

PROCESSING AND EDITING OF DATA

**4-1 Data Editing and Imputation of Item
Nonresponse**

4-2 Maintaining Confidentiality

4-3 Evaluation of Surveys

4-4 Nonresponse Bias Analysis

4-5 Sample Survey Weights

SUBJECT: DATA EDITING AND IMPUTATION OF ITEM NONRESPONSE

NCES STANDARD: 4-1

PURPOSE: To establish guidelines to reduce potential bias, ensure consistent estimates, and simplify analysis, by correcting inconsistent data in a data set (i.e., edits) or substituting values for missing (i.e., imputation) or inconsistent data in a data set (i.e., edits).

KEY TERMS: cross-sectional, cross-sectional imputations, cross-wave imputations, edit, freshened sample, imputation, item nonresponse, key variables, longitudinal, nonresponse bias, overall unit nonresponse, response rate, stage of data collection, unit nonresponse, and universe.

STANDARD 4-1-1: All data must be edited. Data editing is an iterative and interactive process that includes procedures for detecting and correcting errors in the data. Data editing is first done prior to imputation. Data editing must be repeated after the data are imputed, and again after the data are altered during disclosure risk analysis (without jeopardizing the disclosure protections). At each stage, the data must be checked for the following and edited if errors are detected:

1. Credibility, based on range checks to determine if all responses fall within a pre-specified reasonable range;
2. Consistency based on checks across variables within individual records for noncontradictory responses (i.e., no logical inconsistencies);
3. Incorrect flow through prescribed skip patterns;
4. Missing data that can be directly filled from other portions of an individual's record;
5. The omission and/or duplication of records;
6. Internal consistency across records, (e.g., the sum of categories matches the reported total); and
7. Inconsistency between estimates and outside sources.

GUIDELINE 4-1-1A: Editing should use available information and logical assumptions to derive substitute values for inconsistent values in a data file.

GUIDELINE 4-1-1B: When electronic data collection methods are used, data should be edited during, and if necessary after data collection.

GUIDELINE 4-1-1C: Possible actions when inconsistencies and other errors are found include the following:

1. Automated correction within specified criteria,
2. Data verified by respondent, and any automated edits overridden,
3. Corrected data provided by respondent,

4. Corrected data available from elsewhere in the respondent's answers,
5. Corrected data obtained from other sources.

STANDARD 4-1-2: Key variables in data sets used for cross-sectional estimates must be imputed for item nonresponse (beyond overall mean imputation). This applies to cross-sectional data sets and to data from longitudinal data sets that are used to produce cross-sectional estimates from the base year or subsequent freshened samples. (See [appendix A](#) for a discussion of alternative imputation procedures, including the pros and cons of specific approaches).

GUIDELINE 4-1-2A: In census (universe) data collections, it may not be appropriate to impute data in certain situations (e.g., peer analysis situations or when data for a particular establishment—school, university, or library—are being examined individually).

GUIDELINE 4-1-2B: When using non-NCES data sets, it is desirable to impute for missing data in those items being used in NCES publications. This is only appropriate when adequate auxiliary information is available.

GUIDELINE 4-1-2C: Imputation procedures should be internally consistent, based on theoretical and empirical considerations, appropriate for the analysis, and make use of the most relevant data available. If multivariate analysis is anticipated, care must be taken to use imputations that minimize the attenuation of underlying relationships. The Chief Statistician should review imputation plans prior to implementation.

STANDARD 4-1-3: In the case of longitudinal data sets, two imputation approaches are acceptable: cross-wave imputations or cross-sectional imputations. Cross-wave imputations may be used to complete missing data for longitudinal analysis or cross-sectional imputations may be used. ([Guideline 4-1-2C](#) of this Standard applies here, as well.)

STANDARD 4-1-4: In those cases where an item-level nonresponse bias analysis shows that the data are not missing at random, the amount of potential bias must inform the decision to retain or delete individual items (see [Standard 4-4](#)).

STANDARD 4-1-5: In cases where imputation is not used (e.g., items that are not key variables in either cross-sectional or longitudinal analysis), data tables must include a reference to a methodology table or glossary that shows the weighted item response rates for each unimputed variable included in the report (see [Standard 1-3-5](#) for the item response rate formula). For individual variables with item response rates less than 85 percent, the variable must be footnoted in the row or column header. The footnote must alert readers to the fact that the response rate is below 85 percent and that missing data have not been explicitly accounted for in the data.

STANDARD 4-1-6: When imputations are used, documentation indicating the weighted proportion of imputed data must be presented for all published estimates based on NCES data. Information about the amount of imputed data in the analysis can be included in the technical notes and does not have to accompany each table. The range of the amount of imputation used for the set of items included in an analysis must be reported. Also, the specific amount of imputation must be reported for items with response rates less than 70 percent. Items with response rates lower than 70 percent must be footnoted in the tables.

STANDARD 4-1-7: All imputed values on a data file must be clearly identified as such.

GUIDELINE 4-1-7A: Imputed data should be flagged in associated “flag” fields. The imputation method should be identified in the flag. Blanks are not legitimate values for flags.

STANDARD 4-1-8: If unimputed items are used in the estimation of totals or ratios (as in Standard 4-1-3 above), the risks of not using imputed data must be described.

1. Estimated totals using unimputed data implicitly impute a zero value for all missing data. These zero implicit imputations will mean that the estimates of totals will underestimate the true population totals. Thus, when reporting totals based on a unimputed item, the response rate for that item must be footnoted in the data table.
2. Ratios (averages) using unimputed data will implicitly impute the cell ratio for all missing data within the cell. This can cause inconsistencies in the estimates between tables.

SUBJECT: MAINTAINING CONFIDENTIALITY

NCES STANDARD: 4-2

PURPOSE: To protect the confidentiality of NCES data that contain information about individuals (individually identifiable information). For this reason, staff must be cognizant of the requirements of the law and must monitor the confidentiality of individually identifiable information in their daily activities and in the release of information to the public.

KEY TERMS: coarsening, confidentiality, confidentiality edits, on-line analysis tool, data swapping, edits, disclosure risk analysis, individually identifiable data, perturbation techniques, public-use data file, public-use edits, restricted-use data file, stage of data collection, and statistical disclosure techniques.

LEGAL REQUIREMENTS: Four laws cover protection of the confidentiality of individually identifiable information collected by NCES—the Privacy Act of 1974, as amended; the E-Government Act of 2002; the Education Sciences Reform Act of 2002; and the USA Patriot Act of 2001.

Privacy Act of 1974, as amended (5 U.S. Code, Section 552a)—“The purpose of this Act is to provide certain safeguards for an individual against invasion of personal privacy by requiring Federal agencies ...to collect, maintain, use or disseminate any record of identifiable personal information in a manner that assures that such action is for necessary and lawful purpose, that the information is current and accurate for its intended use, and that adequate safeguards are provided to prevent misuse of such information.” A willful disclosure of individually identifiable data is a misdemeanor, subject to a fine up to \$5,000.

E-Government Act of 2002, Confidential Information Protection and Statistical Proficiency Act (CIPSEA 2002) (44 U.S. Code, Section 3501, Subsections 501-513)—Under this law, all individually identifiable information supplied by individuals or institutions to a federal agency for statistical purposes under a pledge of confidentiality must be kept confidential and may only be used for statistical purposes. Any willful disclosure of such information for nonstatistical purposes, without the informed consent of the respondent, is a Class E felony.

Education Sciences Reform Act of 2002 (ESRA 2002) (20 U.S. Code, Section 9573)—Under this law, all individually identifiable information about students, their families, and their schools shall remain confidential. To this end, this law requires that no person may

- a. Use any individually identifiable information furnished under the provisions of this section for any purpose other than statistical purposes for which it is supplied, except in the case of terrorism (see discussion of the Patriot Act);
- b. Make any publication whereby the data furnished by any particular person under this section can be identified; or

- c. Permit anyone other than the individuals authorized by the Commissioner to examine the individual reports.

Further, individually identifiable information is immune from legal process, and shall not, without the consent of the individual concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding, except in the case of terrorism. Employees, including temporary employees, or other persons who have sworn to observe the limitations imposed by this law, who knowingly publish or communicate any individually identifiable information will be subject to fines of up to \$250,000, or up to 5 years in prison, or both (Class E felony).

USA Patriot Act of 2001(Public Law 107-56, Section 508-368)—This law permits the Attorney General to petition a court of competent jurisdiction for an ex parte order requiring the Secretary of the Department of Education to provide data relevant to an authorized investigation or prosecution of an offense concerning national or international terrorism. The law states that any data obtained by the Attorney General for these purposes “...may be used consistent with such guidelines as the Attorney General, after consultation with the Secretary, shall issue to protect confidentiality. This law was incorporated into ESRA 2002.

Federal Statistical Confidentiality Order of 1997 (Code of Federal Regulations, Vol. 62, No. 124, pgs 35044-35050)—This OMB Order provides a consistent government policy for “...protecting the privacy and confidentiality interests of persons who provide information for Federal statistical programs...” The Order defines relevant terms and provides guidance on the content of confidentiality pledges that Federal statistical programs should use under different conditions. The Order provides language for confidentiality pledges under two conditions—first, when the data may only be used for statistical purposes; second, when the data are collected exclusively for statistical purposes, but the agency is compelled by law to disclose the data. Since the USA Patriot Act of 2001 includes a legal requirement that compels NCES to share the data under the conditions specified in the law (see above), the second condition applies to NCES. In this case, the Order instructs the agency to “...at the time of collection, inform the respondents from whom the information is collected that such information may be used only for statistical purposes and may not be disclosed, or used, in identifiable form for any other purpose, unless otherwise compelled by law.”

STANDARD 4-2-1: All NCES staff, without exception, must pledge not to release any individually identifiable data, for any purpose, to any person not sworn to the preservation of confidentiality. Individually identifiable data are confidential and without the written consent of the individual, such data are protected from legal process, except in the case of the authorized investigation and prosecution of terrorism.

STANDARD 4-2-2: All contractors whose activities involve contact with individually identifiable information must provide NCES Project Officers with a list of all staff who will have contact with such data; all such staff must have a signed notarized affidavit of nondisclosure and documentation of a recent OPM midlevel security clearance or an active OPM application for such a clearance on file at NCES. These affidavits, clearance

documentation, and the staff list must be kept current as staff members leave and as new staff members are assigned to NCES projects that require access to individually identifiable information.

STANDARD 4-2-3: All contractor staff with access to individually identifiable information must only use that information for purposes associated with the data collection and analysis specified in the contract.

STANDARD 4-2-4: Respondents must be told in a cover letter or in instructions that “Your answers may be used only for statistical purposes and may not be disclosed, or used, in identifiable form for any other purpose except as required by law.” Furthermore, the routine statistical purposes for which the data may be used must be explained.

STANDARD 4-2-5: All materials having individually identifiable data must be kept secure at all times through the use of passwords, computer security, physical separation of individual identity from the rest of the data, and secure data handling and storage. (See the [Restricted-Use Data Procedures Manual, 2007](http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=96860rev), <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=96860rev>).

STANDARD 4-2-6: When confidentiality edits (that are performed using perturbation techniques) are used for a data file, they must be applied to all analytical files derived from that data file (e.g., public-use files, restricted-use files, and data files supporting on-line analysis tools).

STANDARD 4-2-7: NCES provides access to public-use and restricted-use data through on-line analysis tools that produce tabular estimates, and in some cases regression analysis. When restricted-use data are accessed through an on-line analysis tool, the following conditions must be met:

1. NCES, or any authorized users of NCES data, may not release the exact sample size for restricted-use data files that are distributed through an on-line analysis tool.
2. Only restricted-use data files with Disclosure Review Board (DRB) approved confidentiality edits may be used to produce an on-line analysis tool.
3. An on-line analysis tool may not publish unweighted counts.

The confidentiality protection required in an on-line analysis tool is a function of the type of estimate(s) to be produced. For example, an on-line analysis tool that produces weighted cell counts may require the use of more extensive confidentiality edits.

If a public-use file is released or planned for a data file, any on-line analysis tool created for that data file must be based on public-use data that have undergone perturbation disclosure limitation techniques as part of confidentiality edits.

STANDARD 4-2-8: For public-use data files, NCES minimizes the possibility of a user matching outliers or unique cases on the file with external (or auxiliary) data sources. Because public-use files allow direct access to individual records, perturbation and coarsening disclosure limitation techniques may both be required. The perturbation disclosure limitation techniques, by definition, include the techniques applied in a confidentiality edit (if one is performed) and may include additional perturbation disclosure limitation techniques as well.

Methods for protecting individually identifiable data

Type of protection	Methods	
	Perturbation	Coarsening
Confidentiality edit	Yes	No
Disclosure limitation techniques	Yes	Yes

All public-use files (i.e., the edited restricted-use files) that contain any potentially individually identifiable information must undergo a disclosure risk analysis in preparation for release to the public. The steps are as follows:

1. At an early stage in designing and conducting this analysis, staff must consult the Disclosure Review Board (DRB) for guidance on disclosure risk analysis and on the use of NCES disclosure risk software. Any modifications that are necessary as a result of the analysis must be made, and the entire process must be documented.
2. The documentation of the disclosure risk analysis must be submitted to the DRB. The documentation must include descriptions of the risk of disclosure and the types of edits used to avoid disclosure. Decisions over the type of confidentiality edits must take into account the procedures needed to avoid disclosure of individually identifiable information, age of the data, accessibility of external files, detail and specificity of the data, and reliability and completeness of any external files. The documentation should also include the results demonstrating the disclosure risk after adjustments to the data.
3. The DRB will review the disclosure risk analysis report and make a recommendation to the Commissioner of NCES about the file release.
4. The Commissioner then rules on the release of the data file.

STANDARD 4-2-9: Inasmuch as confidentiality edits are intended to protect individually identifiable data, files that incorporate the results of the DRB approved confidentiality edit plan may be used to produce tables without confidentiality concerns over minimum cell sizes. When this is done:

1. All versions of a data file must reflect the same confidentiality edits. Staff must consult the DRB on the confidentiality plan, data file dissemination plan (restricted, public use, and/or on-line data analysis system), and disclosure risk analysis plan, concurrently.
2. Documentation of the confidentiality edit must be included along with the documentation of the disclosure risk analysis that is submitted to the DRB.

STANDARD 4-2-10: A survey program may decide not to apply confidentiality edits (i.e., perturbation disclosure limitation techniques) to a restricted-use file (and the associated public-use file). In this situation, when tabulations are produced, any table with a cell with 1 or 2 unweighted cases must be recategorized to insure that each cell in the table has at least 3 unweighted cases. This restriction also applies to documentation for public-use files. This rule excludes table cells with zero cases because there are no data to protect in the cell.

EXAMPLE: A principal salary table by race and years of experience may only have 2 Asian respondents with more than 20 years of experience. To implement this standard, one possibility would be to either combine the Asian category with another race group or combine the 20+ years of experience category with the next lower experience category. This process would continue until all cells have either at least 3 unweighted cases or no unweighted cases.

STANDARD 4-2-11: At the discretion of the Director of IES, data security staff may release individually identifiable data to persons for statistical uses compatible with the purposes for which the data were collected. Persons receiving individually identifiable data from NCES shall execute a restricted-use data license agreement, sign affidavits of nondisclosure, and meet such other requirements as deemed necessary in accordance with other confidentiality provisions of the law.

STANDARD 4-2-12: The IES Confidentiality Statute prohibits the use of data released by IES in conjunction with any other information or technique, to identify any individual and to knowingly publish or use such information; subject to the penalties associated with a Class E Felony. Before external data users may gain access to public-use data files, they must agree that they will not use the data to attempt to identify any individual whose data are in the file. This may be accomplished by using the following wording:

“WARNING”

The National Center for Education Statistics (NCES) collects data for statistical purposes. Any effort to determine the identity of any individual in public-use data distributed by NCES is prohibited by law. Violations are subject to Class E felony charges of a fine up to \$250,000 and/or a prison term up to 5 years.

NCES does all it can to assure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might

lead to identification, are omitted or modified in the dataset to protect the true characteristics of individuals. Any intentional identification or disclosure of a person violates the assurances of confidentiality given to the providers of the information. Therefore, users shall:

- Make no use of the identity of any person discovered inadvertently, and advise NCES of any such discovery.
- Not link this dataset with individually identifiable data from other NCES or non-NCES datasets.
- Signify agreement to comply with the stated statutorily based requirements.

REFERENCE

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2007), *Restricted-Use Data Procedures Manual*.
<http://nces.ed.gov/pubs96/96860rev.pdf>

SUBJECT: EVALUATION OF SURVEYS

NCES STANDARD: 4-3

PURPOSE: To provide the necessary information for users of the survey data to understand the quality and limitations of the data and to provide information for planning future surveys or replications of the same survey. The evaluation should also include a systematic assessment of all sources of error for key statistics that will be studied or reported in NCES publications.

KEY TERMS: coverage error, edit, estimation, field test, frame, imputation, item nonresponse, key variables, longitudinal, nonsampling error, overcoverage, pretest, response rate, sampling error, stage of data collection, survey, survey system, undercoverage, unit nonresponse, and variance.

STANDARD 4-3-1: All proposed and ongoing surveys conducted by NCES must include an evaluation component in the survey design plan. The survey evaluation must include the following:

1. Range of potential sources of error;
2. Measurement of the magnitude of sampling error and sources of the various types of nonsampling error expected to be a problem;
3. Studies that identify factors associated with differential levels of error and assess procedures for reducing the magnitude of these errors;
4. Assessment of the quality of the final estimates, including comparisons to external sources, and where possible, comparisons to prior estimates from the same data collection; and
5. Technical report or series of technical reports summarizing results of evaluation studies; for example, a quality profile or total survey error model.

GUIDELINE 4-3-1A: Review past surveys similar to the one being planned to determine what statistical evaluation data have been collected in prior surveys and any potential problems that have been identified. Based on this review, prepare a written summary of what is known about the sources and magnitude of error.

GUIDELINE 4-3-1B: Indicate how each issue will be addressed, including the identification of required data internal and external to the study, a discussion of the comparisons that could be made, the experiments that may be built into the survey, and evaluation methods.

GUIDELINE 4-3-1C: Watch for additional problem areas arising during the course of survey administration and data processing and, where possible, collect and analyze appropriate data to assess the magnitude of the problem.

GUIDELINE 4-3-1D: Analyze data from the survey evaluation prior to or concurrent with the analysis of the survey data so that the results of the evaluation can be taken into account when processing, analyzing, and interpreting the study data.

GUIDELINE 4-3-1E: List 4-3-A may be used to help guide the development of evaluation plans during the survey planning stage and to develop a monitoring system for possible problems that may emerge during data collection and processing. The list identifies five categories of errors and enumerates potential sources of error within each category, methods to measure or evaluate them, and possible modifications for correcting them.

LIST 4-3-A: MEASURING AND EVALUATING ERROR

1. SAMPLE SELECTION, FRAMES AND COVERAGE—ADEQUACY OF FRAME

A. Sources of error:

1. Limitations of the frame—undercoverage/overcoverage of schools or institutions, duplicates, cases of unknown eligibility;
2. Listing error—failure of initial respondents to include or exclude prospective respondents per instruction; and
3. Selection of sampling units and respondent units within sampling units.

B. Evaluation of survey coverage—examples:

1. Comparison of estimated counts to reliable independent sources;
2. Matching studies to earlier versions of the same data source or to other data sources and the use of dual system estimation;
3. Analysis of survey returns for deaths, duplicates, changes in classification, and out-of-scope units; and
4. Field work, such as area listings.

C. Correcting for coverage error—examples:

1. Use a dual-frame approach for survey estimation; and
2. Employ post-stratification procedures.

2. MEASUREMENT ERRORS—DATA COLLECTION

A. Sources of error:

1. Questionnaire design, content, wording and instructions;
2. Length of reference period;
3. Interview mode(s);
4. Interviewers—characteristics, training, and supervision;
5. Respondent rules—self versus proxy respondents;
6. Use of records by respondents;
7. Other respondent effects;
8. Consistency and time-in-sample bias for longitudinal studies;
9. Responses to related multiple measures within a questionnaire;
10. Statistics derived for related measures from different questionnaires within a survey system; and
11. Responses to related measures from multiple respondents in a sampled unit (e.g., parent/student).

- B. Evaluation of measurement errors—examples:
 - 1. Pilot or field test survey and procedures;
 - 2. Cognitive research methods;
 - 3. Reinterview studies;
 - 4. Response variance;
 - 5. Randomized experiments;
 - 6. Behavior coding;
 - 7. Interviewer variance studies;
 - 8. Interviewer observation studies;
 - 9. Record check studies; and
 - 10. Comparisons of related measures within questionnaires, across respondents, and across questionnaires within a survey system.
- C. Correcting for measurement errors—examples:
 - 1. Use the results from a pilot or field test to modify questionnaire and/or procedures;
 - 2. Use input from cognitive research to modify questionnaire;
 - 3. Where possible, use results from comparisons of related measures; and
 - 4. Employ interviewer retraining and feedback.

3. DATA PREPARATION ERROR

- A. Sources of error:
 - 1. Pre-edit coding;
 - 2. Clerical review;
 - 3. Data entry; and
 - 4. Editing.
- B. Evaluation of processing errors—examples:
 - 1. Pre-edit coding;
 - 2. Clerical review verification;
 - 3. Data entry verification;
 - 4. Editing verification for manual edits;
 - 5. Edit rates;
 - 6. Coder error variance estimates; and
 - 7. Rating and scoring error variance estimates.

- C. Correcting for data preparation errors—examples:
 - 1. Resolution of differences identified in verification;
 - 2. Increased training;
 - 3. Feedback during rating and coding; and
 - 4. Edits for lack of internal agreement, where appropriate.

4. SAMPLING AND ESTIMATION ERRORS

- A. Sources of error:
 - 1. Weighting procedures;
 - 2. Imputation procedures; and
 - 3. Sample survey estimation and modeling procedures.
- B. Evaluation of sampling and estimation errors—examples:
 - 1. Variance estimation;
 - 2. Analysis of the choice of variance estimator;
 - 3. Indirect estimates for reporting sampling error—use of generalized variance functions, small area estimates, and regression models;
 - 4. Comparison of final design effects with estimated design effects used in survey planning;
 - 5. Analysis of the frequency of imputation and the initial and final distributions of variables; and
 - 6. Analysis of the effect of changes in data processing procedures on survey estimates.
- C. Correcting for estimation errors—examples:
 - 1. Re-estimation using alternative techniques (e.g., outlier treatments, imputation procedures, and variance estimation procedures); and
 - 2. Explore fitting survey distributions to known distributions from other sources to reduce variance and bias.

5. NONRESPONSE ERRORS

- A. Sources of error:
 - 1. Household/school/institution nonresponse;
 - 2. Person nonresponse; and
 - 3. Item nonresponse.
- B. Evaluation of nonresponse errors—examples (see [Standard 4-4](#)):
 - 1. Comparisons of respondents to known population characteristics from external sources;
 - 2. Comparisons of respondents and nonrespondents across subgroups on available sample frame characteristics or, in the case of item nonresponse, on available survey data;
 - 3. Comparisons of characteristics of early and late responding cases;
 - 4. Follow-up survey of nonrespondents for a reduced set of key variables to compare with data from respondents; and
 - 5. Descriptions of items not completed, patterns of partial nonresponse, and

characteristics of sampling units failing to respond to certain groups of characteristics.

- C. Correcting for nonresponse errors—examples (see Standards 3-2, 4-1, and 4-4):
 - 1. If response rates are low during initial phases of data collection and funds are not available for intensive follow-up of all respondents, take a random subsample of nonrespondents and use a more intensive data collection method;
 - 2. For ongoing surveys, use propensity models to identify nonrespondents that are not missing at random and target respondents with the identified characteristics in subsequent administrations.
 - 3. Use nonresponse weight adjustments for unit nonresponse; and
 - 4. Use item imputations for item nonresponse.
- D. Methods for reducing nonresponse—examples (see Standards 3-2, 4-1, and 4-4):
 - 1. Employ pretest or embedded experiments to determine the efficacy of incentives to improve response rates;
 - 2. Use internal reporting systems to monitor nonresponse during collection;
 - 3. Use follow-up strategies for nonrespondents to encourage participation;
 - 4. Target a set of key data items for collection with unwilling respondents; and
 - 5. For ongoing surveys, consider separate research studies to examine alternative methods of improving response rates.

REFERENCES

- Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., and Sudman, S. (eds.). (1991). *Measurement Errors in Surveys*. New York: John Wiley & Sons.
- Bradburn, N.M., Sudman, S., and Associates. (1979). *Improving Interviewing Methods and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*. San Francisco: Jossey-Bass.
- Brick, J.M., and Kalton, G. (1996). Handling Missing Data in Survey Research. *Statistical Methodology in Medical Research*, 5: 215-238.
- Brick, J.M., Cahalan, M., Gray, L., and Severynse, J. (1994). *A Study of Selected Nonsampling Errors in the 1991 Survey of Recent College Graduates* (NCES 95-640). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Federal Committee on Statistical Methodology. (1978). *An Error Profile: Employment as Measured by the Current Population Survey*. Statistical Policy Working Paper 3. Washington, DC: U.S. Office of Management and Budget.
- Federal Committee on Statistical Methodology. (1988). *Quality in Establishment Surveys*. Statistical Policy Working Paper 15. Washington, DC: U.S. Office of Management and Budget.
- Federal Committee on Statistical Methodology. (1990). *Survey Coverage* (Statistical Policy Working Paper 17). Washington, DC: U.S. Office of Management and Budget.

- Forsman, G., and Schreiner, I. (1991). The Design and Analysis of Reinterview: An Overview. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. New York: John Wiley & Sons, 279-302.
- Gonzalez, M.E., Ogus, J.L., Shapiro, G., and Tepping, B.J. (1975). Standards for Discussion and Presentation of Errors in Survey and Census Data. *Journal of the American Statistical Association*, 70: 351, Part II.
- Groves, R.M., and Couper, M.P. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R.M., Biemer, P.P., Lyberg, J.E., Massey, J.T., Nicholls II, W.L., and Waksberg, J. (Eds.). (1988). *Telephone Methods for Surveys*. New York: Wiley.
- Jabine, T.B., King, K.E., and Petroni, R.J. (1990). *Survey of Income and Program Participation Quality Profile*. Washington, DC: U.S. Department of Commerce, Bureau of the Census.
- Kalton, G. (1998). *Survey of Income and Program Participation (SIPP) Quality Profile*. 3rd Edition. Working Paper 230. Washington, DC: U.S. Bureau of the Census.
- Kulka, R. (1995). The Use of Incentives to Survey "Hard-to-Reach" Respondents: A Brief Review of Empirical Research and Current Research Practices. *Seminar on New Directions in Statistical Methodology*. Statistical Policy Working Paper 23, pp. 256-299. Washington, DC: U.S. Office of Management and Budget.
- Lessler, J.T., and Kalsbeek, W.D. (1992). *Nonsampling Error in Surveys*. New York: John Wiley & Sons.
- Lyberg, L., and Dean, P. (1992). Methods for Reducing Nonresponse Rates: A Review. Paper prepared for presentation at the 1992 meeting of the American Association for Public Opinion Research.
- Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (Eds.). (1998). *Survey Measurement and Process Quality*. New York: John Wiley & Sons.
- Paxson, M.C., Dillman, D. and Tarnai, J. (1995). Improving Response to Business Mail Surveys. In Cox et al. (Eds.) *Business Survey Methods*. New York: John Wiley & Sons.
- Salvucci, S., Walter, E., Conley, V., Fink, S., and Saba, M. (1997). *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Sudman, S., and Bradburn, N. (1974). *Response Effects in Surveys: A Review and Synthesis*. Chicago, IL: Aldine.
- Sudman, S., Bradburn, N., and Schwarz, N. (1996). *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.

U.S. Bureau of the Census and U.S. Bureau of Labor Statistics. (2000). *The Current Population Survey: Design and Methodology*. Technical Paper 63. Washington, DC.

United Nations. (1982). *National Household Survey Capability Programme, Non-Sampling Errors in Household Surveys: Sources, Assessment and Control*. New York: United Nations Department of Technical Cooperation for Development and Statistical Office (preliminary version).

SUBJECT: NONRESPONSE BIAS ANALYSIS

NCES STANDARD: 4-4

PURPOSE: To identify the existence of potential bias due to unit and item nonresponse.

KEY TERMS: base weight, frame, item nonresponse, nonresponse bias, overall unit nonresponse, potential magnitude of nonresponse bias, required response items, response rate, stage of data collection, survey, total nonresponse, unit nonresponse, and wave.

STANDARD 4-4-1: A nonresponse bias analysis is recommended whenever unit or item nonresponse is present to ensure that the estimates based on reported cases are not biased by the missing respondents. Any survey stage of data collection with a unit or item response rate less than 85 percent must be evaluated for the potential magnitude of nonresponse bias before the data or any analysis using the data may be released. (For definitions and calculation formulas see Standard 1-3-2 for unit response rates and Standard 1-3-5 for item response rates.) Estimates of survey characteristics for nonrespondents and respondents are required to assess the potential nonresponse bias. The level of effort required is guided by the magnitude of the nonresponse.

STANDARD 4-4-2: When unit nonresponse is high, nonresponse bias analysis must be conducted at the unit level to determine whether or not the data are missing at random and to assess the potential magnitude of unit nonresponse bias. At the unit level, the nonresponse bias analysis must be conducted using base weights for the survey stage with nonresponse. The following guidelines must be considered in such analysis.

GUIDELINE 4-4-2A: Comparisons of respondents and nonrespondents across subgroups using available sample frame characteristics provide information about the presence of nonresponse bias. This approach is limited because observed frame characteristics are often unrelated or weakly related to more substantive items in the survey.

GUIDELINE 4-4-2B: Formal multivariate modeling can be used to compare the proportional distribution of characteristics of respondents and nonrespondents to determine if nonresponse bias exists and, if so, to estimate the magnitude of the bias. These multivariate analyses are used to identify the characteristics of cases least likely to respond to an interview (such analyses are often referred to as nonresponse propensity models). Cases are coded as either responding to or not responding to the interviews, and multivariate techniques are used to identify which case characteristics significantly relate to unit nonresponse. The predictor variables should have very high response rates. The predictors should include all available information about respondents and their schools or institutions. For those respondents who are not missing at random, the results of detailed propensity models can be used to inform the level of effort to be targeted at similar respondents in future data collections. In the case of ongoing data collections, earlier rounds of data collection can be used to identify possible predictive variables for respondents with shared characteristics with

current nonrespondents. This information can also be used to improve nonresponse weight adjustments or imputation models. This approach may be limited by the extent to which such predictors exist in the data.

GUIDELINE 4-4-2C: Comparisons of respondents to known population characteristics from external sources can provide information about how the respondents differ from a known population. This approach is limited by information available from existing sources on the population of interest. Known population characteristics are often unrelated or weakly related to more substantive items in the survey.

GUIDELINE 4-4-2D: For collections in which successive levels of effort (e.g., increasing number of contact attempts, increasing incentives to respond) are employed to reduce nonresponse, comparisons of characteristics can be made between the later/more difficult cases and the earlier/easier cases to estimate the characteristics of the remaining nonrespondents. This approach may be less effective if overall or total response rates are relatively low or if a collection period is relatively short in duration. In addition, the assumption that nonrespondents are like those respondents who are difficult to reach may not hold.

GUIDELINE 4-4-2E: More intensive methods and/or incentives can be used to conduct a followup survey of nonrespondents on a reduced set of required response items. Comparisons between the nonrespondent follow-up survey and the original survey can be made to measure the potential magnitude of nonresponse bias in the original survey. This approach may be costly and less useful for modeling nonresponse bias if the nonrespondent follow-up survey response rates are also below 70 percent.

GUIDELINE 4-4-2F: The estimated bias can be summarized using the following measures. One measure is the ratio of the bias to the standard error, using the base weight. A second measure is the ratio of the bias to the reported survey mean, using the base weight. If weighting adjustments are used to reduce bias, these measures should also be reported using the final weighted estimates.

STANDARD 4-4-3: When item nonresponse is high, nonresponse bias analysis must be conducted at the item level to determine whether or not the data are missing at random and to assess the potential magnitude of item nonresponse. To analyze potential bias from item nonresponse, the guidelines below must be considered.

GUIDELINE 4-4-3A: For an item with a low total response rate, respondents and nonrespondents can be compared on sampling frame and/or questionnaire variables for which data on respondents and nonrespondents are available. Base weights must be used in such analysis. Comparison items should have very high response rates. A full range of available items should be used for these comparisons. This approach may be limited to the extent that items available for respondents and nonrespondents may not be related to the low response rate item being analyzed.

GUIDELINE 4-4-3B: Formal multivariate modeling can be used to compare characteristics of respondents and nonrespondents to determine if nonresponse bias exists and, if so, to estimate the magnitude of the bias. These multivariate analyses are used to identify the characteristics of cases least likely to respond to an item (such analyses are often referred to as nonresponse propensity models). Cases are coded as either responding to or not responding to the item and multivariate techniques are used to identify which case characteristics significantly relate to item nonresponse. Base weights must be used in such analysis. The predictor variables should have very high response rates. The predictors should include the full range of available information about respondents and their schools or institutions. This approach may be limited by the extent to which such predictors exist in the data.

GUIDELINE 4-4-3C: If the overall response rate is acceptable, item level nonresponse bias analysis may be conducted using data from survey respondents only. Unit-level respondents who answered the low response rate item can be compared to unit-level respondents who did not answer the item. Final weights and unimputed variables should be used in such an analysis. The comparison items should have very high item response rates. This approach may be limited because it does not directly analyze nonresponse bias that may originate because of unit-level nonresponse.

SUBJECT: SAMPLE SURVEY WEIGHTS

NCES STANDARD: 4-5

PURPOSE: To ensure that sample survey data are used to make appropriate inferences about the population from which the sample was drawn. Final weights are constructed as the product of several component weights. The weights reflect the differences between a simple random sampling design and a complex sampling design and between the field results and the sampling design. The weights may also compensate for imperfections in the sampling frame. Weights may also be used to improve the accuracy and/or analytic power of the data.

KEY TERMS: balanced repeated replication, bootstrap, cohort, cluster sampling, freshened sample, half open interval, jackknife, longitudinal, measure of size, multistage sampling, non-response, post-stratification adjustments, probability of selection, trimming, probabilities proportionate to size (PPS), raking, realized sample, strata, stratification, and weight.

STANDARD 4-5-1: Construct weights appropriate for the sample design (e.g., unequal probabilities of selection, stratification, clustering, cross-sectional or longitudinal cohort study) and realized sample (e.g., account for unit nonresponse and adjustments such as poststratification, and raking to known population totals) to allow inferences to the target population. Multistage sample weights must account for all stages of the sample selection process.

GUIDELINE 4-5-1A: For a stratified sample design, sample weights should be the inverse of the probabilities of selection of each sampling unit, taking into account all stages of the sample selection process. For equal probability sampling, the initial (base) weight is determined within each stratum by calculating the ratio of the number of sampling units available on the sampling frame to the number of sampling units selected within each stratum. For probability proportionate to size sampling (PPS) primary sampling units are selected with probabilities proportionate to a measure of size (often used instead of exact sizes N of the elements in a particular strata).

GUIDELINE 4-5-1B: Adjust the weights for unit non-response to minimize bias arising from differences between responding and nonresponding units. These adjustments should be internally consistent, based on theoretical and empirical considerations, appropriate for the analysis, and based on the most relevant data available. Appropriate methods include ratio adjustments based on weighting adjustment classes (or cells) identified by using statistical algorithms (e.g., CHAID: chi-square automatic interaction detector or other statistical procedures such as logistic response propensity modeling).

GUIDELINE 4-5-1C: The nonresponse-adjusted weights can be adjusted further to calibrate the sample to known population totals. For example, poststratification - dividing the population into subgroups such as age/race/sex subgroups - can be used to

adjust weights so that the sample estimates of the subpopulations agree with known or previously derived estimates. This reduces sampling variability and biases resulting from undercoverage. To be effective, the variables that define the poststrata should be correlated with the variables of interest, be well measured in the survey, and control totals must be available for the population as a whole.

GUIDELINE 4-5-1D: As a final weighting step, adjustments such as weight trimming should be made to avoid extremely large weights. (Weight trimming probably should be carried out before post-stratification to ensure sample estimates of the subpopulations agree with known or previously derived estimates).

GUIDELINE 4-5-1E: If analysis of unrepresentative samples is possible from the survey data it can be inappropriate to construct separate weights for every survey component employed in a data collection system (e.g., in a student survey, to create separate weights to adjust for nonresponse in the teacher, parent or transcript components). Determination of the need for creating weights should be based on the pre-determined population which the survey estimates are to represent.

GUIDELINE 4-5-1F: A common technique for analyzing survey data from complex sample designs is the creation of replicate weights that are added to the data file and used to produce estimates of variance that correctly reflect the complex sampling variation. Three such methods used by NCES are balanced repeated replication weighting, jackknife and bootstrap weighting (see [Standard 5.2](#) and [Guideline 5-2-3](#) for more detail on replicate weights).

STANDARD 4-5-2: Weights for longitudinal surveys are more complex than those for cross-sectional surveys, as data are collected from the same sampling units at different time periods. The number and types of weights constructed should reflect the primary unit of analysis and the number of waves (collection periods).

GUIDELINE 4-5-2A: In a multi-wave longitudinal sample survey, cross-sectional weights should reflect the separate waves (collection periods). In addition, for longitudinal analyses between two waves each panel record should be assigned a panel weight or weights to be used for intracohort analysis (within a panel survey cohort) and cross-cohort analysis (between two different panel surveys). The number of panel weights will be determined by the study's design and purpose.

GUIDELINE 4-5-2B: Weighting of a freshened sample in a longitudinal survey must reflect the sampling procedure employed to add new sampling units to the survey. (Typically NCES longitudinal surveys have used the half open interval sampling technique). Also, weights for the freshened sample units should reflect only those waves of data collection in which they have participated.

STANDARD 4-5-3: Final weighting plan specifications, including calculations for how the final weights were derived, and justifications for these calculations must be documented ([Standard 3.4](#)).

GUIDELINE 4-5-3A: The weighting specifications should include a description of the sampling frame and sampling design. (For example, in a two-stage sampling design, schools are selected in the first stage of data collection. In the second stage, students, teachers or some other unit may be selected). Each stage should be described.

GUIDELINE 4-5-3B: When describing the weights, the plan should identify the primary unit of analysis (e.g., students) and the periodicity of the target population.

GUIDELINE 4-5-3C: All final weights constructed should be described including their statistical characteristics as well as their function in analysis of the survey data (e.g. the different cross-sectional and panel weights for a longitudinal survey).

REFERENCES

Cochran, William G. (1977). *Sampling Techniques*. John Wiley and Sons.

Deaton, Angus (1997). *The Analysis of Household Surveys*. Johns Hopkins University Press, World Bank.

Kish, Leslie (1965). *Survey Sampling*. John Wiley and Sons.

Edward L. Korn, Barry I. Graubard "Examples of Differing Weighted and Unweighted Estimates from a Sample Survey" *The American Statistician*, Vol. 49, 1995.

1994 Lohr, S. and Liu, J. "A comparison of weighted and unweighted analyses in the NCVS," *Journal of Quantitative Criminology*, 10, 343-360.