

## **COLLECTION OF DATA**

**3-1 Coverage for Frames and Samples**

**3-2 Achieving Acceptable Response Rates**

**3-3 Monitoring and Documenting Survey Contracts**

**3-4 Documenting a Survey System**

## **SUBJECT: COVERAGE FOR NCES FRAMES AND SAMPLES**

### **NCES STANDARD: 3-1**

**PURPOSE:** To ensure that necessary steps are taken to develop and maintain data collections that are used as sampling frames and that coverage of sampling frames is evaluated and documented.

**KEY TERMS:** capture/recapture, confidentiality, coverage, coverage error, dual-frame estimation, estimation, frame, frame population, freshening, half-open interval, multiplicity estimation, noncoverage, over coverage, supplemental area frame, survey, survey system, target population, under coverage, un-duplication, and variance.

**STANDARD 3-1-1:** Staff responsible for NCES data collections that serve as sampling frames for other NCES surveys must evaluate the coverage of the frame and document coverage rates at least once every 5 years.

**GUIDELINE 3-1-1A:** Frames can be retrospectively compared against alternative frames found inside and outside of the Department of Education, considering total list count comparisons, matching operations, and dual-frame estimation procedures using capture/recapture procedures to estimate noncoverage, and provide an estimation of missing units.

**GUIDELINE 3-1-1B:** Staff responsible for NCES data collections that are used as sampling frames should maintain two-way communications with survey staff who use their collection as a frame. Procedures such as sharing preliminary data files with survey staff in order to develop frames may be instituted. (For example, staff who use an administrative list of public schools for their frame should be alerted when new data are available and each time there is a major change in the list.)

**STANDARD 3-1-2:** NCES data collections that are used as sampling frames for other NCES surveys must strive for coverage rates in excess of 95 percent overall and for each major stratum.

**STANDARD 3-1-3:** Staff using NCES frames for sample surveys must be cognizant of coverage issues and must take the steps necessary to provide satisfactory coverage for the sample survey. If there is not evidence of a coverage rate of at least 85 percent of the target population, then frame enhancements such as frame supplementation or dual-frame estimation must be incorporated into the survey study design.

**GUIDELINE 3-1-3A:** The first time a survey is conducted, background design and coverage work should be done before choosing the frame. Alternative frames, if applicable, should be considered and compared.

**GUIDELINE 3-1-3B:** Coverage errors such as over- and under-coverage, bad contact information, classification, temporal errors, and other listing errors should be minimized before the use of a frame. Techniques such as list supplements, multiplicity estimation, half-open intervals, and un-duplication can be used to reduce these errors and improve coverage of the frame.

**GUIDELINE 3-1-3C:** Any possible changes to frame variables identified by sample survey staff should be reported to the staff responsible for the data collection being used as the frame. For example, the relevant variables to maintain and consider include (1) eligibility (e.g., grade span); (2) contact information (e.g., name, address, and phone number); (3) classification variables (e.g., state and school level); and (4) measures of size (e.g., grade enrollment).

**GUIDELINE 3-1-3D:** To reduce coverage error, whenever a frame has important deficiencies with respect to the measurement unit, dual-frame estimation should be considered to correct these deficiencies. The increased cost associated with dual-frame estimation and the effect dual-frame estimation has on increasing the variance estimates should also be considered when deciding to use dual-frame estimation.

**STANDARD 3-1-4:** For each sample survey, a description of the frame and its coverage must be included in the survey documentation. This description must include, but is not limited to, the target and frame populations (and exclusions thereof); the name and date of the data collection which provided the original frame; any supplementing done to the original frame; limitations of the frame including the timeliness of the frame; and, if applicable, an estimation of the missing units on the frame.

**GUIDELINE 3-1-4A:** Sample survey documentation should include a discussion of coverage issues such as alternative frames that were considered, what was done to improve the coverage of the frame, and how data quality and item nonresponse on the frame may have affected the coverage of the frame.

**GUIDELINE 3-1-4B:** Survey documentation should include any estimation techniques used to improve the coverage of estimates. This would include post-stratification procedures. (For example, a telephone survey could post-stratify estimates of all individuals to account for the exclusion of those without telephones.)

**GUIDELINE 3-1-4C:** NCES survey staff should archive their survey's sampling frames as part of the documentation of the survey system found in [Standard 3-4](#), taking security precautions consistent with confidentiality laws into account. This archiving may be particularly important if a preliminary file was used to develop the frame, or if there is a chance that the frame may be used in the future to further develop research questions.

**STANDARD 3-1-5:** NCES survey staff that use NCES data collections as a frame must share any coverage or usage issues with the NCES data collection staff so that the coverage can be improved for future uses. This standard is related to Guideline 3-1-3B and Guideline 3-1-3C. (For example, after the survey is complete, the survey staff should provide a memo to the NCES data collection staff for the data collection used as a frame, reviewing the major limitations of the coverage or the data quality issues identified.)

**SUBJECT: ACHIEVING ACCEPTABLE RESPONSE**

**RATES NCES STANDARD: 3-2**

**PURPOSE:** To ensure that data collection programs conducted by or for NCES are conducted in a manner that protects the rights of survey respondents to fair treatment and privacy, while at the same time encouraging high rates of response across all strata, since high response rates help ensure that results are representative of the target population.

**KEY TERMS:** confidentiality, imputation, item nonresponse, longitudinal, nonresponse bias, pretest, required response items, response rate, strata, survey, and target population.

**STANDARD 3-2-1:** The data collection must be designed and administered in a manner that protects the rights of the survey respondents, while encouraging respondents to participate.

**GUIDELINE 3-2-1A:** The method of data collection (e.g., mail, telephone, Internet) should be appropriate for the target population and the objectives of the data collection.

**GUIDELINE 3-2-1B:** The data should be collected at the most appropriate time of year.

**GUIDELINE 3-2-1C:** The data collection period should be of adequate and reasonable length to achieve good response rates.

**GUIDELINE 3-2-1D:** When appropriate, respondent incentives should be considered. Propensity models can be used to identify those respondents who are more and less likely to respond; this information can be combined with the use of incentives to increase participation among those cases the least likely to respond. Additional consideration should be given to the use of incentives among the likely nonrespondents who are the most unlike the responding sample.

**STANDARD 3-2-2:** An explanation of the need for data, the goals and objectives of the data collection, the respondent burden, and examples of uses of the data that benefit respondents must be provided to the respondent (Privacy Act of 1974, as amended, 5 U.S.C. 552a).

**GUIDELINE 3-2-2A:** Informational materials for respondents may include a pre-notification letter, brochure, set of questions and answers about the survey, and/or an 800 number to call for more information. These materials describing the data collection should be sent to respondents in advance, when possible.

**GUIDELINE 3-2-2B:** For interviewer-administered data collection programs, training should emphasize techniques for obtaining respondent cooperation and techniques for building rapport with respondents, including respect for respondents' rights, manner, follow-up skills, knowledge of the goals and objectives of the data collection, and knowledge of the uses of the data. Response propensity models can be used to increase field staff awareness of the cases that are most likely to be hard to reach, and at times to identify characteristics of those cases that may be used to target contacts on those times that are most likely to yield cooperation.

**GUIDELINE 3-2-2C:** Prior to conducting a data collection program, endorsements, support, and the active cooperation of interested groups, such as professional organizations, professional associations, education community leaders, and state and local school district officials, should be obtained and communicated to respondents.

**STANDARD 3-2-3:** All NCES data collections must provide information concerning the confidentiality of responses. Privacy and nondisclosure assurances citing the appropriate legislation must be provided, as applicable (see Standard 4-2).

**STANDARD 3-2-4:** In keeping with the goals of the particular data collection effort, respondent burden must be minimized, as required by the Office of Management and Budget clearance process. The following steps aid in minimizing respondent burden, and they minimize item nonresponse and other measurement error:

**GUIDELINE 3-2-4A:** The questionnaire should be pretested to ensure that the established recall periods are reasonable.

**GUIDELINE 3-2-4B:** The questionnaire should be pretested for the difficulty and interpretability of questions.

**GUIDELINE 3-2-4C:** The questionnaire should be pretested for ease in navigation of self-administered questionnaires.

**GUIDELINE 3-2-4D:** Questions should be clearly written and skip patterns easily followed.

**GUIDELINE 3-2-4E:** The questionnaire should be of reasonable length.

**STANDARD 3-2-5:** All data collection programs require some follow-up of nonrespondents to achieve desirable response rates. Follow-up strategies designed to protect the respondents' rights, while achieving acceptable response rates, must be included in the data collection plan.

**GUIDELINE 3-2-5A:** Plan an adequate number of contact attempts.

**GUIDELINE 3-2-5B:** Internal reporting systems that provide timely reporting of response rates and the reasons for nonresponse throughout the data collection should be developed. These systems should be flexible enough to identify important subgroups with low response rates for more intensive follow-ups. Similarly, response propensity models from related data collections may be used to target important subgroups with low response rates and unique characteristics for more intensive follow-ups.

**GUIDELINE 3-2-5C:** For longitudinal surveys, provide appropriate confidentiality assurances, while obtaining as much locating information about respondents as possible during initial contact (e.g., for a student, school address, home address, name of advisor, phone numbers of parents).

**GUIDELINE 3-2-5D:** If response rates are low after the initial phases of data collection, and if further data collection on the full sample is deemed too costly, take a random subsample of nonrespondents and use a more intensive data collection method. This subsample will permit a description of nonrespondents' characteristics, provide data needed for nonresponse bias analysis, and allow for possible weight adjustments or for imputation of missing characteristics.

**GUIDELINE 3-2-5E:** Determine a set of required response items to obtain when a respondent is unwilling to fully cooperate. These items may then be targeted in follow-up to meet the minimum standard for unit response. These items may also be used in a nonresponse bias analysis that compares characteristics of respondents and nonrespondents using the sample data for those items. These required response items may also be used for item nonresponse imputation systems.

## **SUBJECT: MONITORING AND DOCUMENTING SURVEY CONTRACTS**

### **NCES STANDARD: 3-3**

**PURPOSE:** To assist NCES staff in monitoring and documenting survey contract activities.

**KEY TERMS:** edit, estimation, imputation, response rate, stage of data collection, survey, survey system, and variance.

**STANDARD 3-3-1:** The Contracting Officer's Representative (COR) must work to ensure that the contractor meets (a) contract specifications, (b) contract schedules, (c) NCES standards, (d) performance cost controls, and (e) beneficial effort/method of performance criteria in fulfilling the contract. Education Department Directive OCFO:2-108 dated 08/06/2009 includes instructions for *Contracting Monitoring for Program Officials*.

In some instances, the contractor may request technical redirection for unanticipated problems. For simple matters that are clearly within the scope of the contract, such requests may be made verbally. For problems that may require a change in scope, all requests must be in writing and outline the issue(s) and potential options. The COR must use this information in discussions with other NCES senior management in determining the appropriate course of action. All changes in any contract scope of work require action by the Contracting Officer. Whatever course of action is taken, it must be documented and placed in the project files.

**GUIDELINE 3-3-1A:** The COR should maintain close communication with the contractor. Depending on the nature of the survey, the COR should maintain communication through the use of meetings, phone calls, e-mails, visits, and/or an electronic management information system (MIS) for the purpose of tracking and monitoring the progress of the survey.

**GUIDELINE 3-3-1B:** The status of each unit of observation should be kept current and available to the COR at each stage of the data collection process. Critical status events may include, but are not limited to, dates of questionnaire mail out, returns, deletions (out-of-scopes), scan editing, data entry, machine editing, callback(s), and addition to the final data files. The COR should have direct and rapid access to this information.

**GUIDELINE 3-3-1C:** To help decide whether any adjustments or corrective actions are needed, soon after initial startup of field operations, and less frequently thereafter, the COR should evaluate the quality of survey operations by comparing a sample of the original returned questionnaires with the information on the data file for the following purposes:

1. Detect any data processing errors;
2. Learn of any problems with reporting or questionnaire design; and
3. Ensure that editing/update procedures are being correctly implemented.



**GUIDELINE 3-3-1D:** On an as-needed basis, CORs may request a copy or subsample of “completed” records from the current master file and analyze the information for conformance to contract requirements. The extent of the statistical analysis of this subsample should vary with survey objectives. Simple cross-tabulations and frequencies of discrete variables should normally point out internal coding inconsistencies and also provide interim item response rates. Simple descriptive statistics for continuous variables should provide interim item response rates, measures of dispersion, and outliers.

**GUIDELINE 3-3-1E:** The COR should ensure that software used for weighting, imputations, and variance estimation is accurate. This may be done through a series of "checkpoints" imbedded within the program(s). Another alternative is to have the contractor provide printouts from a series of discrete steps, with review by the COR along the way.

**GUIDELINE 3-3-1F:** The COR should keep the CO and NCES management informed of the result of reviews. As an integral part of this work, the COR should offer recommendations for solving any problems, acceptance of deliverables, performance awards, and approval or disapproval of any proposed changes.

**STANDARD 3-3-2:** The COR must review and verify progress reports, invoices, technical products and documentation, written correspondence, and other documents for the following purposes:

1. Ensuring the instruments are correct by reviewing final printed version to be used in the field prior to the fielding of the survey. Alternatively, if the survey is electronic (i.e., CATI, CAPI, or web-based), ensuring the instruments are correct by reviewing the final computer-based instrument.
2. Monitoring adherence to project schedules and requirements;
3. Assuring deliverables meet NCES standards and comply with the conditions of the contract and other quality requirements (e.g., accuracy and completeness);
4. Identifying potential problems that would substantially affect the successful completion of the survey or alter the terms and conditions of the contract (e.g., cost or time increases, quality decreases); and
5. Monitoring interviewers to minimize interviewer falsification of records (e.g., through observation, telephone monitoring of interviews, or computer programs used to detect falsifications)

**STANDARD 3-3-3:** The COR must maintain the following documents in the COR contract file: (a) progress reports, (b) vouchers, and (c) deliverables as required by the contract. Together with the RFP, contract proposal, proposal evaluation, and signed contract, these documents are subject to audit. Also, the COR must document any modifications or changes in (a) key personnel, (b) project schedule, (c) deliverables, and (d) scope of work, and their implications for the project completion date, deliverables, and costs.

**GUIDELINE 3-3-3A:** It is advisable to include in the contract file all correspondence, such as logs of phone conversations, e-mail and written correspondence, and documentation, describing the approval of or decisions made regarding changes.

**GUIDELINE 3-3-3B:** The COR should keep accurate and complete records of contractor performance, such as lateness, unacceptable deliverables, and cost overrun. Actions or decisions taken by the COR or CO to remedy the problems should also be clearly documented.

**STANDARD 3-3-4:** CORs should require that all computer programs (software) are self-documenting. All contracts must define the ownership of computer programs (software) and computer programming used for the creation of data files, data cleaning, data analysis, etc. All computer programming created using government funding shall be considered property of the U.S. government.

**GUIDELINE 3-3-4A:** The programmer should insert "comments" within the program(s) to describe each discrete section of code. Relationships between programs and data files should be flowcharted or described in a separate document. This includes record layouts and file structures.

**STANDARD 3-3-5:** Upon completion and/or termination of the contract, the COR must archive those items specified in Standard 3-4 for Documenting a Survey System and Standard 7-2 for Survey Documentation in Reports.

## **SUBJECT: DOCUMENTING A SURVEY SYSTEM**

### **NCES STANDARD: 3-4**

**PURPOSE:** To ensure that complete documentation is kept on NCES surveys and survey systems and their associated contract deliverables. Documentation includes those materials necessary to understand how to properly analyze data from each survey, as well as the information necessary to replicate and evaluate each survey. In addition, survey system documentation includes information necessary to design and estimate resource requirements of future similar surveys.

**KEY TERMS:** construct validity, coverage, edit, facsimile, frame, imputation, instrument, nonsampling error, public-use data file, response rate, sampling error, strata, survey, survey system, and variance.

**STANDARD 3-4-1.** Survey system documentation must include all information necessary to properly analyze the data. This information shall, at a minimum, include the following:

1. Final data set(s);Final instrument(s) or a facsimile thereof;
2. Definitions of all variables, including all modifications (e.g., truncation, trimming, coding, collapsing);
3. Data file layout;
4. Descriptions of constructed variables on the data file that are computed from responses to other variables on the file;
5. Description of variables used to uniquely identify cases in the data file;
6. Description of sample weights and how to apply them;
7. Description of the strata and primary sampling unit (PSU) identifiers to be used for analysis;
8. Description of how to calculate variances appropriate for the survey design;
9. Description of all editing and imputation methods applied to the data and how to remove imputed values from the data;
10. General description of disclosure avoidance techniques;
11. Descriptions of restrictions on the use of data (e.g., limited to statistical uses; when using restricted data files, all unweighted counts included in any release must be rounded);
12. Descriptions of known data anomalies and corrective actions; and
13. Unweighted frequency counts (if there is an on-line analysis tool supported by a restricted use file, only internal staff and NCES restricted-use data licensees who are authorized to use the specific data set may have access to these unweighted frequency counts).

**GUIDELINE 3-4-1A:** If the data are collected through a web-based collection or through a CATI or CAPI interview, the following information should be included in the documentation of the final instruments:

1. All items in the instrument (e.g., questions, check items, and help screens);
2. Items extracted from other data files to pre-fill the instrument (e.g., dependent data from a prior round of interviewing); and
3. Items that are input to the post data collection processing steps (e.g., output of an automated instrument).

**STANDARD 3-4-2:** To insure that a survey can be replicated and properly evaluated, the survey system documentation must, at a minimum, include the following:

1. Justifications for the items on the survey instrument, including how the final items were selected (including, for example, statistics on construct validity of individual items or sets or items that form scales or tests);
2. All instructions to respondents and/or interviewers either about how to properly respond to a survey item or how to properly present a survey item;
3. Pilot or field test design and results;
4. Description of the data collection methodology;
5. Sampling plan and justifications for why it was implemented, a description of any deviations from the plan, and, if possible, the final sample frame;
6. Selected sample;
7. Description of the magnitude of sampling error associated with the survey, and how it was calculated;
8. Description of the sources of nonsampling error associated with the survey (e.g., coverage, measurement);
9. Unit response rates (weighted and unweighted) and nonresponse bias analysis, if applicable;
10. Overall response rates (weighted and unweighted);
11. Item response rates and nonresponse bias analysis, if applicable;
12. Total response rates.
13. Evaluation reports;
14. Descriptions of models used for indirect estimates or projections, if applicable;
15. Analysis plans; and
16. Time schedule for revised data, if applicable.

**GUIDELINE 3-4-2A:** The survey system documentation should also include the following:

1. Final weighting plan specifications, including calculations for how the final weights were derived, and justifications for why it was implemented;
2. Final imputation plan specifications and justifications for why they were implemented;
3. Data editing plan specifications and justifications for why they were implemented; and

4. Data processing plan specifications and justifications for why they were implemented;

**GUIDELINE 3-4-2B:** Where appropriate, methods for bounding or estimating the nonsampling error from each source identified in the evaluation plan should be developed and implemented.

**GUIDELINE 3-4-2C:** Where possible, nonsampling error estimates and bounds should make use of data from other surveys or from administrative records or censuses, taking into account the limitations of the external data.

**GUIDELINE 3-4-2D:** For recurring surveys, a quality profile report that itemizes all sources of identified error should be produced. Where possible, estimates or bounds on the magnitudes of these errors should be provided; the total error model for the survey should be discussed; and the survey should be assessed in terms of this model.

**STANDARD 3-4-3:** To insure that NCES has sufficient information to design future surveys and to accurately estimate their resource requirements, survey system documentation must include the following:

1. All information germane to the contractual operation of the survey, including the request for proposals or performance work statement used to solicit the contract(s);
2. Independent government cost estimate;
3. Contract(s) used to develop, conduct, and report on the survey;
4. Any modifications to the contract(s);
5. Final contract deliverables, progress reports, and vouchers; and
6. Office of Management and Budget (OMB) clearance package and correspondence with OMB about survey clearance.

**STANDARD 3-4-4:** At a minimum, survey documentation must be stored electronically in a format that can be viewed without proprietary software. Final data sets shall be stored in ASCII format. Additional copies in other formats are allowed, but ASCII versions are required. In addition, substantive reports written to release the data shall also be stored, at a minimum, in the format originally used to produce the report, and PDF or ASCII (see Standard 7-1).

**STANDARD 3-4-5:** All reports, documentation, and public-use data files must be stored on the web, a CD-ROM, or NCES dedicated server space. Restricted data files and associated documentation, including final disclosure avoidance and swapping report, must be transmitted to the Statistical Standards Program for secure storage.