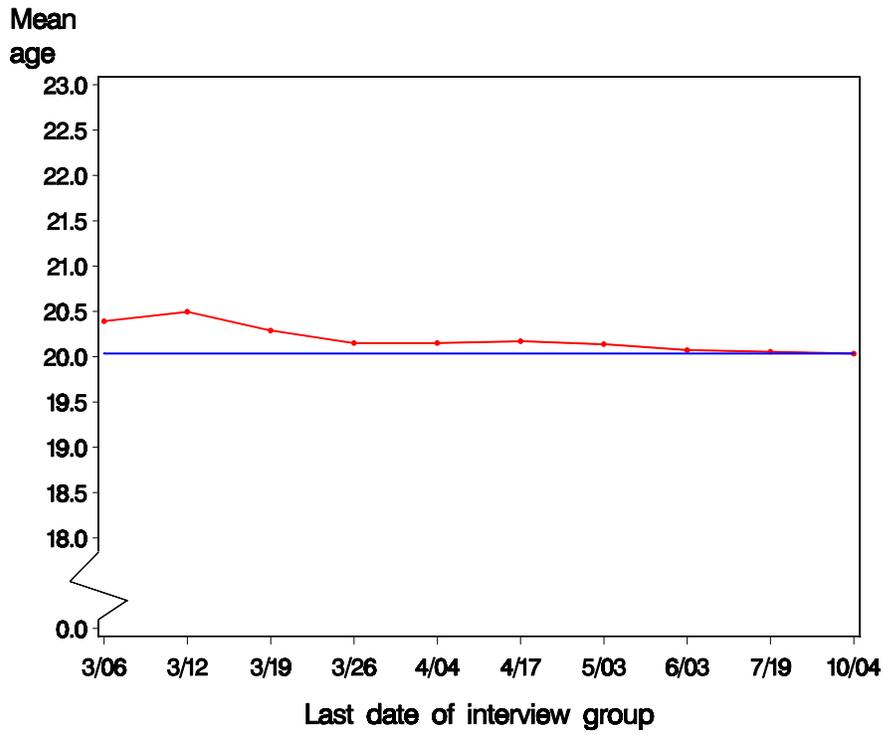
NOTE: The horizontal line gives the mean base-year age for all respondents.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

The cumulative mean base-year age of students is relatively stable throughout data collection (about 20 years of age). Hence, there is no evidence of potential for bias with respect to student age for overall population estimates.

Figure 6.3.—Cumulative percentage of all students who are non-White: 2001
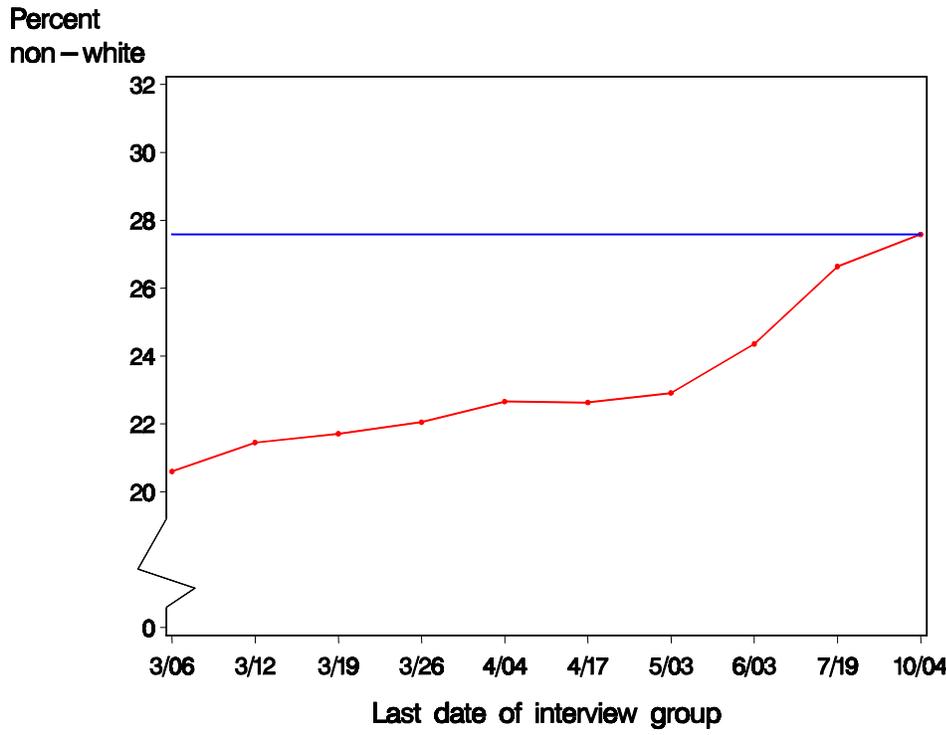
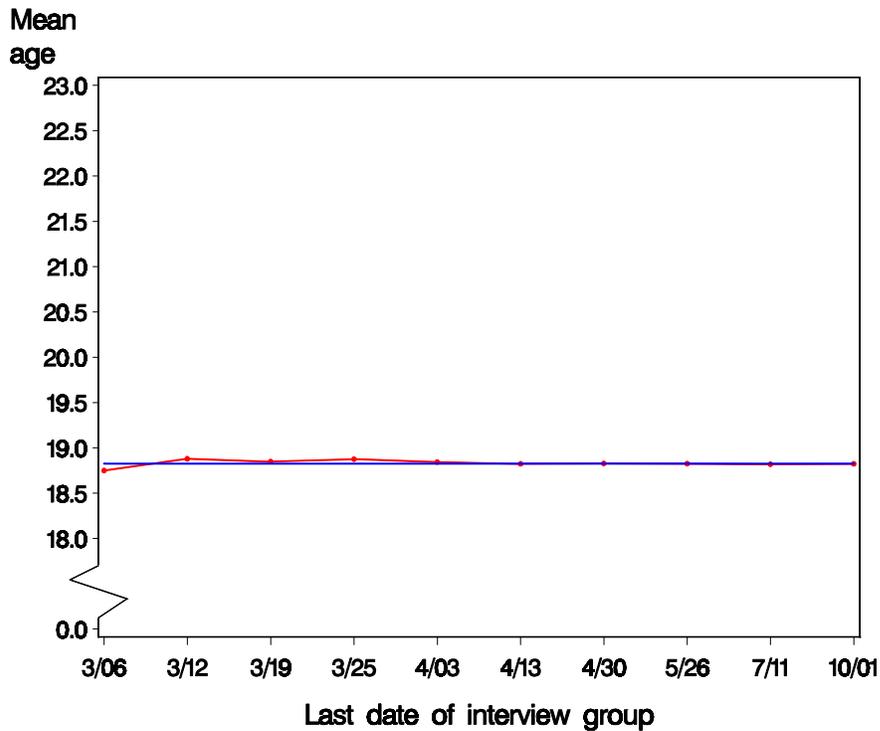NOTE: The horizontal line gives the percentage of all respondents who are non-White.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

The cumulative percentage of all students who are non-White increases throughout data collection. This suggests the potential for bias in overall population estimates with respect to the distribution of students by race because additional respondents would have been more likely to be non-White.

Figure 6.4.—Cumulative percentage of all students who are enrolled in an undergraduate program in spring 2001
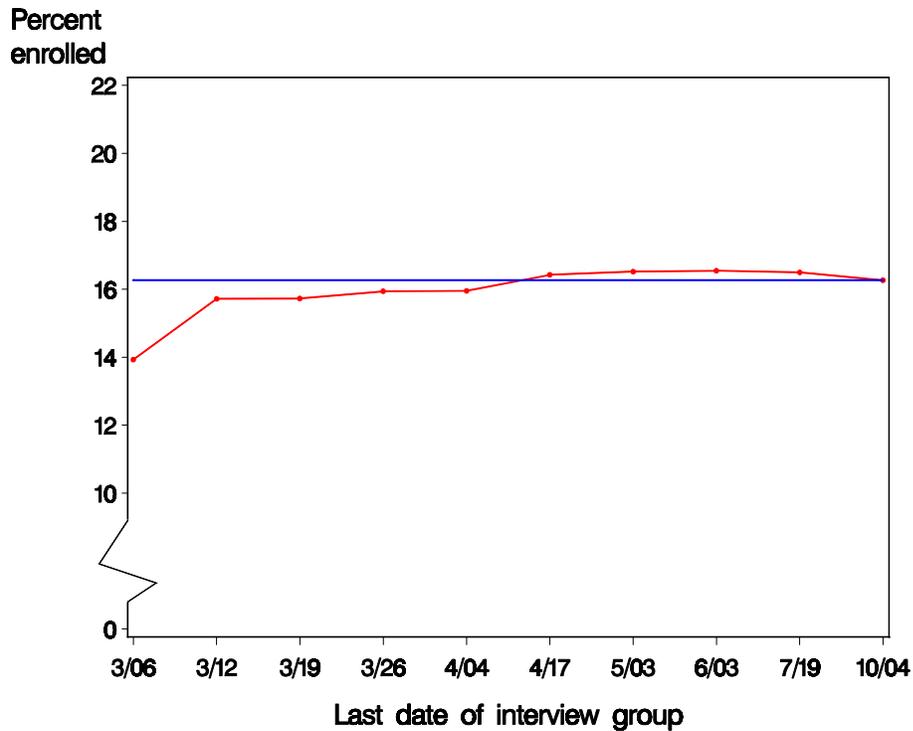
NOTE: The horizontal line gives the percentage of all respondents who were enrolled in an undergraduate program in spring 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

The cumulative percentage of students who were enrolled in an undergraduate program appears to be approaching an asymptote. Hence, there is no evidence of the potential for bias regarding the percentage of students currently enrolled.

Figure 6.5.—Cumulative percentage of all students who attained a degree by June 2001
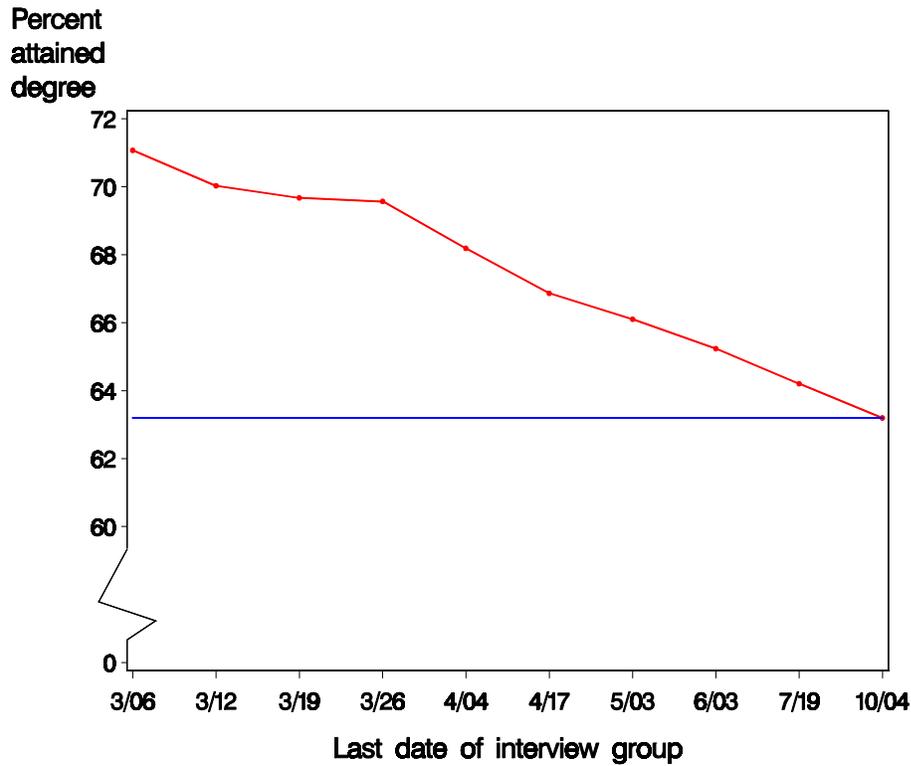
NOTE: The horizontal line gives the percentage of all respondents who had attained a degree by June 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

The cumulative percentage of students who had attained a degree by June 2001 decreases throughout data collection. This result suggests the potential for bias in the overall population estimates because additional respondents would be less likely to have attained a degree.

Figure 6.6.—Cumulative percentage of all students who were employed: 2001
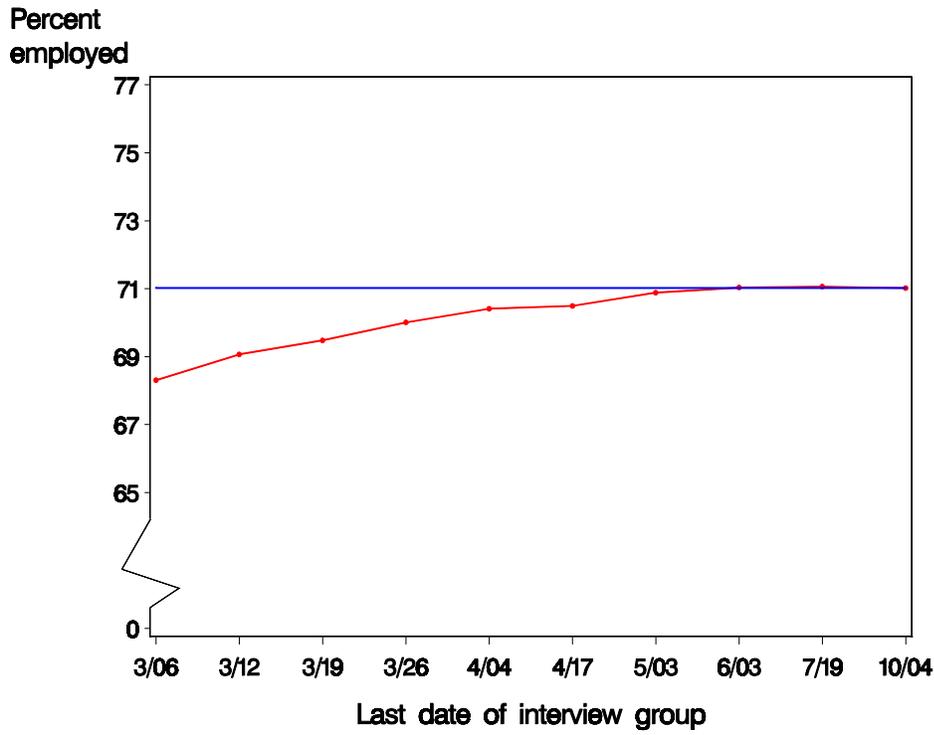
NOTE: The horizontal line gives the percentage of all respondents who were employed in 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

The cumulative percentage of all respondents who were employed approaches an asymptote. This suggests little potential for bias regarding employment status for overall population estimates.

Figure 6.7.—Cumulative mean age in the base year of students in 4-year institutions: 2001

NOTE: The horizontal line gives the mean base-year age for respondents in 4-year institutions.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 4-year institutions in the base year, the cumulative mean base-year age of students is relatively stable throughout data collection (about 19 years of age). Hence, there is no evidence of potential for bias with respect to student age for population estimates for 4-year institutions.

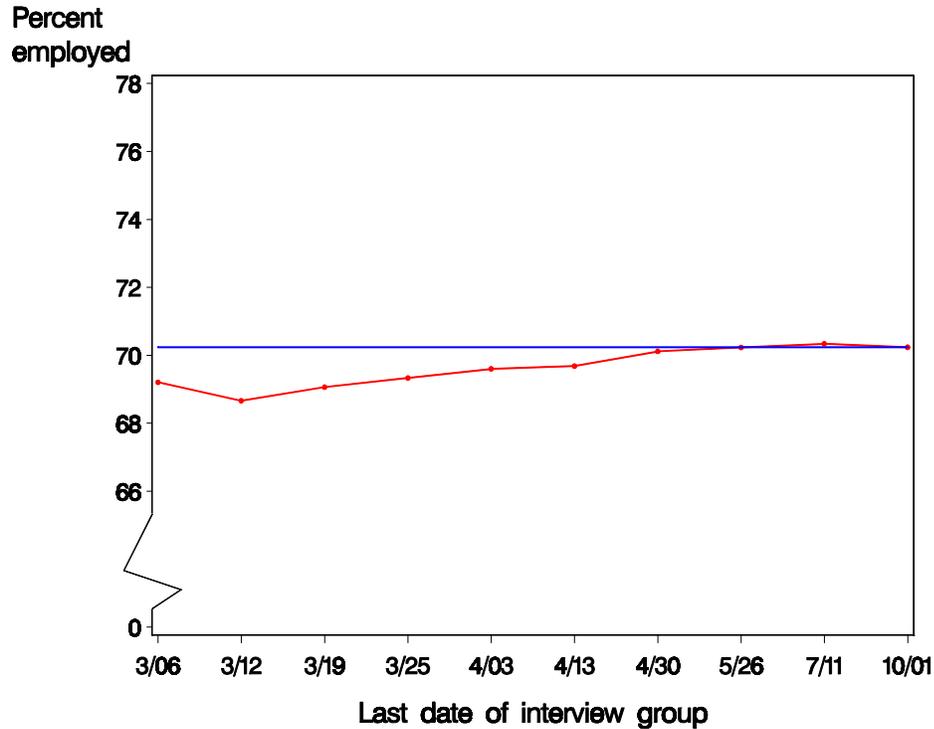**Figure 6.8.—Cumulative percentage of students in 4-year institutions who are non-White: 2001**
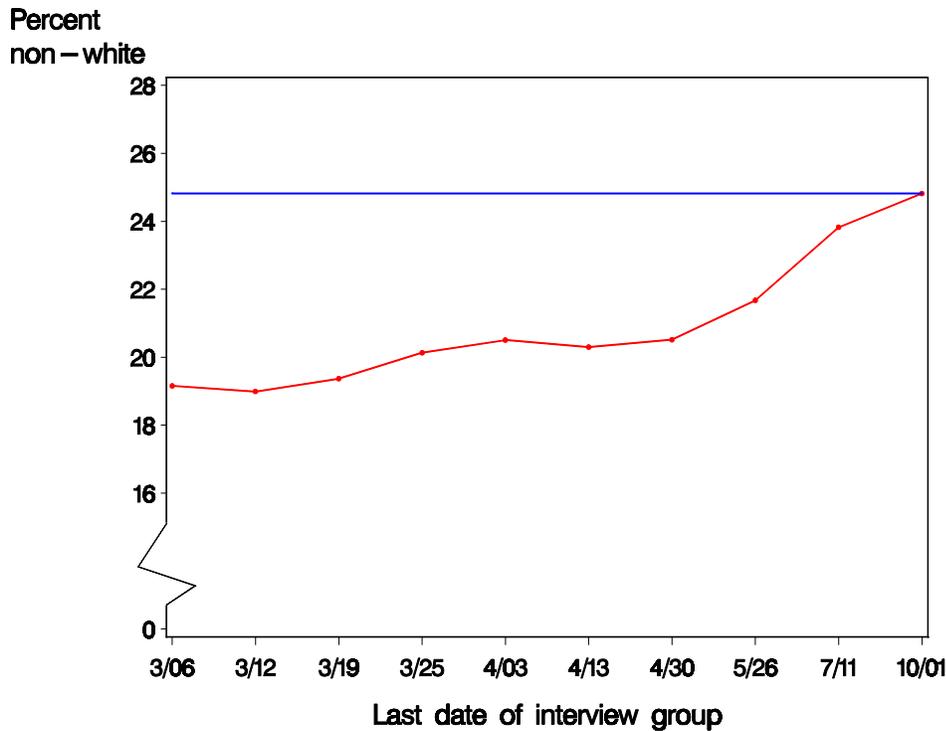
NOTE: The horizontal line gives the percentage of all respondents in 4-year institutions who are non-White.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled in 4-year institutions in the base year, the cumulative percentage of students who are non-White increases throughout data collection. This suggests the potential for bias with respect to the race distribution since additional respondents would be more likely to be non-White.

Figure 6.9.—Cumulative percentage of students in 4-year institutions who were enrolled in an undergraduate program in spring 2001

NOTE: The horizontal line gives the percentage of all respondents in 4-year institutions who were enrolled in an undergraduate program in spring 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 4-year institutions in the base year, the cumulative percentage who were enrolled in an undergraduate program in spring 2001 appears to converge to an asymptote. Hence there is no evidence of potential for bias regarding the distribution of current enrollment in the sample from 4-year institutions.

Figure 6.10.—Cumulative percentage of students in 4-year institutions who attained a degree by June 2001

NOTE: The horizontal line gives the percentage of all respondents in 4-year institutions who had attained a degree by June 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 4-year institutions in the base year, the cumulative percentage of students who had attained a degree by June 2001 decreases throughout data collection. This suggests some potential for bias by degree attainment in the sample from 4-year institutions because additional respondents would be less likely to have attained a degree.

Figure 6.11.—Cumulative percentage of students in 4-year institutions who were employed: 2001
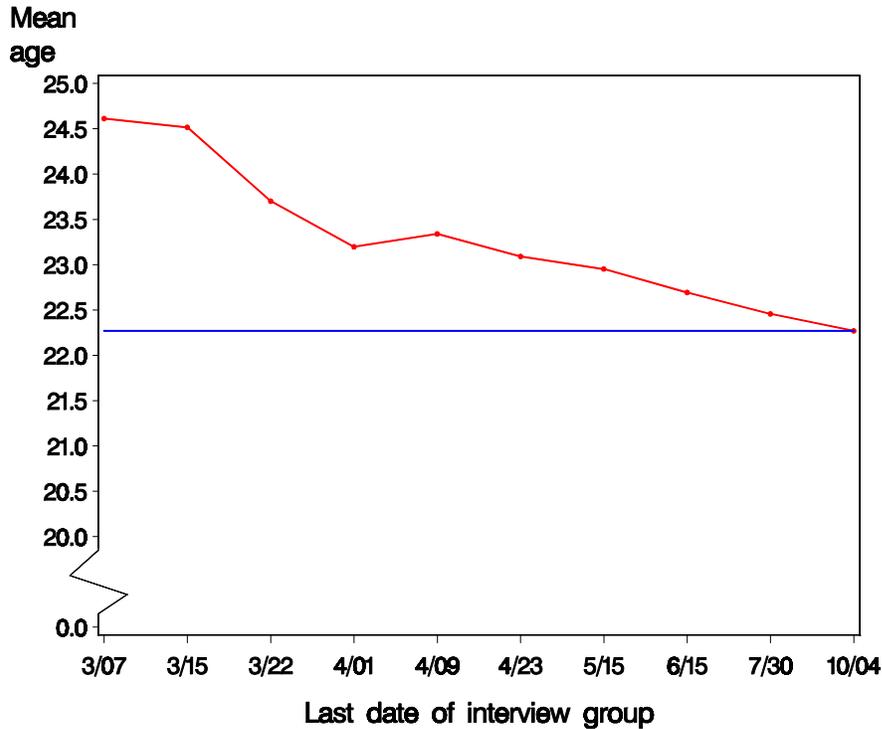

NOTE: The horizontal line gives the percentage of all respondents in 4-year institutions who were employed in 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 4-year institutions in the base year, the cumulative percentage who were employed generally increases throughout data collection but then appears to approach an asymptote. This suggests that there is little potential for bias regarding employment status in the sample from 4-year institutions.

Figure 6.12.—Cumulative mean age in the base year of students in 2-year institutions: 2001
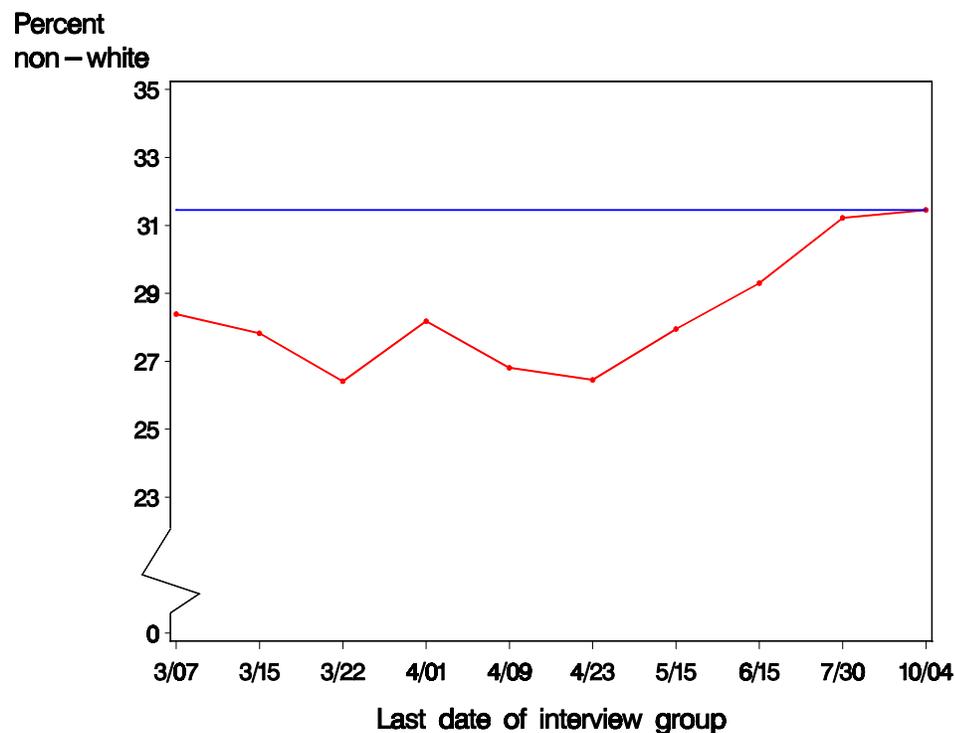

NOTE: The horizontal line gives the mean base-year age for respondents in 2-year institutions.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 2-year institutions in the base year, the cumulative mean base-year age of students decreases, but appears to converge to an asymptotic value of about 22 years of age. Hence, there is no evidence of potential for bias with respect to student age for population estimates for 2-year institutions.

Figure 6.13.—Cumulative percentage of students in 2-year institutions who are non-White: 2001
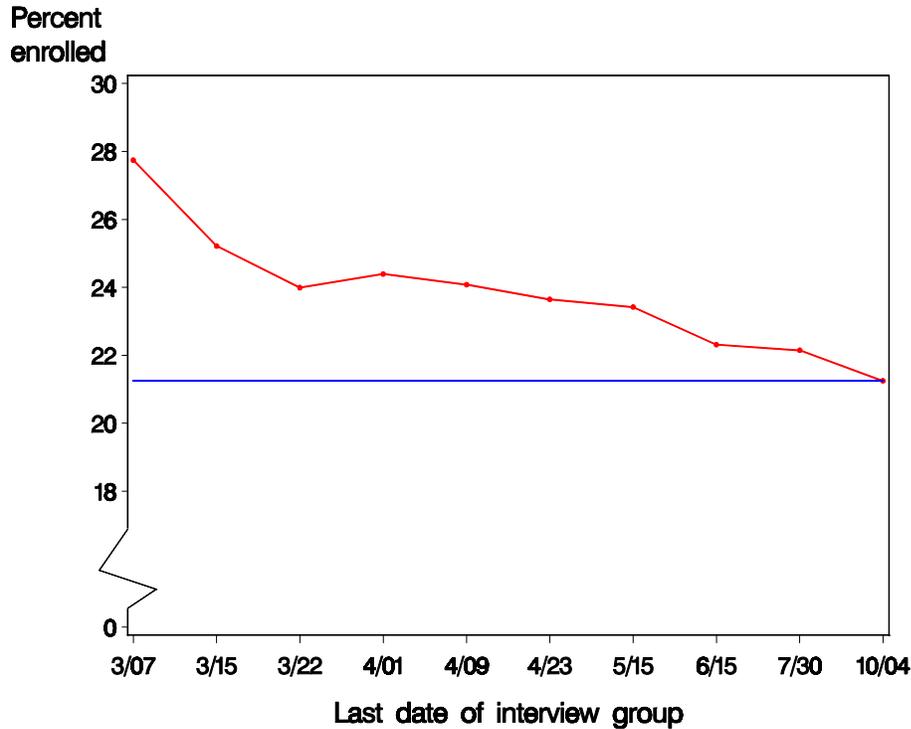

NOTE: The horizontal line gives the percentage of all respondents in 2-year institutions who are non-White.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 2-year institutions in the base year, the cumulative percentage of students who are non-White increases throughout data collection, but does appear to be converging to an asymptote for the last 10 percent of responding students. This suggests that there could be the potential for bias with respect to the race distribution for the sample from 2-year institutions because additional respondents may be more likely to be non-White.

Figure 6.14.—Cumulative percentage of students in 2-year institutions who are enrolled in an undergraduate program in spring 2001
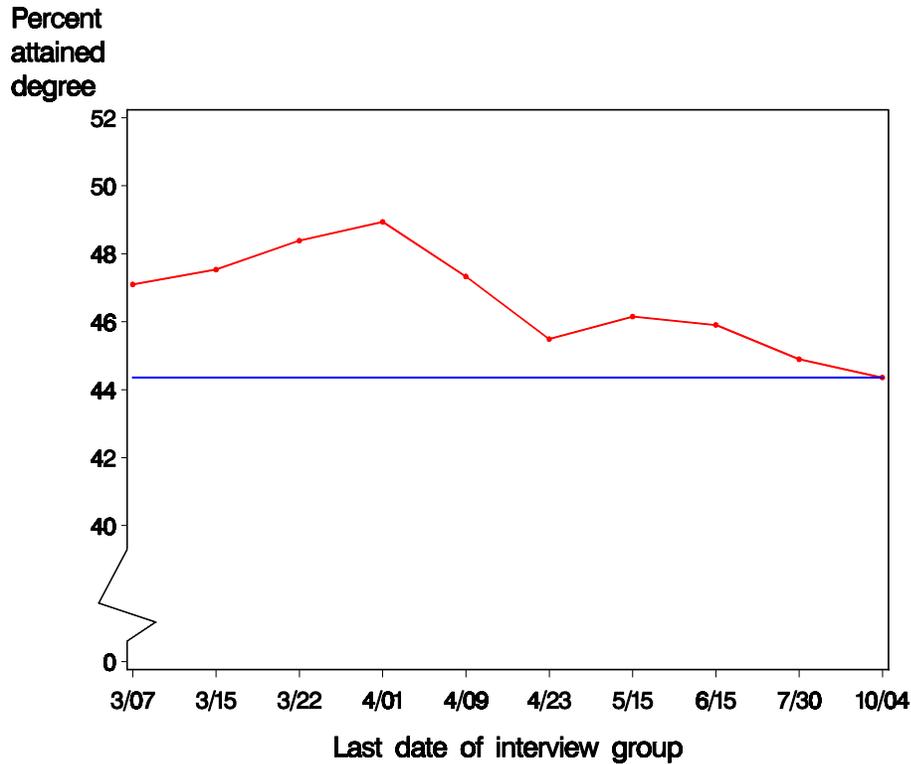

NOTE: The horizontal line gives the percentage of all respondents in 2-year institutions who were enrolled in an undergraduate program in spring 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 2-year institutions in the base year, the cumulative percentage who are enrolled in an undergraduate program decreases throughout data collection. This suggests that there is potential for bias with respect to enrollment status in the sample from 2-year institutions because additional respondents would be less likely to be enrolled.

Figure 6.15.—Cumulative percentage of students in 2-year institutions who attained a degree by June 2001
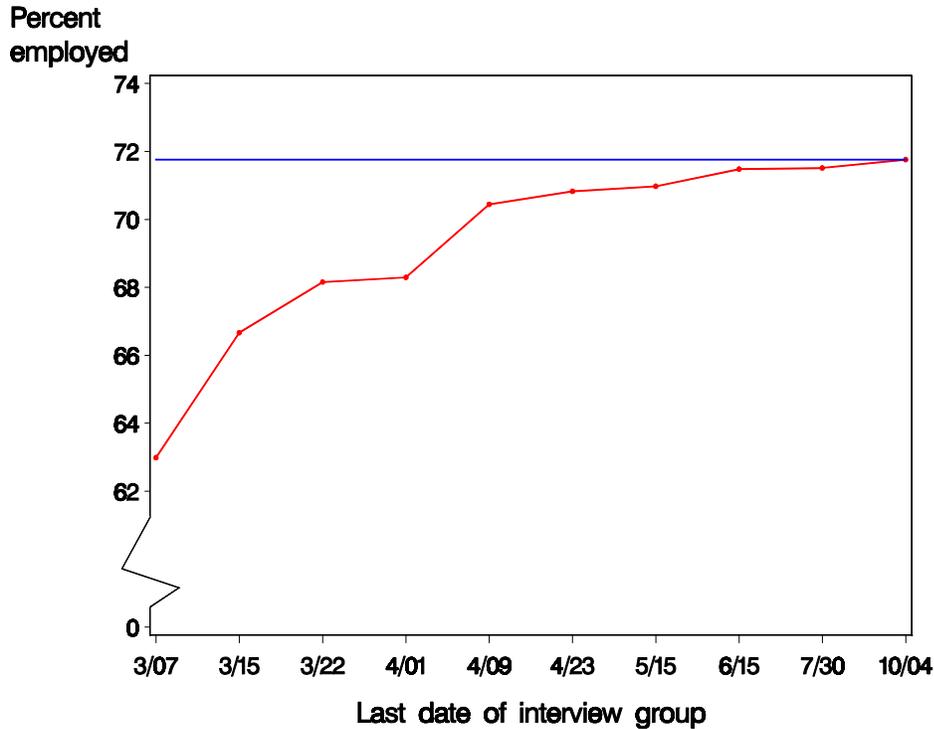

NOTE: The horizontal line gives the percentage of all respondents in 2-year institutions who had attained a degree by June 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 2-year institutions in the base year, the cumulative percentage who had attained a degree decreases throughout data collection. This suggests the potential for bias with respect to degree attainment since additional respondents would be less likely to have attained a degree.

Figure 6.16.—Cumulative percentage of students in 2-year institutions who were employed: 2001
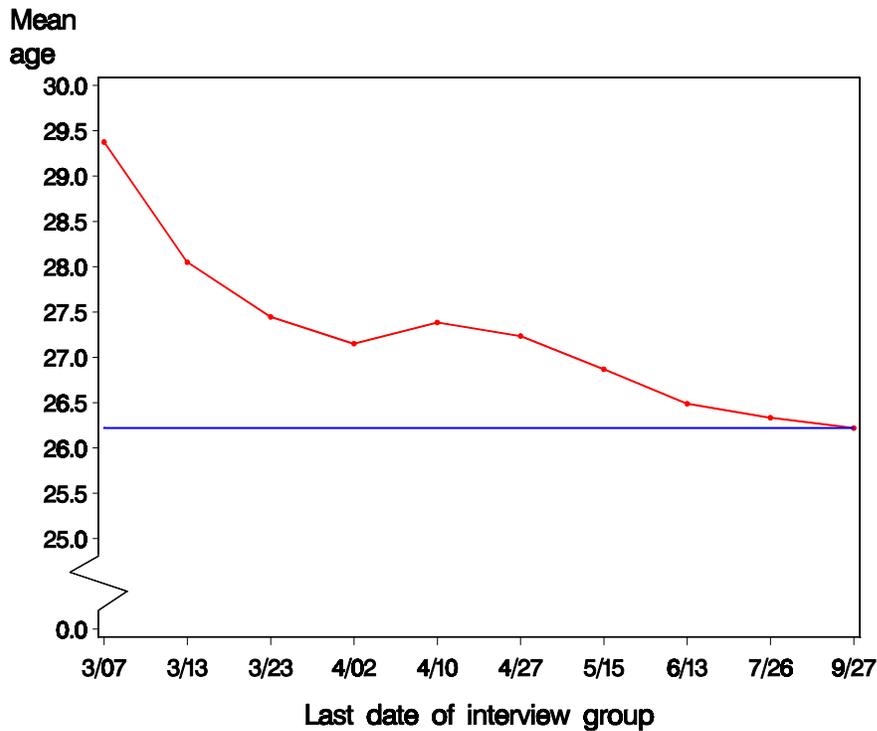

NOTE: The horizontal line gives the percentage of all respondents in 2-year institutions who were employed in 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at 2-year institutions in the base year, the cumulative percentage of students who are employed approaches an asymptotic value. This suggests little potential for bias with respect to employment status in the sample from 2-year institutions.

Figure 6.17.—Cumulative mean age in the base year of students in less-than-2-year institutions: 2001

NOTE: The horizontal line gives the mean base-year age for respondents in less-than-2-year institutions.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at less-than-2-year institutions in the base year, the cumulative mean base-year age of students decreases but appears to converge to an asymptotic value of about 26 years. Hence, there is little evidence of potential for bias with respect to student age for population estimates for less-than-2-year institutions.

Figure 6.18.—Cumulative percentage of students in less-than-2 year institutions who are non-White: 2001

NOTE: The horizontal line gives the percentage of all respondents in less-than-2-year institutions who are non-White.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at less-than-2-year institutions in the base year, the cumulative percentage of students who are non-White increases throughout data collection. This suggests that there is the potential for bias with respect to the race distribution in the sample from less-than-2-year institutions because additional respondents would be more likely to be non-White.

Figure 6.19.—Cumulative percentage of students in less-than-2-year institutions who were enrolled in an undergraduate program in spring 2001

NOTE: The horizontal line gives the percentage of all respondents in less-than-2-year institutions who were enrolled in an undergraduate program in spring 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at less-than-2-year institutions in the base year, the cumulative percentage who were enrolled in an undergraduate program appears to converge to an asymptotic value of about 9 percent. Hence, there is little evidence of potential for bias with respect to the undergraduate enrollment status for population estimates for less-than-2-year institutions.

Figure 6.20.—Cumulative percentage of students in less-than-2-year institutions who had attained a degree by June 2001
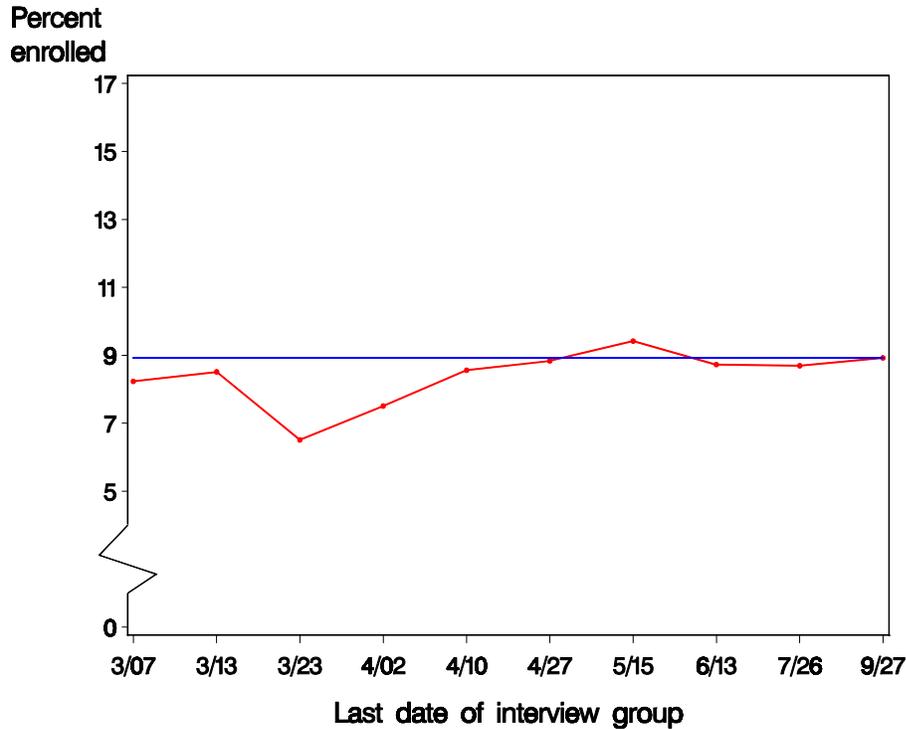

NOTE: The horizontal line gives the percentage of all respondents in less-than-2-year institutions who had attained a degree by June 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at less-than-2-year institutions in the base year, the cumulative percentage who had attained a degree by June 2001 decreases but then approaches an asymptote for the last 20 percent of the nonrespondents. Hence, there is little potential for bias with respect to degree attainment among the sample from less-than-2-year institutions.

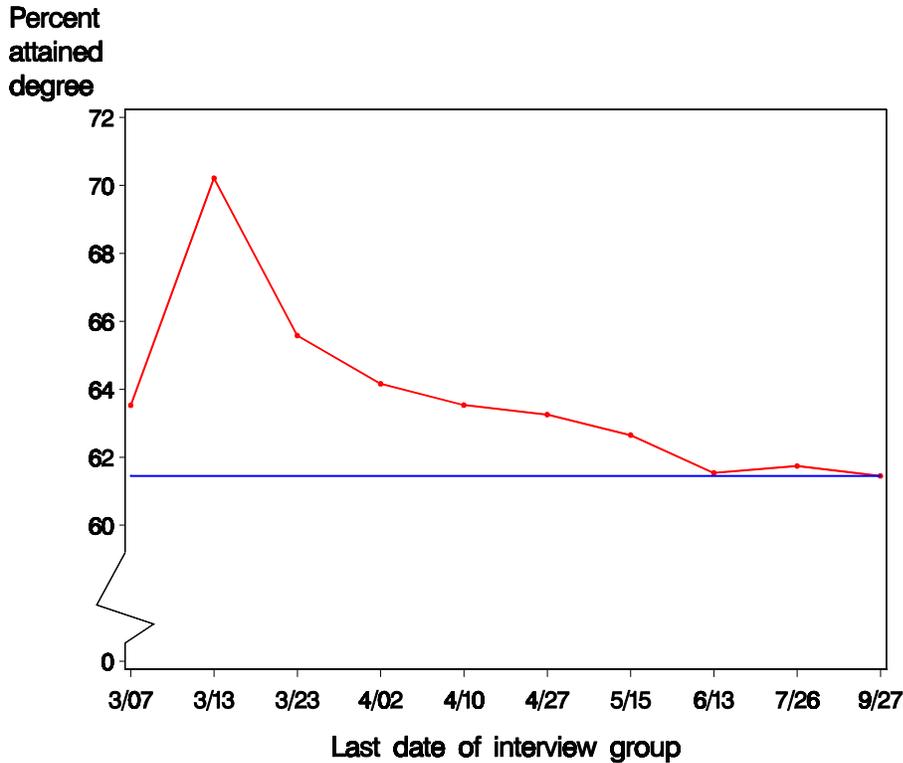**Figure 6.21.—Cumulative percentage of students in less-than-2-year institutions who were employed: 2001**

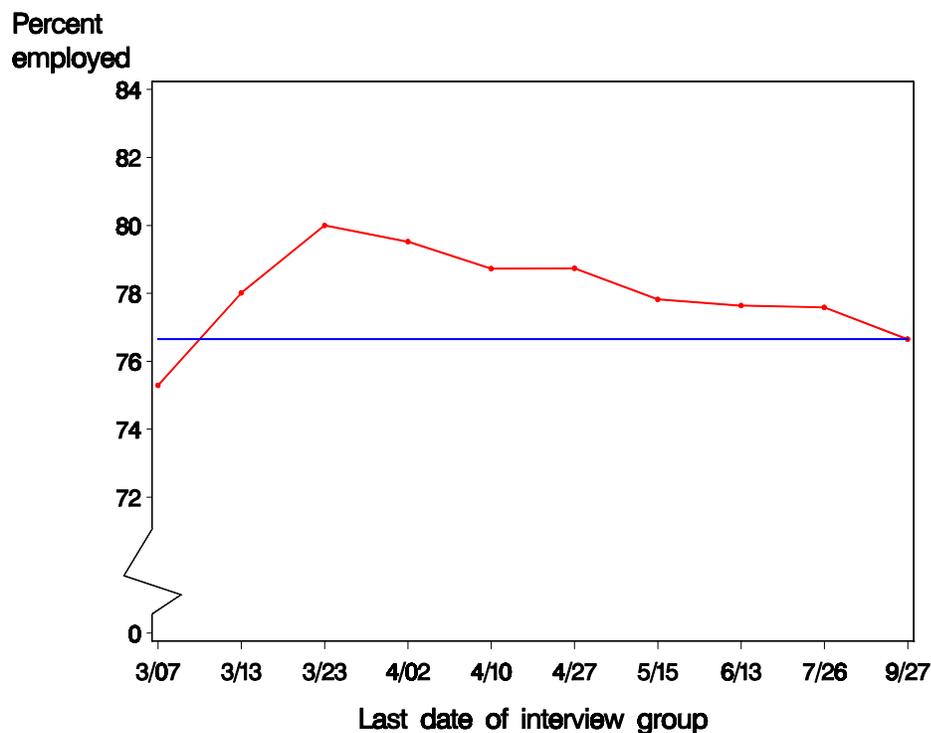NOTE: The horizontal line gives the percentage of all respondents in less-than-2-year institutions who were employed in 2001.

SOURCE: U. S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS: 1996/2001).

Among students enrolled at less-than-2-year institutions in the base year, the cumulative percentage of students who were employed increases sharply, and then decreases, but remains in the 77 percent range for the last 30 percent of the respondents. This suggests that may be some potential for bias with respect to employment status in the sample from less-than-2-year institutions because additional respondents would be less likely to be employed.

In summary, the graphical analyses shown in Figures 6.1—6.21 indicate that the potential for nonresponse bias exists among the following variables:

- institution level, with additional respondents more likely to have attended less-than-4-year institutions in the base year;

112

- race, with additional respondents more likely to be non-White;

- degree status, with additional respondents less likely to have attained a degree by spring 2001; and

- employment status in 2001 (for students who were enrolled in less-than two-year institutions in the base year), with later respondents less likely to be employed.

The analyses do not indicate a potential for nonresponse bias among the following variables:

- age;

- enrollment status in spring 2001; or

- employment status in 2001 for students who were enrolled in 4-year or 2-year institutions in the base year.

The analyses depicted in the figures were conducted using unweighted data to provide a qualitative indication of the potential for bias in the variables analyzed. These and other closely-related variables were included in the formal statistical tests of nonresponse bias in Tables 6.9 – 6.12, and also in the CHAID analyses performed to assist in choosing variables for the weight adjustment models. Table 6.12 indicates that the nonresponse weight adjustments were successful in reducing the nonresponse bias.

c.   ROC Curve

As described in section A, three nonresponse adjustment models were used for computing the final cross-sectional analysis weights for BPS:1996/2001. In order to assess the overall predictive ability of the combined models, a Receiver Operating Characteristics (ROC) curve, shown in **figure 6.22,** was used. A point on an ROC curve is constructed by considering a given predicted probability as a cutoff point for deciding whether a person is a respondent or a nonrespondent. For a given cutoff, a point on the ROC curve is obtained by plotting the proportion of respondents with a predicted probability greater than the cutoff (i.e., true positives) versus the proportion of nonrespondents with a predicted probability greater than the cutoff (i.e., false positives). The points on the ROC curve are then obtained by computing the proportion of true and false positives for the entire range of possible cutoffs.

The area under an ROC curve measures the probability that a randomly chosen pair of observations, one respondent and one nonrespondent, will be correctly ranked. The probability of a correct pairwise ranking is the same quantity that is estimated by the nonparametric Wilcoxon statistic. The null hypothesis associated with the Wilcoxon statistic is that the variable is not a useful discriminator between the populations. This corresponds to the null hypothesis that the predicted response probability of a respondent is just as likely to be smaller than the predicted response probability of a nonrespondent as it is to be greater. Thus, if the null hypothesis is true, the ROC curve will be a diagonal line that reflects the equally likely chance of making a correct or incorrect decision, and the area under the curve will be 0.5. If the null hypothesis is not true, the ROC curve will rise above the diagonal and the area under the curve will be significantly greater than 0.5.

All of the students in the BPS:1996/2001 sample were used in constructing this ROC curve. The student's predicted probability of response was calculated as the product of the predicted probabilities obtained from the three GEM models described in section A: P(located), P(nonrefusal for located students), and P(response for located students who did not refuse). These probabilities were the inverse of the adjustment factors from the GEM models, prior to trimming and smoothing. Since only located students were included in the nonrefusal model, and only nonrefusals were included in the final response model, the predicted probabilities were not directly available for students who were not located or for students who refused. The mean of the predicted probabilities was used for students who were in the models for the probabilities that were not directly available.

As shown in **figure 6.22**, the area under the ROC curve developed for the overall predicted response propensity was about 0.62, which corresponds to a highly significant Wilcoxon test statistic.[11] The curve indicates that in about three of every five randomly chosen pairs of students, one responding and the other nonresponding, the predicted overall response propensity will be greater than that of the nonrespondent. This level of discrimination implies that the variables used in the three models are highly informative but not definitive predictors of a sample student's overall response propensity.

---

[11] Hanley, J. A. and NcNeil, B. J. (1982). "The meaning and use of the area under a receiver-operating characteristic (ROC) curve." *Diagnostic Radiology*, 143:29-36.

Figure 6.22.—ROC curve for overall response propensity
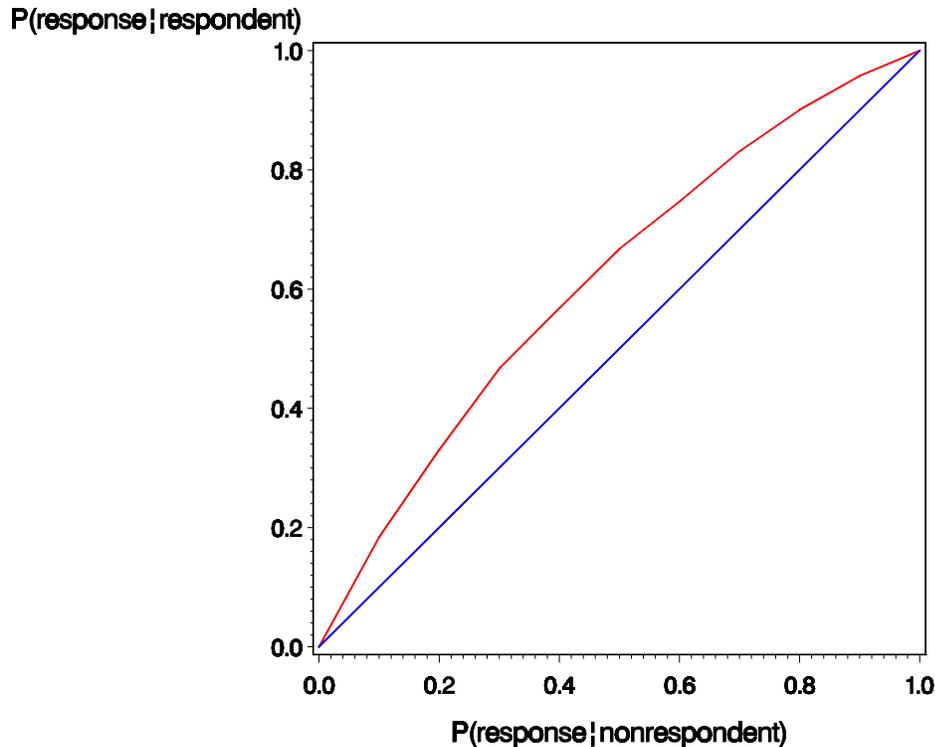
NOTE: Area under the curve = 0.62.
SOURCE: U. S. Department of Education, National Center for Education Statistics Beginning Postsecondary Students Longitudinal Study:1996/2001 (BPS:1996/2001).

D.   Response Rates

1.   Overall Response Rates

The overall BPS:1996/2001 study response rate is an estimate of the proportion of the study population directly represented by the study respondents. Because the BPS:1996/2001 study includes a subsample of both BPS:96/98 and NPSAS:96 nonrespondents, the overall study response rate is the product of the NPSAS:96 institution-level response rate times the BPS student-level response rate. Therefore, the overall BPS study response rates can be estimated directly only for domains defined by institutional characteristics.

Both weighted and unweighted overall study response rates are shown in **table 6.13**, along with their institution and student response rate components. The institution-level response rates shown in this table are the percentage of institutions that provided sufficient data to select the NPSAS student-level sample; these rates were obtained from the NPSAS:96 Methodology

Report (table 3.1). Only the weighted response rates can be interpreted as estimates of the proportion of the BPS study population that are directly represented by the study respondents. **Table 6.13** shows that the student response rate is 83.6 percent and that approximately 76.1 percent of the BPS study population is represented by the respondents. The rate of population coverage does appear to vary by type of institution: the rate is higher for public institutions than for private institutions.

Each weighted student response rate was calculated as the weighted number of respondents divided by the weighted number of eligible students. The weight used in these calculations was the NPSAS:96 base weight that has been adjusted for subsampling the BPS:96/98 nonrespondents; this is the weight variable B01_100. Each overall study response rate was calculated as the product of the NPSAS:96 institutional response rate times the student response rate.

The overall response rates for BPS:1996/2001 are presented in **tables 6. 14** and **6.15**, by type of institution and prior response status. The weighted response rates are higher for students who were interviewed in BPS:96/98 (regardless of their NPSAS:96 response status) than for those who were not interviewed in BPS:96/98. Among those interviewed in both NPSAS:96 and BPS:96/98, the weighted response rate was 79.7 percent, and among those interviewed only in BPS:96/98 (but not in NPSAS:96), the weighted response rate was slightly lower at 76.6 percent. It was lowest among those interviewed in NPSAS:96 but not BPS:96/98 (58.4 percent).

Table 6.13.—Overall BPS:1996/2001 study response rates by type of institution

Institutional sector	Institutions[1]		Students		Overall response rate[2]	
	Response rate		Response rate			
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
All sectors	92.9	91.1	88.3	83.6	82.0	76.1
Public less-than-2-year	93.9	99.6	87.3	83.0	82.0	82.6
Public 2-year	96.4	97.2	86.4	84.6	83.3	82.3
Public 4-year non-doctorate-granting	96.7	96.0	88.9	84.8	86.0	81.4
Public 4-year doctorate-granting	98.4	98.0	90.1	86.0	88.7	84.3
Private not-for-profit 2-year or less	93.3	99.3	83.2	72.4	77.6	71.9
Private not-for-profit 4-year non-doctorate-granting	88.7	83.5	89.2	82.6	79.1	69.0
Private not-for-profit 4-year doctorate-granting	90.9	76.3	91.6	88.2	83.3	67.3
Private for-profit less-than-2-year	81.3	89.8	79.2	71.6	64.4	64.3
Private for-profit 2-year or more	90.5	80.5	83.8	79.1	75.8	63.7

[1] See Table 3.1 in the NPSAS:96 methodology report.
[2] Calculated as the product of the institutional response rate times the student response rate.

NOTE: The weight used is B01_100.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.14.—BPS:1996/2001 response rates by prior response status

Prior response status	Unweighted Response rate	Weighted[1] Response rate	Overall response rate[2] Unweighted	Weighted
Total students	88.3	83.6	82.0	76.1
Interviewed in NPSAS:96 and BPS:96/98	88.6	87.4	82.3	79.7
Interviewed in NPSAS:96 only	65.7	64.1	61.0	58.4
Interviewed in BPS:96/98 only	83.1	84.0	77.2	76.6

[1] The weight used in computing the weighted counts of respondents and nonrespondents and the weighted response rate is B01_100. This weight was applied to the eligible sample members; these are identified by B01ELIG=1.

[2] The overall response rate is the product of the institution response rate from table 6.13 times the student response rate.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

2.   Bias Due to Item Nonresponse

Chapter 4 of this methodology report presents the unweighted response rate for the items with greater than 10 percent nonresponse rate. **Table 6.16** gives these weighted response rates for these same items. This section looks at those items for bias associated with nonresponse to the item. For each of the items with greater than 10 percent nonresponse and at least 50 nonrespondents **tables 6.17–6.23** compare the demographic characteristics of respondents and nonrespondents. Items included are

- cumulative undergraduate GPA,

- Lifetime Learning tax credit (1999) for undergraduates,

- gross annual salary for current job,

- gross annual salary for first postenrollment job,

- gross salary for 2000,

- spouse's gross salary for 2000, and

- total balance due on all credit cards.

Table 6.15.—Overall response rates for BPS:1996/2001 by type of institution and prior response status

Prior response status and institutional sector	Unweighted Response rate	Weighted[1] Respondents	Nonrespondents	Response rate	Overall response rate[2] Unweighted	Weighted
All students						
Institutional sector						
All sectors	88.3	2,779,392	546,328	83.6	82.0	76.1
Public less-than-2-year	87.3	42,777	8,789	83.0	82.0	82.6
Public 2-year	86.4	1,280,241	232,323	84.6	83.3	82.3
Public 4-year non-doctorate-granting	88.9	267,461	48,075	84.8	86.0	81.4
Public 4-year doctorate-granting	90.1	475,097	77,378	86.0	88.7	84.3
Private not-for-profit 2-year or less	83.2	28,506	10,883	72.4	77.6	71.9
Private not-for-profit 4-year non-doctorate-granting	89.2	243,822	51,348	82.6	79.1	69.0
Private not-for-profit 4-year doctorate-granting	91.6	169,731	22,747	88.2	83.3	67.3
Private for-profit less-than-2-year	79.2	172,710	68,617	71.6	64.4	64.3
Private for-profit 2-year or more	83.8	99,048	26,169	79.1	75.8	63.7
NPSAS:96 and BPS:96/98 respondents						
Institutional sector						
All sectors	88.6	2,164,237	310,845	87.4	82.3	79.7
Public less-than-2-year	87.8	42,777	4,760	90.0	82.4	89.6
Public 2-year	86.5	964,393	147,720	86.7	83.4	84.3
Public 4-year non-doctorate-granting	89.2	214,423	29,466	87.9	86.2	84.4
Public 4-year doctorate-granting	90.2	380,695	43,037	89.8	88.7	88.0
Private not-for-profit 2-year or less	84.0	24,171	4,600	84.0	78.4	83.4
Private not-for-profit 4-year non-doctorate-granting	89.6	209,845	25,612	89.1	79.5	74.4
Private not-for-profit 4-year doctorate-granting	91.9	126,041	11,694	91.5	83.5	69.8
Private for-profit less-than-2-year	80.2	127,128	28,941	81.5	65.2	73.1
Private for-profit 2-year or more	84.2	74,763	15,015	83.3	76.2	67.0
NPSAS:96 only respondents						
Institutional sector						
All sectors	65.7	319,900	179,454	64.1	61.0	58.4
Public less-than-2-year	0.0	0	4,029	0.0	0.0	0.0
Public 2-year	77.3	183,783	66,308	73.5	74.5	71.4
Public 4-year non-doctorate-granting	75.0	23,611	10,721	68.8	72.5	66.0
Public 4-year doctorate-granting	63.6	29,672	29,470	50.2	62.6	49.2
Private not-for-profit 2-year or less	42.9	3,465	6,283	35.5	40.0	35.3
Private not-for-profit 4-year non-doctorate-granting	62.5	12,879	16,535	43.8	55.4	36.6
Private not-for-profit 4-year doctorate-granting	87.5	16,456	3,369	83.0	79.5	63.3
Private for-profit less-than-2-year	47.1	27,721	32,242	46.2	38.3	41.5
Private for-profit 2-year or more	69.2	22,313	10,497	68.0	62.7	54.7
BPS:96/98 only respondents						
Institutional sector						
All sectors	83.1	295,255	56,030	84.0	77.2	76.6
Public 2-year	89.5	132,065	18,294	87.8	86.3	85.4
Public 4-year non-doctorate-granting	80.0	29,426	7,887	78.9	77.4	75.7
Public 4-year doctorate-granting	95.2	64,730	4,872	93.0	93.7	91.1
Private not-for-profit 2-year or less	100.0	870	0	100.0	93.3	99.3
Private not-for-profit 4-year non-doctorate-granting	70.6	21,097	9,201	69.6	62.6	58.1
Private not-for-profit 4-year doctorate-granting	78.6	27,234	7,684	78.0	71.4	59.5
Private for-profit less-than-2-year	71.4	17,860	7,434	70.6	58.1	63.4
Private for-profit 2-year or more	75.0	1,972	657	75.0	67.9	60.4

[1] The weight used in computing weighted counts of respondents and nonrespondents and weighted response rate is B01_100. This weight was applied to the eligible sample members; these are identified by B01ELIG=1.

[2] The overall response rate is the product of the institution response rate from table 6.13 times the student response rate.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.16.—Weighted item nonresponse for items with more than 10 percent nonresponse

Item description	Weighted item nonresponse rate[1]
Cumulative undergraduate GPA	17.0
Lifetime Learning tax credit 1999 (undergraduate)	10.2
Gross annual salary for current job	23.4
Gross annual salary for first post-enrollment job	26.5
Gross salary for 2000	14.5
Spouse's gross salary for 2000	23.2
Total balance due on all credit cards	14.9

[1]The weight used is B01AWT.

NOTE: Statistics are based on sample members for whom specific items were applicable and asked. Lifetime Learning tax credit for graduate students and spouse's monthly education loan payment had greater than 10 percent nonresponse but are excluded from this table due to insufficient sample size.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

The tables compare the distributions of respondents and nonrespondents to each item with respect to

- age,
- race/ethnicity,
- gender,
- level of NPSAS:96 institution,
- control of NPSAS:96 institution,
- degree status in 2001,
- employment status,
- highest degree, and
- whether the student is the first generation in postsecondary education.

The bias and the statistical significance of the bias were also estimated. The formulas and methodology described in section C.2 were used for estimating the bias and the statistical significance. The final cross-sectional weight, B01AWT, was used for the calculations. Note that while some variables do show statistically significant biases, the actual bias is generally very small.

Table 6.17.—Comparison of item respondents and nonrespondents for "cumulative undergraduate GPA"

Demographic characteristics	Percent estimate [1]			Estimated bias
	Total	Respondent	Nonrespondent	
Age				
19 or younger	77.43	79.61	66.72	2.1869 *
20 to 23	11.53	10.66	15.80	-0.8727
24 to 29	5.01	4.29	8.52	-0.7179 *
30 to 39	3.49	3.36	4.12	-0.1286
40 or older	2.54	2.08	4.83	-0.4676
Race/ethnicity				
White, non-Hispanic	72.63	74.45	63.75	1.8142 *
Black, non-Hispanic	10.65	9.81	14.78	-0.8431
Hispanic	9.33	8.65	12.65	-0.6785
Asian/Pacific Islander	5.83	5.86	5.67	0.0338
American Indian/Alaska Native	0.89	0.71	1.80	-0.1858
Other	0.66	0.52	1.35	-0.1406
Gender				
Male	46.43	47.76	39.96	1.3215 *
Female	53.57	52.24	60.04	-1.3215 *
Institution level				
4-year	51.40	55.62	30.72	4.2225 *
2-year	45.32	42.77	57.85	-2.5579 *
Less-than-2-year	3.28	1.61	11.43	-1.6645 *
Institution control				
Public	75.73	76.18	73.54	0.4476
Private not-for-profit	19.58	20.84	13.38	1.2647 *
Private for-profit	4.69	2.98	13.07	-1.7123 *
Received degree by June 2001				
Yes	60.54	63.20	47.55	2.6545 *
No	39.46	36.80	52.45	-2.6545 *
Currently employed				
Yes	66.23	65.63	69.21	-0.6045
No	33.77	34.37	30.79	0.6045
Highest degree				
Bachelor's or higher	40.87	46.12	15.14	5.2538 *
Associate	12.25	11.98	13.58	-0.2708
Certificate	7.43	5.10	18.83	-2.3285 *
None	39.46	36.80	52.45	-2.6545 *
First generation in postsecondary education				
Yes	34.31	32.44	43.94	-1.8728 *
No	65.69	67.56	56.06	1.8728 *

* Bias is significant at the $p<0.05/(c-1)$ level, where c is the number of categories within the primary variable.

[1] Weighted estimates, using the BPS:1996/2001 analysis weight B01AWT.

NOTE: Demographic characteristics are from base year data (NPSAS:96) with the exception of the degree and employment categories. Statistics are based on sample members for whom specific items were applicable and asked.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.18.—Comparison of item respondents and nonrespondents for "Lifetime Learning tax credit 1999 (undergraduate)"

Demographic characteristics	Percent estimate [1]			Estimated bias
	Total	Respondent	Nonrespondent	
Age				
19 or younger	81.07	79.81	92.11	-1.2558 *
20 to 23	10.41	11.14	3.99	0.7302 *
24 to 29	3.41	3.59	1.85	0.1777
30 to 39	3.19	3.36	1.74	0.1650
40 or older	1.92	2.11	0.31	0.1830 *
Race/ethnicity				
White, non-Hispanic	75.15	74.28	82.80	-0.8700
Black, non-Hispanic	9.41	9.58	7.90	0.1712
Hispanic	7.93	8.47	3.17	0.5410 *
Asian/Pacific Islander	6.20	6.29	5.43	0.0876
American Indian/Alaska Native	0.80	0.88	0.11	0.0782 *
Other	0.51	0.50	0.58	-0.0080
Gender				
Male	47.91	48.11	46.08	0.2074
Female	52.09	51.89	53.92	-0.2074
Institution level				
4-year	56.69	55.56	66.68	-1.1357 *
2-year	41.44	42.63	31.02	1.1855 *
Less-than-2-year	1.86	1.82	2.30	-0.0498
Institution control				
Public	75.35	75.86	70.81	0.5165
Private not-for-profit	21.94	21.30	27.52	-0.6345 *
Private for-profit	2.72	2.83	1.68	0.1180
Received degree by June 2001				
Yes	66.38	65.32	75.72	-1.0623 *
No	33.62	34.68	24.28	1.0623 *
Currently employed				
Yes	65.76	65.71	66.24	-0.0532
No	34.24	34.29	33.76	0.0532
Highest degree				
Bachelor's or higher	49.38	48.12	60.47	-1.2606 *
Associate	12.12	12.23	11.19	0.1064
Certificate	4.88	4.97	4.07	0.0920
None	33.62	34.68	24.28	1.0623 *
First generation in postsecondary education				
Yes	31.00	31.87	23.28	0.8734 *
No	69.00	68.13	76.72	-0.8734 *

* Bias is significant at the $p < 0.05/(c-1)$ level, where c is the number of categories within the primary variable.

[1] Weighted estimates, using the BPS:1996/2001 analysis weight B01AWT.

NOTE: Demographic characteristics are from base year data (NPSAS:96) with the exception of the degree and employment categories. Statistics are based on sample members for whom specific items were applicable and asked.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.19.—Comparison of item respondents and nonrespondents for "gross annual salary for current job"

Demographic characteristics	Percent estimate [1]			Estimated bias
	Total	Respondent	Nonrespondent	
Age				
19 or younger	66.80	70.21	55.61	3.4124 *
20 to 23	12.53	11.27	16.67	-1.2600
24 to 29	8.30	7.86	9.75	-0.4422
30 to 39	6.71	5.59	10.40	-1.1257 *
40 or older	5.65	5.07	7.57	-0.5845
Race/ethnicity				
White, non-Hispanic	73.86	75.92	67.09	2.0627 *
Black, non-Hispanic	11.67	9.97	17.22	-1.6950 *
Hispanic	9.07	8.90	9.63	-0.1709
Asian/Pacific Islander	3.88	3.94	3.67	0.0639
American Indian/Alaska Native	0.92	0.67	1.75	-0.2520
Other	0.61	0.60	0.64	-0.0087
Gender				
Male	46.73	48.38	41.30	1.6569 *
Female	53.27	51.62	58.70	-1.6569 *
Institution level				
4-year	41.53	45.17	29.57	3.6462 *
2-year	48.74	46.89	54.80	-1.8492 *
Less-than-2-year	9.73	7.94	15.63	-1.7970 *
Institution control				
Public	71.43	71.69	70.59	0.2568
Private not-for-profit	16.53	17.91	11.99	1.3832 *
Private for-profit	12.04	10.40	17.42	-1.6399 *
Received degree by June 2001				
Yes	55.84	58.88	45.87	3.0396 *
No	44.16	41.12	54.13	-3.0396 *
Currently employed				
Yes	99.97	99.97	99.99	-0.0061
No	0.03	0.03	0.01	0.0061
Highest degree				
Bachelor's or higher	34.91	39.58	19.59	4.6729 *
Associate	8.41	8.34	8.64	-0.0706
Certificate	12.52	10.96	17.65	-1.5627 *
None	44.16	41.12	54.13	-3.0396 *
First generation in postsecondary education				
Yes	43.09	40.08	53.44	-3.0093 *
No	56.91	59.92	46.56	3.0093 *

* Bias is significant at the $p<0.05/(c-1)$ level, where c is the number of categories within the primary variable.

[1] Weighted estimates, using the BPS:1996/2001 analysis weight B01AWT.

NOTE: Demographic characteristics are from base year data (NPSAS:96) with the exception of the degree and employment categories. Statistics are based on sample members for whom specific items were applicable and asked.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.20.—Comparison of item respondents and nonrespondents for "gross annual salary for first postenrollment job"

Demographic characteristics	Percent estimate [1]			Estimated bias
	Total	Respondent	Nonrespondent	
Age				
19 or younger	80.03	81.45	76.06	1.4278
20 to 23	9.79	9.15	11.59	-0.6447
24 to 29	4.12	3.85	4.87	-0.2698
30 to 39	3.63	3.05	5.24	-0.5815
40 or older	2.43	2.50	2.24	0.0682
Race/ethnicity				
White, non-Hispanic	74.74	77.57	66.90	2.8234 *
Black, non-Hispanic	9.85	8.43	13.78	-1.4140
Hispanic	8.11	7.51	9.79	-0.6032
Asian/Pacific Islander	5.41	5.14	6.16	-0.2696
American Indian/Alaska Native	1.17	0.62	2.71	-0.5509
Other	0.72	0.73	0.68	0.0144
Gender				
Male	46.01	48.88	38.05	2.8646 *
Female	53.99	51.12	61.95	-2.8646 *
Institution level				
4-year	56.30	58.86	49.17	2.5654 *
2-year	40.71	39.11	45.16	-1.6008
Less-than-2-year	2.99	2.03	5.67	-0.9646 *
Institution control				
Public	72.89	72.78	73.20	-0.1124
Private not-for-profit	22.35	23.35	19.59	0.9953 *
Private for-profit	4.76	3.87	7.21	-0.8829 *
Received degree by June 2001				
Yes	72.51	74.93	65.78	2.4216 *
No	27.49	25.07	34.22	-2.4216 *
Currently employed				
Yes	92.96	94.96	87.41	1.9994 *
No	7.04	5.04	12.59	-1.9994 *
Highest degree				
Bachelor's or higher	54.56	59.20	41.67	4.6387 *
Associate	10.33	9.46	12.76	-0.8736
Certificate	7.62	6.28	11.36	-1.3435
None	27.49	25.07	34.22	-2.4216 *
First generation in postsecondary education				
Yes	32.70	31.28	36.63	-1.4280
No	67.30	68.72	63.37	1.4280

* Bias is significant at the $p<0.05/(c-1)$ level, where c is the number of categories within the primary variable.

[1] Weighted estimates, using the BPS:1996/2001 analysis weight B01AWT.

NOTE: Demographic characteristics are from base year data (NPSAS:96) with the exception of the degree and employment categories. Statistics are based on sample members for whom specific items were applicable and asked.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.21.—Comparison of item respondents and nonrespondents for "gross salary for 2000"

Demographic characteristics	Percent estimate [1]			Estimated bias
	Total	Respondent	Nonrespondent	
Age				
19 or younger	68.57	69.56	62.74	0.9873
20 to 23	12.83	12.72	13.51	-0.1141
24 to 29	7.86	7.91	7.58	0.0476
30 to 39	6.02	5.35	10.01	-0.6758 *
40 or older	4.71	4.47	6.16	-0.2450
Race/ethnicity				
White, non-Hispanic	71.63	72.72	65.16	1.0955 *
Black, non-Hispanic	12.16	10.97	19.16	-1.1862 *
Hispanic	10.35	10.20	11.22	-0.1477
Asian/Pacific Islander	4.50	4.74	3.07	0.2417
American Indian/Alaska Native	0.82	0.80	0.93	-0.0191
Other	0.55	0.57	0.46	0.0158
Gender				
Male	46.44	46.64	45.23	0.2039
Female	53.56	53.36	54.77	-0.2039
Institution level				
4-year	42.29	43.61	34.46	1.3255 *
2-year	49.35	48.67	53.37	-0.6819
Less-than-2-year	8.36	7.72	12.16	-0.6436 *
Institution control				
Public	73.27	73.30	73.14	0.0228
Private not-for-profit	16.16	16.61	13.47	0.4554 *
Private for-profit	10.57	10.09	13.39	-0.4781 *
Received degree by June 2001				
Yes	52.01	52.45	49.41	0.4414
No	47.99	47.55	50.59	-0.4414
Currently employed				
Yes	75.63	75.13	78.55	-0.4958
No	24.37	24.87	21.45	0.4958
Highest degree				
Bachelor's or higher	30.62	31.76	23.91	1.1378 *
Associate	10.29	10.37	9.85	0.0745
Certificate	11.10	10.33	15.65	-0.7709
None	47.99	47.55	50.59	-0.4414
First generation in postsecondary education				
Yes	41.08	40.17	46.57	-0.9076 *
No	58.92	59.83	53.43	0.9076 *

* Bias is significant at the p<0.05/(c-1) level, where c is the number of categories within the primary variable.

[1] Weighted estimates, using the BPS:1996/2001 analysis weight B01AWT.

NOTE: Demographic characteristics are from base year data (NPSAS:96) with the exception of the degree and employment categories. Statistics are based on sample members for whom specific items were applicable and asked.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.22.—Comparison of item respondents and nonrespondents for "spouse's gross salary for 2000"

Demographic characteristics	Percent estimate [1]			Estimated bias
	Total	Respondent	Nonrespondent	
Age				
19 or younger	49.19	51.94	40.11	2.7431 *
20 to 23	15.07	15.53	13.54	0.4619
24 to 29	13.34	13.79	11.87	0.4434
30 to 39	12.22	10.06	19.38	-2.1627 *
40 or older	10.18	8.69	15.10	-1.4858
Race/ethnicity				
White, non-Hispanic	79.32	82.39	69.13	3.0760 *
Black, non-Hispanic	8.81	6.69	15.81	-2.1137
Hispanic	8.17	7.48	10.46	-0.6923
Asian/Pacific Islander	2.20	1.94	3.05	-0.2564
American Indian/Alaska Native	0.72	0.84	0.32	0.1192
Other	0.79	0.65	1.23	-0.1328
Gender				
Male	35.37	34.08	39.62	-1.2847
Female	64.63	65.92	60.38	1.2847
Institution level				
4-year	28.42	30.68	20.92	2.2634 *
2-year	59.51	58.93	61.45	-0.5839
Less-than-2-year	12.07	10.39	17.64	-1.6795 *
Institution control				
Public	75.51	76.27	73.01	0.7548
Private not-for-profit	10.74	11.45	8.37	0.7152 *
Private for-profit	13.75	12.28	18.62	-1.4700 *
Received degree by June 2001				
Yes	44.81	45.52	42.46	0.7088
No	55.19	54.48	57.54	-0.7088
Currently employed				
Yes	74.29	74.14	74.78	-0.1482
No	25.71	25.86	25.22	0.1482
Highest degree				
Bachelor's or higher	18.98	20.62	13.57	1.6347 *
Associate	10.34	10.23	10.70	-0.1094
Certificate	15.49	14.67	18.20	-0.8165
None	55.19	54.48	57.54	-0.7088
First generation in postsecondary education				
Yes	50.76	47.28	62.43	-3.4788 *
No	49.24	52.72	37.57	3.4788 *

* Bias is significant at the $p<0.05/(c-1)$ level, where c is the number of categories within the primary variable.

[1] Weighted estimates, using the BPS:1996/2001 analysis weight B01AWT.

NOTE: Demographic characteristics are from base year data (NPSAS:96) with the exception of the degree and employment categories. Statistics are based on sample members for whom specific items were applicable and asked.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

Table 6.23.—Comparison of item respondents and nonrespondents for "total balance on all credit cards"

Demographic characteristics	Percent estimate [1]			Estimated bias
	Total	Respondent	Nonrespondent	
Age				
19 or younger	70.02	71.10	63.82	1.0881
20 to 23	11.86	11.55	13.60	-0.3070
24 to 29	8.82	8.85	8.61	0.0357
30 to 39	5.46	5.14	7.29	-0.3209
40 or older	3.85	3.35	6.67	-0.4960
Race/ethnicity				
White, non-Hispanic	73.22	73.23	73.16	0.0112
Black, non-Hispanic	12.34	11.61	16.54	-0.7366
Hispanic	9.17	9.52	7.19	0.3477
Asian/Pacific Islander	4.26	4.52	2.77	0.2628
American Indian/Alaska Native	0.47	0.50	0.35	0.0219
Other	0.53	0.62	0.00	0.0929
Gender				
Male	42.79	42.84	42.46	0.0575
Female	57.21	57.16	57.54	-0.0575
Institution level				
4-year	42.28	44.06	32.19	1.7735 *
2-year	50.65	49.47	57.37	-1.1807
Less-than-2-year	7.07	6.47	10.44	-0.5929
Institution control				
Public	76.47	76.13	78.42	-0.3417
Private not-for-profit	14.48	15.28	9.89	0.8068 *
Private for-profit	9.05	8.58	11.70	-0.4651
Received degree by June 2001				
Yes	48.88	49.07	47.76	0.1959
No	51.12	50.93	52.24	-0.1959
Currently employed				
Yes	71.84	71.36	74.58	-0.4817
No	28.16	28.64	25.42	0.4817
Highest degree				
Bachelor's or higher	28.21	30.08	17.63	1.8608 *
Associate	9.67	9.01	13.43	-0.6619
Certificate	10.99	9.99	16.70	-1.0030
None	51.12	50.93	52.24	-0.1959
First generation in postsecondary education				
Yes	39.95	38.51	48.19	-1.4332 *
No	60.05	61.49	51.81	1.4332 *

* Bias is significant at the $p<0.05/(c-1)$ level, where c is the number of categories within the primary variable.

[1] Weighted estimates, using the BPS:1996/2001 analysis weight B01AWT.

NOTE: Demographic characteristics are from base year data (NPSAS:96) with the exception of the degree and employment categories. Statistics are based on sample members for whom specific items were applicable and asked.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Students Longitudinal Study: 1996/2001 (BPS:1996/2001).

In summary, that there were only seven items in the entire BPS:1996/2001 interview with nonresponse in excess of 10 percent (and at least 50 nonrespondents) indicates very little overall bias due to item nonresponse. Among the seven items, there were significant differences in distributions of the demographic variables between the total sample and the respondents to the variable. However, these differences, while statistically significant due to the large BPS:1996/2001 sample size, were generally small and all less than 5.3 percent. Therefore, while some demographic characteristics were significantly associated with response to these questionnaire items, the amount of bias is fairly small.