Using Weighting Adjustments to Compensate for Survey Nonresponse

Gary R. Pike
Executive Director
Information Management & Institutional Research
Indiana University Purdue University Indianapolis
355 N. Lansing St., AO 139
Indianapolis, IN 46202-2896
pikeg@iupui.edu

Paper presented at the annual meeting of the Association for Institutional Research, Kansas City, Missouri, June 2007.

Using Weighting Adjustments to Compensate for Survey Nonresponse

Weighting adjustments are used in some studies to compensate for biased estimators produced by survey nonresponse. Using data from the 2004 National Survey of Student Engagement (NSSE) and the NSSE poststratification weighting algorithm, this study found that weighting adjustments were needed for some, but not all institutions. Unfortunately, no simple criterion for determining when weighting adjustments were needed could be identified. In addition, weighting adjustments reduced the precision of estimators for a majority of the institutions. Based on these findings, institutions and researchers concerned about compensating for nonresponse would be well advised to carefully evaluate the need for, and consequences of, weighting before employing weighting adjustments.

Surveys of faculty and students are an important method of data collection in higher education research. Dey and his colleagues identified approximately 200 published studies and reports that were based on national surveys (Dey et al., 1997). My review of research published in the *Journal of College Student Development, Journal of Higher Education, Research in Higher Education,* and *Review of Higher Education* during 2005 revealed that slightly more than 60% of the studies made use of survey data (Pike, 2007). In addition, the National Center for Postsecondary Improvement found that three-quarters of the colleges and universities responding to its questionnaire reported using surveys in their assessment efforts (Peterson, Einarson, Augustine, & Vaughan, 1999).

Survey costs usually require that researchers interview or send questionnaires to samples of individuals, analyze the results, and make inferences about the population based on the responses of those sampled (Groves, 1989). As a consequence, obtaining a sufficient number of responses is critical to ensuring the accuracy and appropriateness of the generalizations being made. Unfortunately, response rates have been declining in surveys of the general populace and in surveys of faculty and students (de Leeuw & de Heer, 2002; Jones, 1996; Porter & Whitcomb, 2005). Dey (1997) reported that response rates on surveys conducted as part of the Cooperative Institutional Research Program (CIRP) declined from 60% in 1961-1965 to 21% in 1987-1991.

A serious problem posed by nonresponse is that differences in response rates across subgroups may produced biased estimators of population parameters (Kish, 1965). As a consequence, the results of many current surveys may be systematically over- or underestimating levels of student engagement and satisfaction (Porter & Whitcomb, 2005). Weighting adjustments can effectively compensate for bias due to differences in response rates across population subgroups, but the corrections frequently come at the cost of less precise estimators (Kalton, 1983a; Vartivarian, 2004). Unfortunately, relatively little practical information is available about when weighting adjustments are needed or what the consequences of weighting may be (Mandell, 1974). Given the lack of clear guidance about weighting, it is not surprising that relatively few published studies use weights to adjust for sampling designs and/or nonresponse (Hahs-Vaughn, 2006).

The present research examined the use of weighting adjustments to compensate for problems created by survey nonresponse. Drawing on data from the 2004 administration of the National Survey of Student Engagement (NSSE), this study focused on the conditions under which the poststratification weighting adjustments used by NSSE staff are needed to compensate for nonresponse bais. In addition, the study examined the effects of weighting adjustments on the precision of sample statistics as estimators of population parameters.

Background

The importance of having a large number and/or high proportion of respondents is widely recognized, and survey-research specialists use a variety of approaches to increase response rates (Dillman, 2000). These approaches include simple and attractive questionnaire design, persuasive communication to encourage participation, multiple follow-up contacts, and incentives or rewards for participation (Dillman, Eltinge, Groves, & Little, 2002; Groves & Couper, 1998; Groves, Singer, & Corning, 2000; Porter & Whitcomb, 2003). Although proper questionnaire design and administration can significantly increase response rates, some level of nonresponse is inevitable. In the face of the inevitability of survey nonresponse, the critical questions are what are the consequences of nonresponse and what can be done to minimize the adverse effects of nonresponse.

Consequences of Survey Nonresponse

Survey nonresponse can be a serious problem because it may produce biased estimators of survey variables (Dillman, Eltinge, Groves, & Little, 2002; Oh & Scheuren, 1983; Tremblay, 1986; Vartivarian, 2004). There are at least two ways in which nonresponse can lead to biased estimators. First, nonresponse bias can occur when respondents and nonrespondents differ on survey variables. Consider the following example of a campus survey of student engagement and satisfaction: If respondents are more involved in campus activities and more satisfied with their college experiences than nonrespondents, estimators (e.g., sample means) of engagement and satisfaction will be biased—overstating levels of student engagement and satisfaction with college. This type of bias will be small if the proportion of nonrespondents is small or if the differences between respondents and nonrespondents on survey variables are small. However, high nonresponse rates increase the likelihood of bias even when there are small differences between respondents and nonrespondents. Jones (1996) noted that for a modest difference between respondents and nonrespondents (e.g., effect size = 0.25), a 50% response rate gives a researcher a 48% chance of correctly estimating the population mean within a 95% confidence interval.

Bias is not limited to estimators of population means. When the parameter of interest is the population total, survey nonresponse will always produce a sample total that underestimates the population total (Groves & Couper, 1998). When the difference between two sample means is used as the estimator of the difference between subgroup means in the population, bias will be influenced by the nonresponse rates for the subgroups and differences between the population means of respondents and nonrespondents (Kalton, 1983b). In some instances biases will be additive; in other cases the biases may cancel out. Nonresponse can even affect regression (i.e., slope) coefficients, influencing both the numerator and denominator of slope calculations (Groves & Couper, 1998). If the relationship between x and y is stronger for respondents than nonrespondents, the regression coefficient will be overestimated. If the relationship between x and y is stronger for nonrespondents than for respondents, the regression coefficient will be underestimated.

Studies of the extent to which differences between respondents and nonrespondents pose a threat to the validity of survey research are equivocal, although the preponderance of evidence suggests that these differences are relatively minor for many types of surveys. Moore and Tarni (2002) reviewed two studies that used multiwave follow-up procedures with surveys. From one wave to the next, they found that there were significant differences between respondents and nonrespondents on key survey variables. In contrast, Curtin, Presser, and Singer (2000), Keeter et al. (2000), and Merkle and Edelman (2002) found very few meaningful differences between respondents and nonrespondents in surveys of the general population. Based on their review of research using public opinion surveys, Groves, Presser, and Dipko (2004) concluded that respondents and nonrespondents do not differ significantly with respect to survey variables.

Research on respondents and nonrespondents in higher education surveys have reached similar conclusions. In a study of three samples of freshmen with substantially different response rates, Hutchison, Tollefson, and Wigington (1987) did not find significant differences in responses to survey questions. Similarly, Kuh (2001) reported that a follow-up telephone survey of nonrespondents to the National Survey of Student Engagement (NSSE) found only minor differences in responses to survey questions. He also noted that the observed differences may have been an artifact of the telephone survey eliciting more positive response (due to social demand) than the original paper-and-pencil survey.

A second type of nonresponse bias occurs when response rates differ across subgroups, and the subgroups differ in their mean responses to survey questions (Jones, 1996; Kalton, 1983a). In the campus survey example described previously, if females are more involved and more satisfied with college than males *and* if the response rate for females is greater than the response rate for males, survey results will overestimate levels of involvement and satisfaction in the population. Differences among groups will also be biased if the groups being compared are related to, but not the same as, the subgroups with different response rates. For example, if full-time students are disproportionately female and females are more involved and satisfied with college, full-time part-time differences may be exaggerated.

The extent to which differences in the response rates of subgroups affect the results of correlational studies is a matter of some debate. Dey (1997) found that differences in response rates had substantial effects on means, but almost no effect on regression results. In contrast, DuMochel and Duncan (1983) argued that differences in response rates may affect regression results, particularly when important explanatory variables are omitted from the regression model. They recommended that adjustments for nonresponse be used in those instances.

Several studies of nonresponse in general-population surveys have found that socio-demographic characteristics are related to survey participation. Specifically, age, gender, and socioeconomic status have been linked to the likelihood of participating in surveys (Goyder, Warriner, & Miller, 2002; Groves, Cialdini, & Couper, 1992; Groves & Couper, 1998). Similarly, studies of nonresponse in higher education surveys have found differences in response rates to be associated with gender, ethnicity, and academic ability (Dey, 1997; Porter & Umbach, 2006; Porter & Whitcomb, 2005). Differential response rates are also present in NSSE. A national norming study found that females and full-time students had higher response rates than males and part-time students (Kuh et al., 2001).

Weighting Adjustments for Nonresponse

The differences between the two sources of nonresponse bias are important in deciding whether the use of weighting adjustments is appropriate for a given study. Weighting adjustments can compensate for bias when there are differences in the response rates of subgroups, those subgroups differ significantly on the survey variables, and there are *not* meaningful differences between respondents and nonrespondents. Weighting adjustments for nonresponse can never compensate for bias that is attributable to differences in the underlying population means for respondents and nonrespondents (Kalton, 1983b).

Case weighting tends to be the method of choice for educational researchers who use weighting adjustments to compensate for survey nonresponse (Hahs-Vaughn, 2006; National Center for Education Statistics, 2002). When case weighting is used to adjust for nonresponse, cases are grouped into classes based on auxiliary information about survey respondents (Bethlehem, 2002). The decision about what classes to use is not trivial. As Oh and Scheuren (1983) observed, the tendency to select weighting classes based on convenience, rather than appropriateness, frequently leads to a failure to adjust adequately for nonresponse. Appropriate weighting classes are formed from characteristics that are strongly related to survey variables and the response (i.e., respond or not respond) variable (Chapman, 1976; Kalton, 1983a). In the campus survey example, described previously, weighting classes based on gender would be appropriate because females are more likely to respond and are more involved and satisfied than males. Problems can arise when the goal of a study is to make inferences about a variety of survey questions and the relationships between survey questions and weight classes vary. In these situations, a clear choice of classification variables will be difficult. Kalton (1983a) recommended identifying a core set of survey questions in advance and then using those questions to guide the selection of weighting classes.

The number of classes used in weighting adjustments is another important consideration. Ideally, all characteristics that are related to the survey and response variables would be used to form weighting classes (Gelman & Carlin, 2002). However, elaborate classification schemes based on many respondent characteristics substantially increase the likelihood that some classes will contain no respondents. If there are no respondents in a class, weights cannot be calculated. Having few respondents in a class also creates problems because the weights will be unstable and vary substantially from one class to another. The net effect of weighting will be to increase the variance of the weighted estimator relative to the unweighted estimator (Groves & Couper, 1998; Kalton, 1983a; Thomsen, 1973; Vartivarian, 2004). The weighted estimator will be unbiased, but less precise than the unweighted estimator.

Once weighting classes have been defined, survey researchers must select the weighting method to be used. The two most frequently used approaches are population weighting and sample weighting. The choice of one weighting method over the other usually depends on what data are available to researchers. When population weighting adjustments are used, the respondent sample is weighted so that the weighted sample distribution is the same as the distribution of the population across classes. Population weighting adjustments require that survey researchers have data about the distribution of the population and respondents across weighting classes. Data about the distribution of nonrespondents is not needed. Population weighting adjustments are applied to the scores of individual respondents (i.e., case weights) in order to “weight up” the number of respondents to the number of individuals in the population. This process allows researchers to compensate for problems of noncoverage (i.e., some individuals being left out of the sampling frame) in addition to problems of nonresponse (Kalton, 1983a). In some cases, weighting a sample up to the size of the population is undesirable because it substantially increases statistical power (see Thomas, 2006; Warwick & Lininger, 1975). In those instances, population weighting adjustments that preserve the respondent sample size should be used (see Kalton, 1983b).

In some studies, researchers will not have access to data about the population, but they will have access to data about respondents and nonrespondents. In those studies, sample weighting would be an appropriate method of compensating for nonresponse. Sample weighting adjustments weight respondents within classes so that the profile of respondents across classes is equivalent to the profile of the entire survey sample. The general formula for sample weighting of class means is identical to the formula for population weighting, except for the calculation of weights. Sample weights represent the proportion of the sample in a given class, rather than the proportion of the population in that class. When sample weighting is applied to respondents' scores, the size of the weighted sample of respondents will be the total sample size. As with population weighting, it may be important for the weighted sample size to correspond to the number of respondents. In that case, a modified weighting formula, based on relative weights, should be used.

The effect of weighting adjustments on the precision of estimators (i.e., sampling variances and standard errors of means) is a matter of considerable debate. Kish (1965) argued that weighting adjustments for nonresponse always decrease the precision of estimators. Kalton (1983a) and Kalton and Kasprzyk (1986) argued that weighting adjustments frequently increase the variances and standard errors of estimators, thereby decreasing precision. They were quick to acknowledge, however, that a loss of precision does not occur in all instances and noted that loss of precision is more likely when class means are nearly equal, response rates for classes vary significantly, and/or weighting classes contain relatively few elements. Vartivarian (2004) also argued that loss of precision does not inhere in weighting adjustments. In fact, she argued that nonresponse weighting can increase the precision of estimators in those instances where weighting adjustments are effective in reducing bias.

Thus, the effectiveness of weighting adjustments in compensating for problems created by survey nonresponse remains an open question. Weighting can offset bias resulting from differences in response rates across classes when differential response rates are coupled with differences in class means on survey variables. However, differential response rates and differences in class means do not always require weighting adjustments (Vartivarian, 2004). In addition, reduction in bias may be offset by loss of precision in point estimators. What is lacking in the literature, and what is needed by survey researchers, are practical guidelines for employing weighting adjustments (Mandell, 1974). The usefulness of criteria grounded in practice can be seen in the weighting procedures employed in NSSE.

NSSE 2004 Weighting Adjustments

As previously noted, the NSSE norms report found that females and full-time students were significantly more likely to respond to the survey than were males and part-time students (Kuh et al., 2001). In addition, the survey results presented in the norms report showed that responses to survey items differed by sex and enrollment status. Similar results have been reported in other studies using NSSE data (Kuh et al., 2007; Pike, 2004). In general, females and full-time students report higher levels of engagement than males and part-time students. Thus, responses to the NSSE survey are likely to overstate levels of student engagement in the population. Because this source of bias is correctable, NSSE staff employed a form of poststratification weighting in the 2004 institutional and national reports (National Survey of Student Engagement, 2004). Specifically, students' responses to survey questions were weighted for sex and enrollment status using Little's (1993) poststratification weighting algorithm. Survey respondents were cross-classified in four weighting classes: part-time male, full-time male, part-time female, and full-time female. Weights for a given class (*h*) were calculated using the formula

$$W_h = rP_h / r_h$$

Where *r* is the total number of institutional respondents, P_h is the proportion of the population in a given weighting class, and r_h is the number of respondents in the weighting class. Separate weights were calculated for first-year students and Seniors. If data on sex or enrollment status were not available for a student, no weight was assigned to that student. Approximately 5% of the students were not assigned weights and not included in the calculation of benchmark scores due to missing data. In order to address problems created by small cell sizes in weighting classes, weights of unity (1.00) were assigned when there were less than five respondents in a weighting class (r_h) (National Survey of Student Engagement, 2004).

Although the NSSE reports provide information about weighting procedures, and case weights are included in the data files sent to institutions, information about the need for and consequences of weighting is not provided. As a result, colleges and universities do not have clear guidelines for determining when weighting adjustments are appropriate. To address this shortcoming, the present research examined the need for and consequences of NSSE weighting adjustments. The descriptive analyses focused on three questions:

1. Do the NSSE weighting adjustments significantly affect institutional benchmark scores?
2. What institutional and survey-response characteristics are associated with differences in weighted and unweighted benchmark scores?
3. What effect do weighting adjustments have on the precision of benchmark scores, and what institutional and survey-response characteristics are related to the precision of weighted benchmark scores?

Research Methods

Data Source

The data for this study came from the NSSE 2004 administration of *The College Student Report*. The initial sample consisted of approximately 560,000 first-year and senior students attending 473 four-year colleges and universities. The institutions that participated in the survey are very similar to the national profile in terms of geographic region and urban-rural locale. Public institutions and master's universities were overrepresented, whereas baccalaureate-general colleges were underrepresented among participating institutions (National Survey of Student Engagement, 2004).

Students at 200 colleges and universities had the option of responding via a paper-and-pencil questionnaire or via the web, and 175 schools opted for web-only administration. In 2004, NSSE introduced Web+ administration which included multiple electronic contacts and mailing a paper-and-pencil survey to selected nonrespondents. A total of 98 institutions selected this method of administration. Approximately 13% of the respondents completed the paper version of the survey, and 87% used the web (National Survey of Student Engagement, 2004). Generally, administration mode does not affect NSSE results, except that Web respondents tend to report greater use of electronic technology (Carini et al., 2003).

Only seniors were included in the current analyses to ensure sufficient numbers of part-time students. As a further restriction, only institutions with both expected and observed counts of at least five students in every weighting class were included in the study. This requirement eliminated from the study all institutions with weights of unity (1.00). One hundred twenty-nine institutions met these criteria. Of the total, 34% were doctoral/research universities, 48% were master's universities, and 11% were baccalaureate colleges. The remaining institutions were specialized colleges or universities that were not included in the six most prominent Carnegie 2000 classification types. Approximately 73% of the institutions included in the study were public colleges and universities. Undergraduate enrollment ranged from slightly more than 800 students to more than 37,000 students. Average undergraduate enrollment was approximately 10,000 students. The average response rate for seniors was slightly more than 34% and ranged from 10% to 60%.

Measures

The five NSSE benchmarks were the survey variables used in this study. The benchmarks represent clusters of activities that research shows are linked to positive educational outcomes. The Academic Challenge benchmark focuses on activities that demonstrate an institution emphasizes the importance of academic effort and sets high expectations for student performance, particularly in the areas of writing and higher-order thinking. Active and Collaborate Learning benchmark questions ask students to report on the extent to which they are required to think about and apply what they are learning and to work with other students to solve problems and master difficult material. Student-Faculty Interaction items ask students to report on how often they interact with faculty inside and outside the classroom. The Enriching Educational Experiences benchmark covers a wide range of educationally purposeful learning activities inside and outside the classroom. It also includes students' reports of their diversity experiences and experiences with technology. The final benchmark, Supportive Campus Environment, focuses on students' perceptions of institutional commitment to student success and the quality of students' interactions with peers, faculty, and administration (Kuh et al., 2001).

Two measures were used to form weighting classes. Sex, whether a student was female or male, was taken from institutional data reported to NSSE, and full-time part-time enrollment status was measured using student reports from the survey. The procedures used to form weighting classes in this study were identical to those used by NSSE staff (National Survey of Student Engagement, 2004). The four weight classes used in the study were males enrolled part time, males enrolled full time, females enrolled part time, and females enrolled full time. The need for weighting adjustments was represented by the difference between weighted and unweighted benchmark scores. Because the direction of the difference between weighted and unweighted scores was not important, the absolute value of the difference between weighted and unweighted scores served as the response variable.

Both institutional characteristics and survey-response characteristics were used to answer the second research question and identify the factors associated with the need for weighting adjustments. Five variables were used to represent institutional characteristics in this study. The five variables were the institutions' Carnegie 2000 classifications, public/private control, undergraduate enrollment (in thousands of students), the percent of undergraduates in 2004 who were female, and the percent of undergraduates who were full-time students. Data for all five measures came from the Integrated Postsecondary Education Data System (IPEDS).

Three variables were used to represent the characteristics of institutions' responses to the NSSE survey. The first measure, the institutional response rate was calculated using student-level data provided by the NSSE staff. The second survey-response variable was calculated from the student level data and provided a measure of disproportional response rates across weight classes. The disproportional-response measure was the sum of the absolute differences between the proportions of respondents and the proportions of the population in weight classes. For example, if the population of seniors at an institution was equally divided among the weight classes of part-time males, full-time males, part-time females, and full-time females, and the proportions of respondents were 0.05, 0.20, 0.25, and 0.50, respectively, the disproportional-response score would be 0.50[1]. A higher score on this measure represents greater disproportional response.

The third survey-response variable was a measure of differences in unweighted benchmark scores across weight classes. Differences in benchmark scores across weight classes were represented by the effect size of the group differences. Effect sizes were calculated by testing for differences in unweighted benchmark scores using a oneway analysis of variance. Levels of the independent variable in the analysis were the four weighting classes. Separate analyses were conducted for each institution and estimates of explained variance (i.e., Eta^2) were used as the effect-size measures.

The design-effect correction was the measure of the relative precision of weighted and unweighted scores used to answer the final research question. First, sampling variances and standard errors were calculated for the weighted and unweighted institutional means. The formulas used to calculate variances and standard errors for the unweighted means were the traditional measures of precision for samples under conditions of random selection (Kalton, 1983b). The formulas used to calculate variances and standard errors for the weighted means were developed by Holt and Smith (1979) for poststratification weighting. Specifically, the formula used to calculate sampling variances was

$$\operatorname{var}(\bar{y}) = \sum \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{s_h^2}{n_h}$$

Where N_h was the number of students in the population for a given weighting class, N was the number of students in the population, n_h was the number of respondents in a weighting class, and s_h^2 was the raw-score variance for a weighting class. Little (1993) recommended that this formula be used to calculate sampling variances for his weighting algorithm. The standard error of the mean was defined as the square root of the sampling variance.

Next, design-effect corrections (DEFT) were calculated for each institution (Hahs-Vaughn, 2006). The design-effect correction was the ratio of the standard error of the weighted estimator to the standard error of the unweighted estimator. A design-effect correction greater than 1.00 indicated that the weighted estimator was less precise than the unweighted estimator,

---

[1] $|0.05 - 0.25| + |0.20 - 0.25| + |0.25 - 0.25| + |0.50 - 0.25| = 0.50$

whereas a design-effect correction of less than 1.00 indicated that the weighted estimator was more precise than the unweighted estimator.

Three variables were used to represent the factors influencing precision that were identified by Kalton (1983a) and Kalton and Kasprzyk (1986): (1) differences in weight-class means (i.e., the measure of effect size used in the second phase of the study); (2) disproportional response; and (3) the presence of small cell sizes (i.e., less than 10 students). A single variable, the magnitude of the difference between weighted and unweighted means was used to represent Vartivarian's (2004) claim that precision is influenced by the need for weighting adjustments.

Data Analysis

The data analysis was carried out in three phases corresponding to the study's research questions. Whether weighting adjustments significantly affected institutional benchmark scores was assessed using paired t-tests to determine if weighted and unweighted benchmark scores were significantly different. Tests were conducted for all five senior benchmarks. Institutions were the units of analysis in this and subsequent phases of the data analysis.

The second phase of the data analysis examined the relationships between differences in weighted and unweighted benchmarks and institutional and survey-response characteristics. Analyses in the second phase of the study were limited to differences in weighted and unweighted scores on the Student-Faculty Interaction benchmark. This benchmark was selected because the results of the first phase of the study revealed that the variance in score differences was greatest for this benchmark.

Because of problems of multicolinearity, the unique effects of all institutional and survey-response characteristics on weighted-unweighted scores differences could not be assessed simultaneously. As a result, the relationships between mission and control and weighted-unweighted score differences were assessed using ANOVA procedures. Next, weighted-unweighted score differences were regressed on the remaining institutional and survey-response characteristics. Variables not significantly related to absolute score differences were dropped and the model was re-estimated. The final model served as a guide for predicting the conditions under which weighting adjustments substantially affected Student-Faculty Interaction benchmark scores.

Initially, aggregate measures (i.e., the mean and standard deviation) of the design-effect corrections were calculated and examined the third phase of the study. These aggregate measures provided an indication of whether weighted or unweighted scores were more precise overall. Next, design-effect corrections were regressed on the three variables identified by Kalton (1983a) as influencing precision and the variable identified by Vartivarian (2004) as influencing precision. Based on initial results, nonsignificant factors were dropped and the model was re-estimated. The final model provided an indication of the factors significantly related to increases and decreases in the precision of weighted estimators.

Results

Differences between Weighted and Unweighted Scores

The results of the paired t-tests indicated that weighted and unweighted scores differed for four of the five institutional benchmarks. Weighted and unweighted scores for the Supportive Campus Environment benchmark were not significantly different. Table 1 displays means and standard deviations for the weighted and unweighted benchmark scores. The table also includes means and standard deviations for differences between weighted and unweighted scores, correlations between weighted and unweighted scores, and t-test results. An examination of the results presented in Table 1 revealed that, on average, weighted scores were slightly lower than unweighted scores. Mean differences for the four benchmarks producing statistically significant results ranged from -0.35 to -0.44. Although the weighted and unweighted scores differed significantly, they were highly correlated with correlations ranging from 0.96 to 0.99. Variability in differences between weighted and unweighted scores was greatest for the Student-Faculty Interaction benchmark. For this reason, subsequent analyses focused on SFI scores.

Insert Table 1 about here

Factors Affecting Differences in Weighted and Unweighted Scores

Analysis of variance results indicated that institutional mission, represented by Carnegie classification ($F = 2.06$; $df = 5{,}123$; $p > 0.05$) and public-private control ($F = 0.71$; $df = 1{,}127$; $p > 0.05$) were not significantly related to absolute differences between weighted and unweighted SFI scores. When difference scores were regressed on institutional and survey-response characteristics, the association was statistically significant ($F = 25.27$; $df = 6{,}122$; $p < 0.05$) and explained a substantial proportion of the variance in the difference scores ($R^2 = 0.56$). The results of the regression analyses are presented in Table 2. An examination of the results for the full model revealed that only two survey-response characteristics were significantly related to absolute-difference scores: the degree to which response rates were disproportional to the population across weighting classes and the effect size for raw-score differences across weighting classes. Based on these results, a reduced model, including only the disproportional-response and effect-size measures was specified and tested. The association between absolute-score differences and the two survey-response measures was statistically significant ($F = 61.80$; $df = 2{,}126$; $p < 0.05$) and explained half of the variance in the differences between institutions' weighted and unweighted SFI scores ($R^2 = 0.50$).

Insert Table 2 about here

Using the coefficients from the reduced model, expected differences in weighted and unweighted SFI benchmarks were calculated for various combinations of disproportional response and effect size. These expected differences are presented in Table 3. The values displayed in the table show that even small differences in raw scores across weighted classes can produce large differences between weighted and unweighted scores when there is substantial disproportional response. For example, when the effect size of differences in raw scores across weight classes is only 0.05, but there are large discrepancies in the proportions of actual and expected respondents (0.75), the difference between weighted and unweighted scores is expected to be 2.28, or more than one-half of a standard deviation in unweighted SFI scores (3.91). Likewise, large score differences can occur when disproportional response is low, but raw-score differences across weight classes are large. An effect size of 0.35 and a disproportional-response coefficient of 0.10 produce a coefficient (2.07) that is greater than half of a standard deviation in unweighted SFI sores.

Insert Table 3 about here

Factors Affecting the Precision of Weighted Scores

An examination of the standard errors of the mean for the SFI institutional benchmarks revealed that, on average, the weighted estimators were slightly less precise than the unweighted estimators. The average standard error for the weighted means was 1.76, whereas the average standard error for the unweighted means was 1.71. The ratio of the two standard errors (i.e., the design-effect correction) was 1.03. It is important to note that there was substantial variation in standard errors across institutions. The standard errors for weighted means ranged from 1.16 to 3.13, and the standard errors for unweighted means ranged from 1.03 to 2.70. The magnitude of the design-effect corrections also varied substantially, ranging from 0.79 to 1.59. These data indicate that for some institutions, weighted benchmarks were 20% more precise than unweighted benchmarks. For other institutions, weighted SFI benchmark scores were almost 60% less precise than unweighted benchmarks.

Multiple-regression results for the four-variable model indicated that the four variables were significantly related to the precision of weighted scores ($F = 14.609$; $df = 4{,}125$; $p < 0.05$). The model explained 32.0% of the variance in the design-effect corrections. However, analysis of the parameter estimates for the four variables in the model revealed that only the effect-size and disproportional response measures were significantly related to design-effect corrections. A model containing only these two variables was specified and estimated. Results again indicated that the variables were significantly related to the precision of weighted scores ($F = 27.002$; $df = 2{,}127$; $p < 0.05$). The model accounted for 30.0% of the variance in DEFT scores.

Table 4 presents the results of both regression analyses. An examination of the standardized parameter estimates for the final model reveals that the effect size of differences in weighting class means was negatively related to design-effect corrections ($\beta = -0.466$). That is, as the difference in unweighted means increases across weighting classes, the precision of weighted estimators increases. In contrast, the coefficient representing differences in response rates across weighting classes (i.e., disproportional response) was positively related to DEFT scores ($\beta = 0.406$). The direction of the relationship was such that greater differences in response rates were associated with a loss of precision in parameter estimates.

Insert Table 4 about here

Discussion

Findings

The principal findings of this study can be summarized as follows:

- First, weighted and unweighted scores differed significantly for the Academic Challenge (AC), Active and Collaborative Learning (ACL), Student-Faculty Interaction (SFI), and Enriching Educational Experiences (EEE) benchmarks. Weighted and unweighted institutional scores for the Supportive Campus Environment (SCE) benchmark were not significantly different. Weighted scores were, on average, lower than unweighted scores. However, the differences were not uniform. For several institutions, the differences between weighted and unweighted scores were trivial (i.e., less than one-tenth of a point). For other institutions, weighted institutional benchmark scores were substantially higher than unweighted scores. Weighted and unweighted benchmark scores also were highly correlated, with the correlations for all five benchmarks being greater than 0.95.

- Second, the magnitude of the differences between weighted and unweighted institutional SFI scores was not related to general institutional characteristics, such as Carnegie classification or public/private control. Neither was the magnitude of score differences related to institutional population characteristics, such as the size of the undergraduate student body, percent of undergraduates who were female, or percent of undergraduates who were full-time students. The magnitude of the score differences was also unrelated to the institutional survey response rate. However, measures of disproportional response rates across weighted classes and the magnitude of differences in benchmark scores across weighting classes were significantly and positively related to the magnitude of the differences between weighted and unweighted institutional benchmark scores.

- Third, the standard errors of the mean for weighted SFI benchmark scores were, on average, slightly higher than the standard errors for unweighted scores. Once again, substantial variability was observed. For some institutions, weighted scores were markedly less precise than unweighted scores. For other institutions, weighted scores were more precise than unweighted institutional benchmarks. Overall, the relative precision of weighted SFI scores was significantly related to the magnitude of the difference between weighted and unweighted scores. Two variables were significantly and uniquely related to the relative precision of weighted institutional benchmarks—disproportional response and differences in unweighted SFI scores across weight classes. Higher levels of disproportional response were associated with decreased precision, whereas significant differences in SFI scores were associated with increased precision.

Limitations

Care should be taken not to over generalize the findings of this study. These results are specific to the *College Student Report* and the poststratification weighting algorithm used by NSSE. Different results might have been obtained had a different survey or a different method of weighting been used. In addition, many of the results reported in this research are specific to the Student-Faculty Interaction benchmark. Although the omnibus results were essentially the same for the Academic Challenge, Active and Collaborative Learning, Student-Faculty Interaction and Enriching Educational Experiences benchmarks, it is impossible to determine whether the findings for the second and third phases of this study can be generalized to these benchmarks.

Only one year of data was analyzed in this study. Although the results from the 2004 NSSE survey are generally consistent with results from other NSSE administrations, the findings reported in this research might have differed in unknown ways if data from other years were analyzed. The generalizability of the findings may have been further limited by the fact that a subset of the institutions participating in NSSE 2004 was included in the study. Although the restriction imposed in this research favored colleges and universities with large undergraduate populations, the restriction was essential to ensure that the findings were not an artifact of efforts by NSSE staff to compensate for extremely small cell sizes in weighting. The results of this study are also specific to the two variables used to form weighting classes: sex of the respondent and full-time/part-time enrollment status. Limiting the analyses to these classification variables was clearly justified because sex and enrollment status are the variables used by NSSE staff in poststratification weighting adjustments. Nevertheless, use of different classification variables might have produced different results.

Implications

Despite these limitations, the results of the present study have important implications for theory and practice in survey research. First and foremost, the finding that weighted and unweighted institutional benchmark scores differed significantly strongly suggests that weighting adjustments are needed to compensate for the biasing effects of survey nonresponse. Because differences in response rates across the weighting classes were the same for all five of the institutional benchmarks, the absence of a statistically significant difference between weighted and unweighted Supportive Campus Environment benchmarks indicates that unweighted SCE benchmark scores were not different across weighting classes. This finding supports the claim made by Kalton (1983a) that nonresponse bias requires more than differences in response rates across subgroups. Significant differences in survey-variable means across the same subgroups are also required.

For the four institutional benchmarks with statistically significant differences between weighted and unweighted scores, the weighted benchmarks were, on average, lower than the unweighted scores. This finding is consistent with the results reported in previous studies. Specifically, females and full-time students tend to be more engaged and more likely to respond to the NSSE survey than males and part-time students (Kuh et al., 2001; Kuh et al., 2007; Pike, 2004). The finding also suggests that unweighted comparison-group scores may tend to over estimate levels of student engagement. As a result, institutions interested in comparing their institution's NSSE benchmark scores to the benchmark scores of peer institutions would be wise to rely on weighted benchmark scores.

Although unweighted NSSE benchmarks more often than not overestimate levels of student engagement, results were markedly different for some institutions. The fact that weighted and unweighted scores were virtually identical for several institutions indicates that weighting may not be appropriate for all colleges and universities. Unfortunately, no simple criterion is available to determine when weighting is needed. Weighting adjustments may be needed for some institutions with relatively high response rates, but not needed for other institutions with low response rates. Colleges and universities would be well advised to calculate both weighted and unweighted benchmark scores and then to compare the two scores to determine if the weighted scores differ significantly from unweighted benchmarks. If the two scores are substantially different, it is reasonable to conclude that weighting adjustments are needed.

Weighting adjustments are not a panacea. Overall, weighted NSSE benchmarks tend to be slightly less precise than unweighted scores. This suggests that there may be a tradeoff between compensating for bias due to survey nonresponse and precision of estimation. In some cases weighting adjustments substantially reduce the precision of estimators; in other cases weighting may actually improve the precision of estimators. In still other cases, weighting has little or no effect on precision. Because the relationship between weighting adjustments and the precision of estimators appears to be complex, institutions would be well advised to calculate sampling variances and standard errors for both weighted and unweighted estimators and then compare the relative precision of the estimators before relying on weighting adjustments to compensate for survey nonresponse.

Because weighting adjustments effectively compensate for nonresponse bias that is attributable to disproportional response rates and differences in survey variables across subgroups, but have little effect on institutional means when disproportional response and subgroups differences are not present, it is empting to adopt the practice of always weighting NSSE scores to compensate for the possibility of nonresponse bias. The danger in this course of action lies in the potentially deleterious effects of weighting on precision of measurement. The findings of this study indicate that when response rates differ markedly across population subgroups, but survey means for the subgroups are not significantly different; weighting will have relatively little effect on institutional benchmark scores. However, the precision of the weighted institutional benchmarks will be less than the precision of the unweighted benchmarks. The net result is a loss of precision in estimating population parameters without any measurable gain in bias reduction. Determining whether the current practice of weighting NSSE benchmark scores for all institutions actually results in less accurate estimators for some institutions was beyond the scope of the present study. Nevertheless, it seems likely that at least some institutions that participated in the NSSE survey had substantially different response rates across weighting classes, but relatively small differences in benchmark scores across those classes. For those institutions, weighted institutional benchmarks would actually provide less accurate indicators of student engagement than unweighted benchmarks.

In addition to cautioning institutions to proceed carefully when employing weighting adjustments, the results of this research have important implications for theories of sampling and nonresponse weighting. Most obviously, the results of the present research confirm Kalton's (1983a) claim that weighting adjustments can compensate for nonresponse bias resulting from disproportional response and differences in survey response variables across weighting classes. What is most significant about these findings is that *both* disproportional response *and* differences in survey variables are required before weighting adjustments are needed.

Particularly intriguing were the results regarding the relative precision of weighted and unweighted estimators. Rather than supporting Kish's (1965) claim that loss of precision is inherent in weighting adjustments, the results of this research revealed that weighting adjustments may actually improve the precision of estimators. Furthermore, the results of this study generally supported Kalton's (1983a) claim that weighting reduces the precision of estimators when there are substantial differences in response rates across subgroups and few differences in subgroup means. No evidence was found to support Vartivarian's (2004) claim that gains in precision occur when weighted and unweighted scores differ substantially. Instead, these results suggest that precision may be enhanced when there are small differences in subgroup response rates, but large differences in subgroup means. In this instance weighting may be desirable to improve precision without substantially affecting the point estimate of the population mean.

It is also significant that the results of the present research underscore the need to use weighting adjustments to compensate for bias in point estimators, but suggest that adjustments may not be needed in correlational studies. The present research found that the correlations between weighted and unweighted institutional benchmark scores were all in excess of 0.95. The near-perfect correlations between weighted and unweighted scores indicate that the two sets of scores can be used interchangeably without affecting regression results in meaningful ways. This finding is consistent with the results reported by Dey (1997) and helps explain the reasons behind Dey's findings. It should be recognized, however, that weighted and unweighted scores do not have to be highly correlated. Additional research identifying the factors that influence the magnitude of correlations between weighted and unweighted scores would shed light on the conditions under which weighting adjustments are need in correlational studies.

Conclusion

Even a cursory review of literature reveals that surveys are extensively used in higher education research and assessment, but little attention is paid to the effects of survey nonresponse on bias in and the precision of estimators of population parameters. The results of the present research demonstrate that survey nonresponse can lead to biased institutional benchmarks when there are substantial differences in response rates across population subgroups and benchmark scores differ for those subgroups. Unfortunately, the effects of nonresponse on bias are relatively complex and do not yield simple criteria for determining when weighting adjustments are needed to compensate for nonresponse bias. Calculating and comparing both weighted and unweighted institutional benchmark scores appears to be the best method of determining when weighting adjustments are warranted.

Although weighting adjustments do compensate for the type of nonresponse bias discussed in this research, weighting adjustments come at a cost. Weighting adjustments may substantially reduce the precision of population estimators. Ironically, there are situations in which unbiased (i.e., weighted) estimators may be less accurate than biased estimators. Higher education researchers and assessment professionals engaged in survey research should carefully evaluate the need to weighting adjustments and the consequences of those adjustments. When warranted, weighting adjustments should be used to compensate for bias due to survey nonresponse. When weighting adjustments are not warranted, they should be avoided.

References

Bethlehem, J. G. (2002). Weighting nonresponse adjustments based on auxiliary information. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 275-287). New York: Wiley.

Carini, R.M., Hayek, J.H., Kuh, G.D., Kennedy, J.M., & Ouimet, J.A. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education, 44,* 1-19.

Chapman, D. W. (1976). A survey of nonresponse imputation procedures. *Proceedings of the Social Statistics Section, American Statistical Association, 1976*(1), 245-251.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly, 64,* 413-428.

de Leeuw, E., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 41-54). New York: Wiley.

Dey, E. L. (1997). Working with low survey response rates: The efficacy of weighting adjustments. *Research in Higher Education, 38,* 215-227.

Dey, E. L., Hurtado, S., Rhee, B., Inkelas, K. K., Wimsatt, L. A., & Guan, F. (1997). *Improving research on postsecondary student outcomes: A review of the strengths and limitations of national data resources.* Stanford, CA: National Center for Postsecondary Improvement.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2d ed). New York: Wiley.

Dillman, D. A., Eltinge, J. L., Groves, R. M., & Little, R. J. A. (2002). Survey nonresponse in design, data collection, and analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 3-26). New York: Wiley.

Du Mouchel, W. H., & Duncan, G. J. (1983). Using survey weights in multiple regression analyses of stratified samples. *Journal of the American Statistical Association, 78,* 535-543.

Gelman, A., & Carlin, J. B. (2002). Poststratification and weighting adjustments. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 289-302). New York: Wiley.

Goyder, J., Warriner, K., & Miller, S. (2002). Evaluating socio-economic status (SES) bias in survey nonresponse. *Journal of Official Statistics, 18,* 1-11.

Groves, R. M. (1989). *Survey errors and survey costs.* New York: Wiley.

Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly, 56,* 475-495.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys.* New York: Wiley.

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68,* 2-31.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-salience theory of survey participation: Description and an illustration. *Public Opinion Quarterly, 64,* 299-308.

Hahs-Vaughn, D. L. (2006, Fall). *Weighting omissions and best practices when using large-scale data in educational research.* (Air Professional File No. 101). Tallahassee, FL: Association for Institutional Research.

Holt, D., & Smith, T. M. F. (1979). Post stratification. *Journal of the Royal Statistical Society, Series A (General), 142*(1), 33-46.

Hutchinson, J., Tollefson, N., & Wigington, H. (1987). Response bias in college freshmen's responses to mail surveys. *Research in Higher Education, 26,* 99-106.

Jones, J. (1996). The effects of non-response on statistical inference. *Journal of Health and Social Policy, 8,* 49-62.

Kalton, G. (1983a). *Compensating for missing survey data.* Ann Arbor: University of Michigan Institute for Social Research.

Kalton, G. (1983b). *Introduction to survey sampling* (Sage University Paper series on Quantitative Applications in the Social Sciences, 07-035). Beverly Hills, CA: Sage.

Kalton, G., & Kasprzyk, D. (1986). The treatment of missing survey data. *Survey Methodology, 12,* 1-16.

Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly, 64,* 125-148.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Kuh, G. D. (2001). *The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties.* Bloomington, IN: Indiana University Center for Postsecondary Research. Retrieved February 24, 2006 from the national Survey of Student Engagement Web site: http://nsse.iub.edu/html/psychometric_framework_2002.cfm.

Kuh, G. D., Hayek, J. C., Carini, R. M., Ouimet, J. A., Gonyea, R. M., & Kennedy, J. (2001). *NSSE technical and norms report.* Bloomington, IN: Indiana University Center for Postsecondary Research.

Kuh, G. D. Kinzie, J., Cruce, T., Shoup, R., & Gonyea, R. M. (2007, January). *Connecting the dots: Multi-faceted analyses of the relationships between student engagement results from the NSSE, and the institutional practices and conditions that foster student success.* (Revised final report prepared for the Lumina Foundation for Education grant #2518). Bloomington, IN: Indiana University Center for Postsecondary Research.

Little, R. J. A. (1993). Post-stratification: A modeler's perspective. *Journal of the American Statistical Association, 88,* 1001-1012.

Mandell, L. (1974). When to weight: Nonresponse bias in survey data. *Public Opinion Quarterly, 38,* 247-252.

Merkle, D. M., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 243-257). New York: Wiley.

Moore, D. L., & Tarnai, J. (2002). Evaluating nonresponse error in mail surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 197-211). New York: Wiley.

National Center for Education Statistics. (2002). *Statistical standards.* Retrieved January 3, 2006 from the NCES Web site: http://nces.ed.gov/statprog/2002/stdintro.asp.

National Survey of Student Engagement (2004). *Student engagement: Pathways to collegiate success.* Bloomington, IN: Indiana University Center for Postsecondary Research. Retrieved November 20, 2006 from http://nsse.iub.edu/html/report-2004.cfm.

Oh, H. L., & Scheuren, F. (1983). Weighting and adjustment for unit nonresponse. In W. G. Madow, I. Olkin, & D. B. Rubin (Eds.), *Incomplete data in sample surveys, volume 2, theory and bibliographies* (pp. 143-184). New York: Academic Press.

Peterson, M. W., Einarson, M. K., Augustine, C. H., & Vaughan, D. S. (1999). *Institutional support for student assessment: Methodology and results of a national survey.* Stanford, CA: National Center for Postsecondary Improvement.

Pike, G. R. (2004). Measuring quality: A comparison of *U. S. News* rankings and NSSE benchmarks. *Research in Higher Education, 45,* 193-208.

Pike, G. R. (2007). Adjusting for nonresponse in surveys. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. XXII, pp. 411-449). Dordrecht, The Netherlands: Springer.

Porter, S. R., & Umbach, P. D. (2006). Student survey response rates across institutions: Why do they vary? *Research in Higher Education, 47,* 229-247.

Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education, 44,* 389-407.

Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The role of demographics, engagement, and personality. *Research in Higher Education, 46,* 127-152.

Thomas, S. L. (2006). Sampling: Rationale and rigor in choosing what to observe. In C. F. Conrad & R. C. Serlin (Eds.), *The SAGE handbook for research in education: Engaging ideas and enriching inquiry* (pp. 393-404). Thousand Oaks, CA: Sage.

Thomsen, I. (1973). A note on the efficiency of weighting subclass means to reduce the effects of non-response when analyzing survey data. *Statistisk Tidskrift, 11,* 278-283.

Tremblay, V. (1986). Practical criteria for definition of weighting classes. *Survey Methodology, 12,* 85-97.

Vartivarian, S. L. (2004). *On the formation of weighting class adjustments for unit nonresponse in sample surveys*. (Doctoral dissertation, University of Michigan, 2004). Ann Arbor, MI: ProQuest. (UMI No. 3137952).

Warwick, D. P., & Lininger, C. A. (1975). *The sample survey: Theory and practice*. New York: McGraw-Hill.

Table 1

Comparison of Weighted and Unweighted Institutional Benchmarks

Benchmark	Weighted	Unweighted	Difference	t	Correlation
Academic Challenge	54.9	55.3	0.4	6.34*	0.96
	(2.6)	(2.6)	(0.7)		
Active & Collaborative Learning	48.6	49.0	0.4	5.42*	0.98
	(3.7)	(3.6)	(0.8)		
Student-Faculty Interaction	38.3	38.7	0.4	3.93*	0.97
	(4.1)	(3.9)	(1.0)		
Enriching Educational Experiences	34.7	35.2	0.5	5.27*	0.98
	(4.4)	(4.3)	(0.9)		
Supportive Campus Environment	56.7	56.8	0.1	1.77	0.99
	(5.2)	(5.1)	(0.7)		

* $p < 0.05$

Standard deviations of weighted and unweighted benchmark scores are in parenthesis.

Table 2

Institutional Characteristics and Survey-Response Variables Associated with Differences in Weighted and Unweighted Benchmarks

Explanatory Variables	Beta Model 1	Beta Model 2
Undergraduate Enrollment (1,000s)	0.067	
Percent of Female Undergraduates	-0.119	
Percent of Full-Time Undergraduates	0.111	
Institutional Response Rate	-0.103	
Measure of Disproportional Response	0.614*	0.562*
Measure of Score Differences	0.444*	0.365*
Squared Multiple Correlation	0.556	0.495

* $p < 0.05$

Table 3

Predicted Differences in Weighted and Unweighted Benchmark Scores

Disproportional	Effect Size				
Response	0.10	0.15	0.20	0.25	0.30
0.10	0.61	0.90	1.19	1.49	1.78
0.15	0.76	1.05	1.34	1.64	1.93
0.20	0.91	1.20	1.50	1.79	2.08
0.25	1.06	1.35	1.65	1.94	2.23
0.30	1.21	1.50	1.80	2.09	2.38
0.35	1.36	1.66	1.95	2.24	2.53
0.40	1.51	1.81	2.10	2.39	2.68
0.45	1.66	1.96	2.25	2.54	2.83
0.50	1.82	2.11	2.40	2.69	2.98
0.55	1.97	2.26	2.55	2.84	3.13
0.60	2.12	2.41	2.70	2.99	3.28
0.65	2.27	2.56	2.85	3.14	3.43
0.70	2.42	2.71	3.00	3.29	3.58
0.75	2.57	2.86	3.15	3.44	3.74

Table 4

Variables Related to the Relative Precision of Weighted Benchmarks

Explanatory Variables	Beta Model 1	Beta Model 2
Measure of Disproportional Response	0.616*	0.531*
Measure of Score Differences	-0.179*	-0.205*
Less than 10 Respondents in a Cell	0.124	
Difference between Benchmark Score	-0.103	
Squared Multiple Correlations	0.320	0.300

* $p < 0.05$