NCSES Supplemental Guidelines for New Surveys or Major Revisions of Existing Surveys

Supplemental Guidelines

NCSES Supplemental Guidelines for New Surveys or Major Revisions of Existing Surveys

Background

The Office of Management and Budget (OMB) issued two sets of guidance for surveys in 2006: *Agency Survey and Statistical Information Collections* and Standards and Guidelines for Statistical Surveys

Agency Survey and Statistical Information Collections details OMB's review process and provides advice for improving information collection designs. *Standards and Guidelines for Statistical Surveys* provides 20 standards and numerous guidelines that apply to federal censuses and statistical surveys.

The OMB standards codify the professional principles and practices that federal agencies are required to adhere to, and they detail the level of quality and effort expected in all statistical activities. The OMB standards are not intended to substitute for the extensive body of literature on statistical and survey theory, methods, and operations. For reference, a copy of the table of contents for the OMB standards is provided in appendix A.

Federal agencies are required to adhere to all standards for every statistical survey, even for surveys that have already received OMB approval. "For each statistical survey in existence when these standards are issued and for each new survey, the sponsoring and/or releasing agency should evaluate compliance with applicable standards" (OMB 2006, p. 4).

The OMB standards and guidelines are not designed to be completely exhaustive of all efforts that an agency may undertake to ensure the quality of its statistical information. Agencies are encouraged to develop additional, more detailed standards focused on their specific statistical activities (OMB 2006, p. 4).

NCSES Guidance for a Major Redesign of an Existing Survey, Design of a New Survey, or Improvement of a Continuing Survey

This NCSES supplemental guidance provides additional detail to guidance found in guideline 1.1 (Survey Planning) in OMB's *Standards and Guidelines for Statistical Surveys* , which details work involving the development of a new survey, a major revision of an existing survey, or update and improvement of a continuing survey. See appendix B for the complete OMB guideline 1.1.2. "Planning is an important prerequisite when designing a new survey or survey system, or implementing a major revision of an ongoing survey."

Table 1 summarizes the correspondence between the recommendations found in OMB guideline 1.1.2 and the sections in this set of NCSES supplemental guidelines. Note that activities detailed in "Part C: Data Collection and Processing Aspects" correspond generally to OMB guideline 3.

This guidance provides activities that should be included when developing a new survey, reviewing and revising an existing survey, or updating and improving a continuing survey. The purpose is to

lay the groundwork for a successful new or major redesign by planning for and building into NCSES contracts the activities, time, and flexibility needed.

The activities detailed in NCSES's supplemental guidance have been divided into three parts depending upon the focus. Part A focuses on the conceptual and measurement side of the survey process. Part B focuses on assessments of the statistical side of the process. Part C focuses on assessments of the data collection processes. The division of the activities roughly follows the survey process outlined by Groves et al. (2004). For graphical representations of the survey process extracted from Groves et al. (2004), see appendix C. Note that several activities included in the supplemental guidelines may be used for multiple purposes. Other organizations of the activities could have been given, and the reader is invited to think of the activities in a way that best fits their own framework and context for the work.

The guidance is meant to provide a useful reminder of the issues and activities that need to be considered when developing plans. It should not be viewed as complete, as comprehensive, or as a simple checklist of to-do items. The activities provided in this guidance are highly recommended in part or all, but each survey is unique with a unique set of circumstances, so individual surveys should be planned accordingly. Each issue and activity is designed to contribute to the overall success of the survey. All are aimed at reducing one or more types of potential survey errors.

The guidance does not distinguish which activities are better suited for developing new surveys or revising or improving existing surveys because all activities may be applicable in all situations. The survey manager must make their best judgment about which activities to include under which circumstances.

TABLE 1. Correspondence between NCSES supplemental guidance and guideline 1.1.2 in the Office of Management and Budget's *Standards and Guidelines for Statistical Surveys*

Guideline 1.1.2 – Planning is an important prerequisite when designing a new survey or survey system, or implementing a major revision of an ongoing survey. Key planning and project management activities include the following:

NCSES Part A – Justification for survey

NCSES Part A – Review of related studies, surveys, and reports NCSES Part A – Review of all survey items

NCSES Part A – Plan for pretesting NCSES Part B – Plan for quality assurance

NCSES Parts B and C – Plan for evaluating survey procedures

NCSES Part A – Analysis plan

NCSES Part B – Estimate of resources NCSES Part A – Dissemination plan

Part A: Concepts and Question Development

The activities in this section are aimed at the methodology or measurement side of a survey. The focus here is on developing the survey concept and determining effective methods of measurement, which often takes the form of a question. The idea is to figure out and bridge the differences between what we would like to know about (concepts) with what we can collect (measures).

Concepts, sometimes referred to as constructs, are the high-level questions of interest that the survey is trying to address. They are the elements of the information that are of interest and are often abstract in nature.

Measurements are ways to gather information about the concepts/constructs. They are concrete and are often in the form of a survey question about a specific point of data.

The activities in this section focus on (1) getting input from various sources and (2) developing and fine-tuning questions or measures. Likewise, there are several different sources for getting input or information. There are many different types of methods for eliciting information.

Examples of sources and methods are shown in figure 2. Sources and methods can be mixed and matched in a variety of combinations depending upon cost constraints, timing, and availability of resources. For example, focus groups can be used with data users or with respondents.

Some of the more common combinations used in NCSES survey redesigns are described in Part

The survey manager is encouraged to use combinations of sources and methods as best fit the circumstances.

TABLE 2. Examples of sources of information and methods to elicit information for new or redesigned surveys

Sources	Methods
Media – books, journals, articles, white papers.	Literature reviews
Experts	Panels
Respondents – current or potential	Workshops
Data users	Focus groups
Professional societies	Case studies, record-keeping studies, respondent studies, usability studies
Institutional records – higher education, establishments, companies	Cognitive interviews
Other government agencies – United States or international	Surveys
Contractors – current or present	Debriefings
	Professional conferences, meetings, workshops

Commonly used activities to develop or re-evaluate information needs for a survey are described in items 1 through 5.

Commonly used activities to access understanding and implementation of the developed measures and survey instruments are described in items 6 through 13.

Activities to develop or re-evaluate information needs and to develop specific measures and survey questions

1. **Literature review –** A literature review is one way to get an environmental scan of the topic of interest and to gain an understanding of the past and current topology of the topic, including what is currently known or measured and what types of information or data are missing. Types of media or materials typically covered in a literature review include other survey instruments, journal articles, books, conference proceedings, and websites.

2. **Expert panel** – An expert panel consisting of leaders in the area of interest can provide current and relevant information on a variety of issues facing a new survey or a major revision of an existing survey. Suggested topic areas for the expert panel include the following:
   a. How they think about the field (e.g., What does it look like through their eyes?)
   b. Where they gather their information (e.g., How do they stay current?)
   c. Identify the current state of the field including hot topics (e.g., What issues do they face today? What is important today?)
   d. Emerging issues on the horizon for the field (e.g., Where do they see the area in 5 years, in 10 years?)
   e. Are there any unique circumstances that apply in this area?
   f. Advice on specific pre-identified issues (e.g., state of current affairs for postdoctorates, globalization of R&D)
   g. Potential measures or metrics (e.g., return on investment for R&D)
   h. Identify target audiences and products for dissemination
   i. Views or concerns with confidentiality, security, or sensitivity of data
   j. What is needed but is missing from the currently available data
   k. Suggestions for improvements for existing surveys
   l. Suggestions of other experts to consult

3. **Data user workshops*** – Current and potential data users can provide valuable information about the uses and needs (1.1.2.1). Typical questions to cover in workshops include the following:
   a. Identify how and why current data are used. How does their organization use the data? Do they combine it with our sources?
   b. Identify issues/concerns or problems with current data. Is it measuring what they think it is measuring?
   c. Identify data gaps – What data do they have a need for that is not available? Do they need different cuts of the data?
   d. Set data priorities – Most users want more and more data. Which data are more important than others and why?
   e. Suggest other potential users – Who else might use the data?

[* Worksheets or surveys sent to data users prior to a workshop to collect input for the workshop are subject to generic clearance and require OMB approval.]

4. **Focus groups with respondents –** Focus groups with individuals are used to obtain their perspective on the issues. Groves et al. (2004) notes that the more targeted the topics for the focus groups, generally the more useful the information obtained. Focus groups can provide the following:
   a. Help with understanding of complex topics and uses of common nomenclature and terms
   b. Views on confidentiality concerns and data security
   c. Views on the sensitivity of different types of requested data
   d. Understanding of what can and cannot be answered or what will or will not be answered readily
   e. What is important to them concerning this topic

5. **Record keeping studies –** Record keeping studies are often aimed at establishment surveys but can be used with individuals. The goal of a record keeping study is to understand what is generally kept in a set or book of records that a respondent can refer to for answers. Examples of a set or book of records are a checkbook or immunization record for an individual, an accounting or human resources system for establishments, or a registrar system for institutions. Questions of interest include the following:
   a. What specific data items are kept in the set or system?
   b. Who has access to the set or system?
   c. How hard is it to get the desired information out of the system?
   d. Do all the data of interest reside in one system or are there several systems?
   e. Do different entities keep the same kind of information in their records?

Activities to access and fine tune implementation of the developed measures and survey instrument

6. **Expert reviews of survey instrument** – Outside experts from a variety of areas can provide initial insights into survey concepts, survey wording, and instrument design. Examples of experts include the following:
   a. Accountants - financial information
   b. Human resource managers – employment data
   c. R&D managers – R&D technical information
   d. Lawyers – contracts, patents or tech transfer data
   e. Registrars – student enrollment data
   f. Specific subject area specialist – e.g., biologist for biological topics
   g. Cognitive or survey methodologist – layout and flow of instrument

7. **Cognitive interviews** – Cognitive interviews provide a way to find out how respondents understand and answer the questions. They can be used to find out how respondents define terms. Cognitive interviews can take several forms:
   a. Concurrent think aloud
   b. Retrospective think aloud

c. Paraphrasing
d. Probing

8. **Usability testing** – Usability testing provides a way to find out how easy it is for respondents to answer the survey. Are they able to follow the flow of the questions, do they put their answers in the right place, what format do they provide answers (e.g., date), do they get discouraged before the end, how do they respond to error messages, and other questions related to ease of use. Usability testing is often targeted at Web-based surveys, but it is more generally applied to any self-administered survey. The goal of the testing is to make the survey easy to understand, easy to fill out, and easy to return/submit.

9. **Pretest, field test, or pilot test*** – Pretests are generally small-scale rehearsals for the full data collection efforts. The purpose of the field or pilot test is to evaluate the complete survey design, including the survey instrument, the data collection procedures and systems, and the statistical methods (frame, sample selection, imputation, and estimation). On a qualitative level, it provides a chance to run through the whole process on a smaller scale to learn where procedures or the survey might need to be changed, where procedures are missing, or where redundancies are occurring. On a quantitative level (see number 11. Quantitative evaluations), data from the pilot provide an opportunity to evaluate the survey items based on the actual data collected.

   [* Data collected from a pilot or field test are subject to OMB approval and may be subject to the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA). See http://ciser.cornell.edu/NYCRDC/helpful_links/CIPSEA.pdf.]

10. **Debriefing studies*** – Debriefing studies are used to learn more about any issues or problems that respondents had answering the survey. They provide a way to study response error in the survey, to evaluate understanding of the survey questions, to access availability of the information, to check the validity of the survey measure to reflect the intended construct, to evaluate the reliability (consistency and stability of responses) of the survey measure, and to discover other issues surrounding the respondents' experiences with the instrument and procedures. Good sources of information include respondents, current or past contractors, and data users.

    [* Data collected via respondent debriefing surveys are subject to generic clearance and require OMB approval and may be subject to CIPSEA.]

11. **Quantitative evaluations*** – Unlike the preceding activities, quantitative evaluations are used to assess survey responses based on the actual data collected during a pilot or field test of larger numbers of respondents. These evaluations may be designed to look for high rates of missing data, out-of-range values, inconsistencies with other questions, problems with skip patterns, context or order issues, response bias, or items with little variation in response (straightlining). Behavior coding, randomized experiments, and split-ballot

experiments are methods of quantitative evaluations that might be used. The results of the quantitative evaluations are generally applicable to the population of interest.

[*CIPSEA may apply.]

12. **Validity assessments** – Groves et al. (2004) define validity as "the extent to which the survey measure accurately reflects the intended construct." Validity can be evaluated in two ways. The first way is to use data external to the survey. For example, when the item is a factual item, such as in the number of students in classroom A at school B, external data from the teacher's Blackboard site or a personal visit to the classroom to count the students might be used to validate the response. In this case, validity can be evaluated by calculating the correlation between the survey responses and the record values obtained from the external source. The second way is to use data from related items on the survey. In this case, validity can be evaluated by looking at the relationships between answers of related survey items via correlation or other more sophisticated modeling approaches.

13. **Reinterview studies** – Reinterview studies are a specific type of debriefing study that are worth mentioning separately because of their focus on reliability of the survey questions or measures. Reinterview studies are ones in which survey respondents are asked to answer the same questions a second time so two responses are collected from the same respondent thus allowing for the calculation of several different statistics relating to the consistency and variability of the responses.

    Reliability is a measure of variability and is used to talk about the consistency of a measurement (1) across time or (2) across items designed to measure the same concept. Reliability is computed as the ratio of the variance of true values to the variance of reported values.

    Another common measure of variability and consistency is the "simple response variance," computed as the squared difference between the original response and the reinterview response. When an item has high reliability, it has low simple response variance.

    A third measure of the same concept is the index of consistency computed as (1 – reliability).

    [* Data collected via reinterview studies are subject to generic clearance and require OMB approval, and they may be subject to CIPSEA.]

Deciding which activity to undertake

The different activities described above help to identify different types of issues. When deciding which activities to undertake, it is helpful to keep in mind three main goals of survey questions (Groves et al, 2004, pp. 250–251). Generally, several different activities should be planned since no one activity provides all the information needed.

- Content
    - Do the questions ask for the right information? Experts, subject matter specialists, and data users can provide this information. Panels, focus groups, and workshops work well for gathering this information. Attendance at professional conferences or meetings may also provide insight.
    - Do respondents or potential respondents understand the question as intended? Are the questions sensitive in nature? Focus groups, cognitive interviews, respondent debriefings, re-interview studies, field tests, or quantitative evaluations may be used to obtain this information.
    - Is the information available from respondents? Respondents or potential respondents are the best source for this. When appropriate, do not forget to ask about their record keeping systems. Site visits, record keeping studies, focus groups, cognitive interviews, or quantitative evaluations may be used to obtain this information.
- Cognitive
    - Are the questions understandable? Do the respondents understand what is being asked? Do they interpret the questions that same way? Are any of the terms vague or ambiguous? Focus groups, debriefings, cognitive interviews, and expert reviews can all be used to elicit this information.
    - Can the respondents answer the questions? Are the response categories appropriate and do they make sense? Quantitative evaluations and respondent debriefings (interviews or surveys) can be used to help elicit this information.
- Usability
    - How well do the different modes of the survey instrument work in practice and together? Do respondents know where to start? Can they navigate through the questionnaire appropriately, including correctly following skip patterns? Is the layout confusing? Do they skip over critical information? Usability tests and cognitive tests are often used to assess this. For instruments accessed via computer, capture and analysis of metadata can be very insightful.

The following caveats apply to the activities discussed above in Part A.

- In most cases, information obtained via activities 1 to 8, 10, and 13 cannot be used to infer to the whole population of interest because they generally involve small numbers or may not cover all groups of interest.
- Activities 1 to 5 may not be the best venue for evaluating specific question wording or for understanding how a respondent arrives at an answer (Groves et al. 2004).
- Activity 3 worksheets or surveys are subject to generic clearance and require OMB approval

- Pretests, field tests, or pilot tests (activity 9) can be thought of as dress rehearsals for full productions. They are a testing ground for all aspects of the survey: the questionnaire, survey processes, and survey methods.

- Results from the quantitative evaluations (activity 11) can often be used to infer to the whole population of interest, particularly when the pilot or field data collection is appropriately designed for these purposes (see "Part B: Statistical Aspects").

- Typically, activities 6 through 8 are done before data collection. Activities 10 through 13 are generally done after data collection.

- Debriefing activities may be conducted after a pilot test, a full-scale implementation, or a regular survey collection.

- Activities 9 through 13 may be subject to CIPSEA.

Part B: Statistical Aspects

The following activities are aimed at the statistical aspects of the survey and how well they support the survey goals. The activities focus on assessments of the different statistical aspects.

The assessments cover the following:

- The sample frame(s) in terms of coverage and completeness

- Sample design (sample selection and estimation techniques) in terms of efficiency (precision and cost) and robustness (fitness for use)

- Nonresponse, such as response rates and response bias

- Techniques for handling nonresponse, such as imputation or reweighting in terms of effectiveness

- Estimates of variance

- Limitations of the data for inference or analysis

Assessments of the statistical aspects should occur during the development of pre- and post-data collection activities. Assessments at the pre-data collection development stage are aimed at reducing risk, gaining efficiencies, or acquiring robustness in methods and techniques that are best suited for the specific population of interest. Assessments during the post-collection stage are aimed at measuring how well the methods and techniques came to produce the desired results. Post-collection assessments should be used to improve the statistical methods for the next survey round. For one-time surveys, the post-survey assessments can be used to make improvements for similar surveys.

It is assumed that the following statistical activities are routinely being undertaken as part of the survey or development of a new survey:

- Developing a clear statement of survey goals and concepts of interest

- Clearly defining the target population
- Developing the sample frame
- Developing the sampling and estimation plans (Note that sample design covers both sample selection and estimation techniques since they are intimately linked and should not be developed or considered separately.[1] Estimation techniques cover point, interval, bias. and variance estimates.)
- Measuring nonresponse (unit and item)
- Determining ways to handle nonresponse and nonresponse bias

Broad descriptions of the essential assessments are provided below. For new surveys, the assessment descriptions can be used to help guide the development of each component. All assessments should be built into plans and statements of work for new surveys or revisions of existing surveys.

Sample frame assessment – A sample frame assessment is focused on coverage, both over- and under-coverage, completeness, and access to the sample frame. Coverage is relative to the target population of interest. Completeness is relative to the data items on the frame. A key and necessary component to this assessment is a clear and unambiguous definition of the target population of interest. A sample frame assessment should be conducted during pre-data collection activities.

A sample frame assessment, at a minimum, should address the following:

- Over-coverage of ineligible units
- Under-coverage of omitted units
- Ease and periodicity of access to frame
- Cost to access frame
- Inclusion of key variables, including measure of size and stratification variables, as appropriate
- Cleanliness (accuracy) of information on frame
- Completeness of information on frame
- Timeliness (up to date) of information on frame
- Ability to match (for multiple sources or for list frame development)

[1] That is, certain estimation techniques are valid only for certain sample selection methods.

- In cases where several lists or sources exist but no one source contains all the variables of interest, can the lists or sources be matched at the unit level accurately?

- In cases where several lists or source exist but no one source covers all the units of interest, can the lists or sources be matched to remove overlap among units (duplicate units)?

- Limitations

Sample design assessment – A sample design assessment is focused on the efficiency of the design to provide the desired level of precision for the estimates at the lowest possible cost. The aim of an economic sample design is to (1) minimize the variance (or mean square error) for a fixed cost or (2) to minimize the cost for a fixed variance.

Cost and sample sizes are connected. In most cases a smaller sample size will result in a lower cost; however, under some circumstances, such as cluster or multistage sampling, a larger sample size may result in a lower cost for the desired precision.

Surveys with many variables of interest pose additional challenges to sample design. A design that is optimal for one variable of interest is generally not optimal for other variables of interest. Thus, a more robust design may be chosen over an optimal design with the potential loss of precision for some variables but an overall gain in precision for the majority of items or for key items. Achieved precision should be measured for all estimates produced, not just for key variables, and compared to the desired precision.

Survey sampling techniques fall into two broad camps, design based (D) or predictive/model based(P) (Särndal, Swensson, and Wretman 1992; Valliant, Dorfman, and Royall 2000). Regardless of the choice, a sample design assessment is a critical tool for developing and assessing the efficiency of the chosen design.

Estimation techniques must be compatible with sample selection techniques. A couple of cautions: (1) when complex design-based sample selection techniques are used,[2]estimators developed for simple random sampling generally cannot be used.

Complex design-based sample selection techniques include stratified sampling, two-stage sampling, cluster sampling, systematic sampling, PPS (probability proportionate to size sampling), or adaptive sampling and (2) when predictive/model-base techniques are used, evaluation of the working model prior to making estimates and inferences is critical.

A pre-data collection sample design assessment should address the following:

- Goals of the design

- Comparison of alternative sample designs

[2] The exception is in the case of self-weighting designs.

- Comparison of sample sizes for one or more desired levels of precision
- Robustness of selected design
- Correspondence between selected working model and balanced sampling strategy (P only)
- Estimation techniques, including variance techniques
- Assessment of the sampling variance
- Limitations

A post-collection sample design assessment should include all areas discussed above plus the following:

- Reassess balance of final sample (P only)
- Evaluate appropriateness of original working model (P only)
- Achieved precision

Assessment of nonresponse* – A nonresponse assessment is focused on (1) the amount and type of nonresponse and (2) patterns of unit and item nonresponse potentially leading to nonresponse bias. Unit and item response rates should be calculated according to American Association for Political Opinion Research standards (AAPOR 2016). See appendix D for NCSES preferred calculation. Item response rates should be calculated for all published variables and variables of interest.

As appropriate, response rates in terms of dollars or other units may also be calculated and provide additional insight. Identified patterns of unit or item nonresponse may indicate problems with questionnaire design (e.g., poor layout or wording), survey procedures (e.g., poor or inconsistent follow-up), sample design (e.g., coverage, frame), or other design issues. Any nonresponse adjustments should consider patterns of nonresponse.

Current OMB guidelines require a nonresponse bias study for all surveys with less than an 80% response rate; however, recent research indicates that nonresponse bias can occur in surveys with higher response rates and may not occur in studies with lower response rates. Plan for a nonresponse bias analysis if the expected unit response rates is below 80% (guideline 1.3.4) Plan for a nonresponse bias analysis if the expected item response rate is below 70% for any items used in reports (guideline 1.3.5).

Please note that OMB will generally not approve surveys with unit response rates below 60%. Therefore, in cases where prior experience suggests the potential for an overall unit response rate of less than 60%, the decision to proceed with data collection must be made in consultation with NCSES's chief statistician.

Assessment of techniques for handling nonresponse and nonresponse bias – Non-response adjustments techniques fall into two broad categories: (1) reweighting and (2) imputation.

Within these categories, several options exist, and "Any adjustment method is based explicitly or implicitly on models for the survey items and the response probabilities. Consequently, it is important to validate these models thoroughly, within the limits of the data" (Groves et al. 2002). At a minimum, assessments of nonresponse adjustments should address the following:

- Nonresponse bias in the resulting point estimator
- Variance inflation of adjusted estimator compared to idealized full-sample estimator; this is akin to a design effect for the adjusted estimator
- Variance estimator bias ("Direct application of standard full-sample variance estimation methods to adjusted estimators will lead to negatively biased variance estimators" [Groves et al. 2002].)

Two other useful techniques tools for assessing the overall effect of any nonresponse adjustments are (1) confidence interval coverage rates and widths and (2) power cover analyses. They provide an overall view of the competing effects since, in general, any given adjustment method will not perform uniformly over the three criteria described above.

Assessments described in Part B should be conducted on a regularly scheduled basis for all on-going surveys. New surveys or major revisions of existing surveys should plan time and resources for all Part B assessments. When specific improvement activities are planned for on- going surveys, applicable assessments are to be included.

Part C: Data Collection and Processing Aspects

The following activities are aimed at the data collection and processing aspects of the survey and how well they are performing.

The assessments cover the following:

- Functionality testing of electronic instruments and systems
- Multi- or mixed-mode effects
- Contact strategies
- Coding
- Edits

Assessments should take place prior to and after data collection and processing activities. Assessments conducted prior to the data collection and processing are generally aimed at identifying areas for improvement based on prior survey cycle issues or at changes in procedures due to changing survey requirements or environments. Assessments conducted after the data collection and processing are generally aimed at measuring how well the systems, methods, techniques, or processes came to produce the desired results. Assessments after data collection should be aimed at

correcting any systematic issues that were identified and to improve the activities for the next survey round. For one-time surveys, the post-survey assessments can be used to make improvements for similar surveys.

It is assumed that the following collection and processing activities are routinely being undertaken as part of the survey or development of a new survey:

- Instrument design, may be multi-mode
- Respondent contact strategy
- Data capture systems
- Coding methods
- Editing techniques

Broad descriptions of the assessments are provided below. For new surveys, the assessment descriptions can be used to help guide the development of each component. All assessments should be built into plans and statement of works for new surveys or revisions of existing surveys.

Functionality testing of electronic data collection instruments – Functionality testing covers all types of electronic data collection instruments including, but not limited to, Web based, computer assisted (telephone and personal interview), and spreadsheets or workbooks. Functionality testing ensures all features and functions of the electronic instruments are working as specified, such as built-in edits, skip patterns, automatic summing, Web links, and interactive voice response (originally touch-tone data entry). Functionality testing also covers the underlying data capture system and all interfaces with the system. Functionality testing should ensure that the data are being correctly loaded and saved into the correct variables in the data capture system.

Assessment of multi-mode effects – Multi-mode effects may occur when questionnaire materials are developed in several different modes (e.g. Web, telephone, and mail). Many surveys are increasingly using multi-mode efforts to increase response rates and to accommodate different styles for different types of respondents. However, similarly constructed or even exactly the same questions may elicit different responses depending upon the mode. An assessment of multi- or mixed-mode effects focuses on the consistency of responses across the different modes. What impact do the different modes have on the data? Do respondents answer with the same intent across the different modes?

Respondent contact strategy assessment – A contact strategy assessment is focused on the overall effectiveness of respondent contacts throughout the survey process. All contacts with respondents should be included in the contact strategy. Contacts include e-mail notifications, phone conversations, mailings, Web announcements, and other “discussions or conversations” with respondents. They include pre- survey contacts, initial launch communications, reminders and thank yous, nonresponse follow-ups, flyers, brochures, out-going telephone contacts, and scripts for incoming telephone calls.

A respondent contact strategy assessment, at a minimum, should evaluate the following in how well they support the survey goals (e.g., response rates, quality):

- Number and types or modes of contacts
- Timing or periodicity of contacts
- The language used in the contact material vs. the goal vs. the elicited response
- Costs, for the overall strategy and for individual waves or modes of contacts

Assessment of coding – A coding assessment is focused on how consistent nonnumeric data are translated and summarized into numeric data. Coding can range from a very simple process to a very complex one. Simple cases involve assigning codes to categorical responses. Complex cases involve assigning codes to write-in or verbatim responses. Errors or miscoding can arise in either case. An assessment of the coding, at a minimum, should address the following:

- The category framework used to classify – Do the categories still make sense? Do they reflect the current environment? Are the categories comprehensive and complete? Are the criteria for classifying clear and consistent?
- The accuracy/consistency of the coding – Do the coders, who may be respondents, make consistent decisions when faced with the same types of information? Are the coders able to map the responses to the codes in a one-to- one fashion? How often does miscoding occur? Are the errors systematic in nature?

Assessment of edits – An edit assessment is focused on how well the edit methods improve the quality of the data where quality is defined in terms of accuracy. Edits come in a variety of flavors including range, ratio, comparison, consistency and outlier. Edits may be done in real time as data are collected or at the back end after data are submitted. Edits should not change the data when the data are true and should minimally distort any reported data. Edits are aimed at reducing errors in the data, but in some cases, edits may introduce error into the data.

An edit assessment, at a minimum, should address the following:

- Evaluate the edit counts. That is, which edits are used the most and which are not used at all? Are the edits that are used the most trying to solve a problem somewhere else in the system? Is the data being over edited? Are the edits that are not used at all meaningless? Are there unexpected changes in the edit counts across cycles or across different types of respondents?
- Evaluate the magnitude of the edits. Do the edits cause large changes in the data? Why? Is an edit trying to fix a problem that could be solved by a change in the questionnaire or in the processing? Are a large number of edits causing small changes in the data?
- Evaluate anomalies in the data. How are they identified? Are they one-off cases or is there a systematic issue?
- Evaluate the repeatability of the edits. Are all the edits replicable? If not, why not?

Assessments described in Part C should be conducted on a regularly scheduled basis for all on-going surveys. New surveys or major revisions of existing surveys should plan time and resources for all Part C assessments. When specific improvement activities are planned for on- going surveys, applicable assessments are to be included.

Part D: Iterative Nature

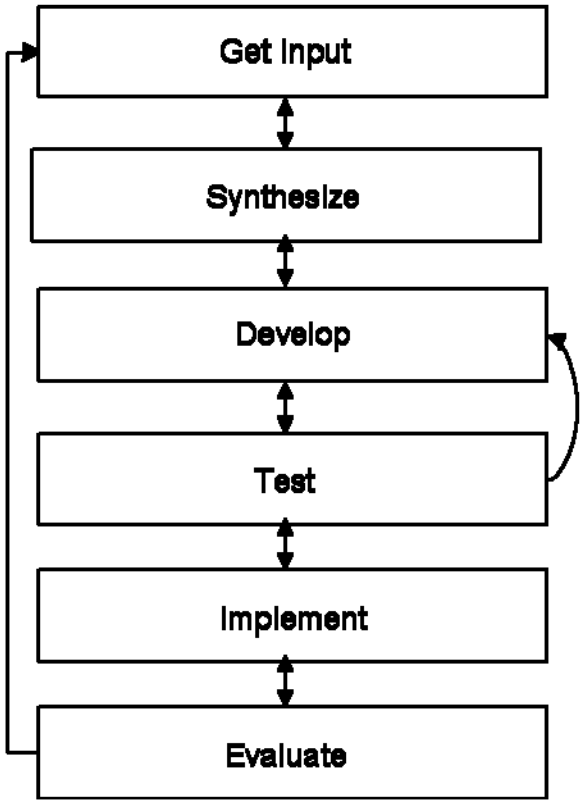
Survey design work is iterative by nature. It is helpful to keep this in mind when designing a new survey, embarking on a major redesign of an existing survey, or improving a continuing survey because it provides a basis for implementing and evaluating both large and small (substantial and incremental) changes. A depiction of the iterative nature is shown in figure 1, which illustrates the feedback loops in the decision-making process.

Decisions are made for each process in a survey through the following steps:

1. Getting input
2. Synthesizing the input
3. Making a decision and developing the process
4. Testing the new or revised process
5. Implementing the survey
6. Evaluating the final results

The arrows between the steps indicate that iterations may be made between decision steps in the process before moving to the next step in the process. Note an evaluation can feed into the next iteration of the current step or into the first part of the next step. For example, the evaluation of a round of cognitive interviews may feed into another round or iteration of cognitive interviews when it results in changes to the questionnaire being tested, or the evaluation of the cognitive interview may feed into the pilot test of the questionnaire.

FIGURE 1. Feedback loops in decision making



Part E: Implementation

It is expected that this guidance will be used when NCSES staff are planning for and developing a statement of work for a new survey, a major revision of an existing survey, or improvements to a continuing survey. The type and number of activities included in the plan and the statement of work will depend upon the specific circumstances.

Justifications for all activities, both included and not included, are to be documented using the attached template, see appendix E, and discussed with the program director and other relevant NCSES senior staff prior to finalizing the plans and the statement of work for a new survey, a major revision of an existing survey, or improvements to a continuing survey.

These guidelines go into effect immediately, 10 December 2008 and any new activities planned subsequent to this data related to updating and improvement of a continuing survey, major revision of an existing survey, or design of a new survey are subject to these guidelines.

Part F: Bibliography

Biemer PP, Lyberg LE. 2003. *Introduction to Survey Quality.* Hoboken, NJ: Wiley & Sons, Inc.

Cochran WG. 1977. *Sampling Techniques.* 3rd ed. Hoboken, NJ: Wiley & Sons, Inc.

Dillman DA. 2007. *Mail and Internet Surveys: The Tailored Design Method.* 2nd ed. Hoboken, NJ: Wiley & Sons, Inc.

Groves RM, Dillman DA, Eltinge JL, Little RJA. 2002. *Survey Nonresponse.* Hoboken, NJ: Wiley & Sons, Inc.

Groves RM, Fowler FJ Jr, Couper MP, Lepkowski JM, Singer E, Tourangeau R. 2004. *Survey Methodology.* Hoboken, NJ: Wiley & Sons, Inc.

Hansen MH, Hurwitz WN, Madow, WG. 1993. *Sample Survey Methods and Theory, Volume I: Methods and Applications.* Hoboken, NJ: Wiley & Sons, Inc.

Hansen MH, Hurwitz WN, Madow WG. 1993. *Sample Survey Methods and Theory, Volume II: Theory.* Hoboken, NJ: Wiley & Sons, Inc.

Kish L. 1965. *Survey Sampling.* Hoboken, NJ: John Wiley & Sons, Inc.

Lohr SL. 1999. *Sampling: Design and Analysis.* Duxbury Press.

Office of Management and Budget (OMB). 2006. *Office of Management and Budget Standards and Guidelines for Statistical Surveys*.

Särndal DE, Swensson B, Wretman JH. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Tourangeau R, Rips LJ, Rasinski K. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

Valliant R, Dorfman AH, Royall RM. 2000. *Finite Population Sampling and Inference.* Hoboken, NJ: Wiley & Sons, Inc.

Wolter KM. 2003. *Introduction to Variance Estimation.* New York: Springer-Verlag.

Appendix A

Table of Contents from **Standards and Guidelines for Statistical Surveys**

OFFICE OF MANAGEMENT AND BUDGET STANDARDS AND GUIDELINES FOR STATISTICAL SURVEYS

September 2006 Table of Contents

Appendix B

OMB Guideline 1.1.2 from Standards and Guidelines for Statistical Surveys

Guideline 1.1.2: Planning is an important prerequisite when designing a new survey or survey system, or implementing a major revision of an ongoing survey. Key planning and project management activities include the following:

1. A justification for the survey, including the rationale for the survey, relationship to prior surveys, survey goals and objectives (including priorities within these goals and objectives), hypotheses to be tested, and definitions of key variables. Consultations with potential users to identify their requirements and expectations are also important at this stage of the planning process.

2. A review of related studies, surveys, and reports of Federal and non-Federal sources to ensure that part or all of the survey would not unnecessarily duplicate available data from an existing source, or could not be more appropriately obtained by adding questions to existing Federal statistical surveys. The goal here is to spend Federal funds effectively and minimize respondent burden. If a new survey is needed, efforts to minimize the burden on individual respondents are important in the development and selection of items.

3. A review of the confidentiality and privacy provisions of the Privacy Act, the Confidential Information Protection and Statistical Efficiency Act of 2002, and the privacy provisions of the E-Government Act of 2002, and all other relevant laws, regulations, and guidance, when planning any surveys that will collect individually-identifiable data from any survey participant.

4. A review of all survey data items, the justification for each item, and how each item can best be measured (e.g., through questionnaires, tests, or administrative records). Agencies should assemble reasonable evidence that these items are valid and can be measured both accurately and reliably, or develop a plan for testing these items to assess their accuracy and reliability.

5. A plan for pretesting the survey or survey system, if applicable (see Section 1.4).

6. A plan for quality assurance during each phase of the survey process to permit monitoring and assessing performance during implementation. The plan should include contingencies to modify the survey procedures if design parameters appear unlikely to meet expectations (for example, if low response rates are likely). The plan should also contain general specifications for an internal project management system that identifies critical activities and key milestones of the survey that will be monitored, and the time relationships among them.

7. A plan for evaluating survey procedures, results, and measurement error (see Section 3.5).

8. An analysis plan that identifies analysis issues, objectives, key variables, minimum substantively significant effect sizes, and proposed statistical tests (see Section 5.1).

9. An estimate of resources and target completion dates needed for the survey cycle.

10. A dissemination plan that identifies target audiences, proposed major information products, and the timing of their release.

11. A data management plan for the preservation of survey data, documentation, and information products as well as the authorized disposition of survey records.

Appendix C

Summary of Survey Process

Groves et al. (2004) provides a nice discussion of the lifecycle of a survey in Chapter 2. Figure 2.5, copied below, shows two sides of the survey process for a quality perspective. On the left is the conceptual and measurement side. On the right is the statistical (representation) side.

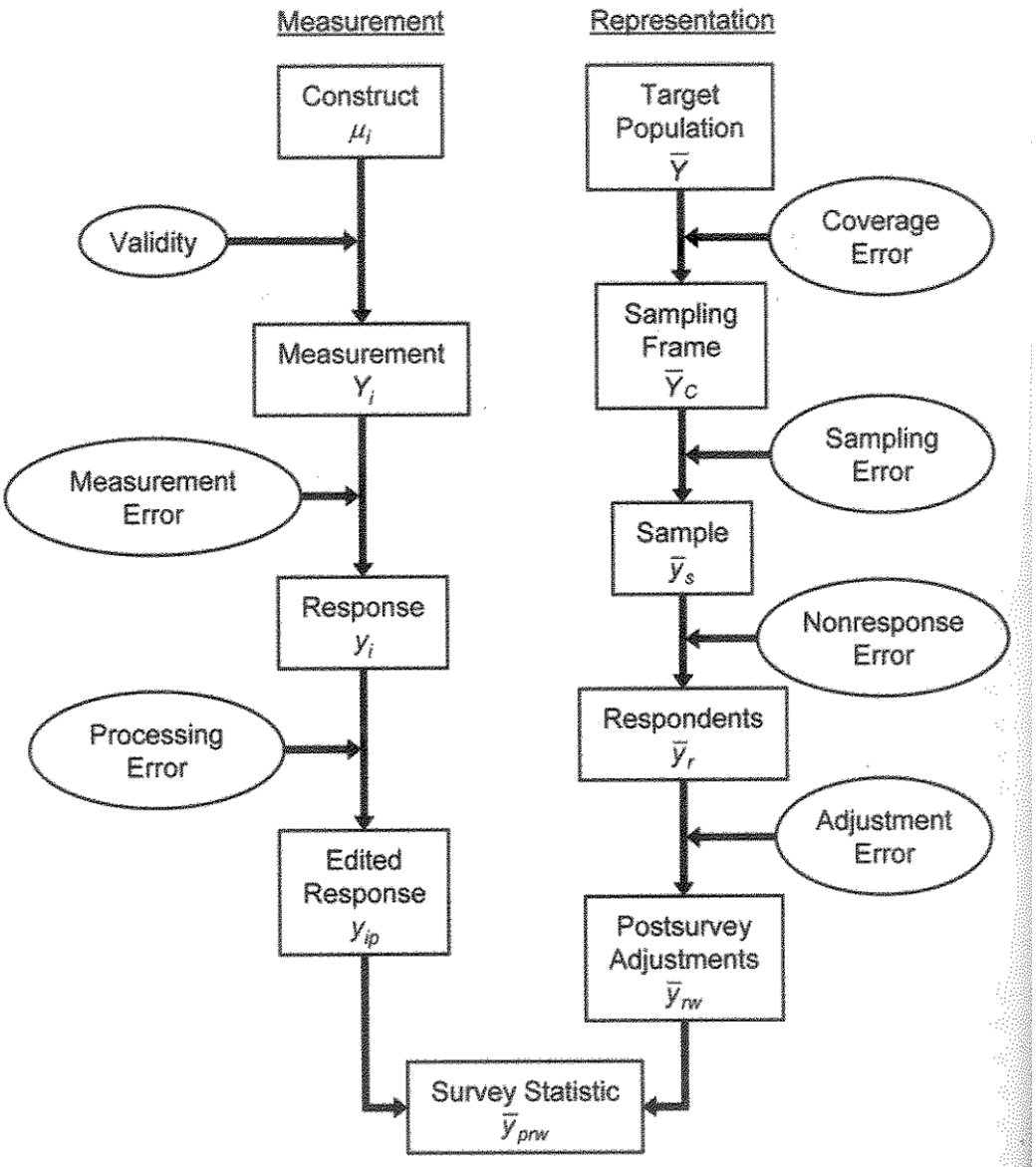


Figure 2.5 Survey lifecycle from a quality perspective.

Groves et al. (2004), in the same chapter, also illustrate a survey from a process perspective. Figure 2.4, copied below, shows this perspective. It is useful to think of both perspectives during the design phases.

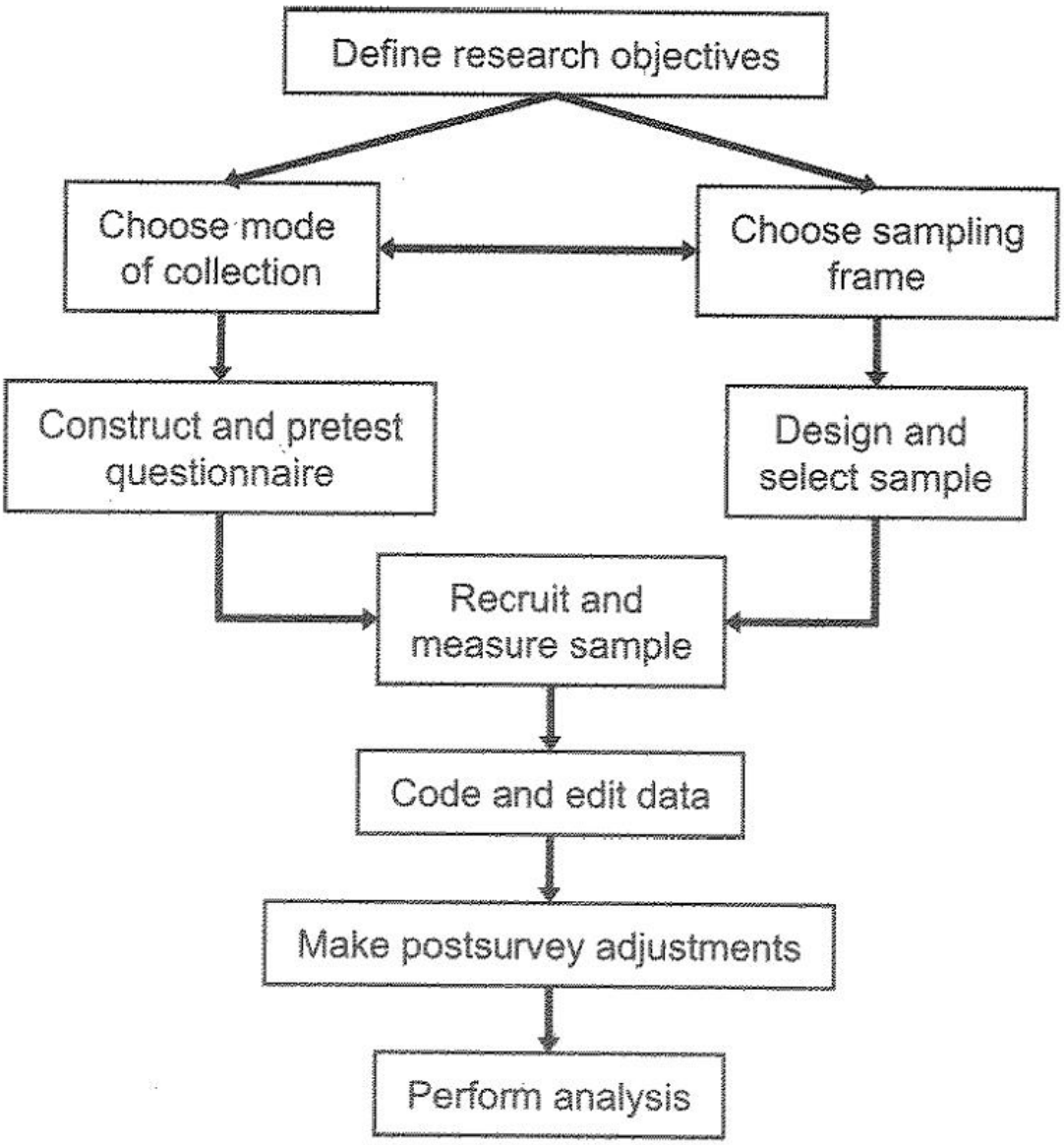


Figure 2.4 A survey from a process perspective.

Appendix D

Response Rate Calculations

The NCSES preferred calculation for response rates is **RR4** as given in American Association for Public Opinion Research (AAPOR). 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys,* pp. 61–62. Available at https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.

The cited pages for response rate calculations are copied below.

Calculating Outcome Rates from Final Disposition Distributions

Numerous outcome rates are commonly cited in survey reports and in the research literature. The same names are used to describe fundamentally different rates and different names are sometimes applied to the same rates. As a result, survey researchers are rarely doing things in a comparable manner and frequently are not even speaking the same technical language. As Groves and Lyberg (1988) have noted, "(t)here are so many ways of calculating response rates that comparisons across surveys are fraught with misinterpretations." Among the more common terms utilized are response, cooperation, refusal, and contact.

As defined by CASRO (Frankel, 1983) and other sources (Groves, 1989; Hidiroglou, et al., 1993; Kviz, 1977; Lessler and Kalsbeek, 1992; Massey, 1995), the response rate is the number of complete interviews with reporting units divided by the number of eligible reporting units in the sample. Using the final disposition codes described above, several response rates are described below:

RR = Response rate

COOP = Cooperation rate

REF = Refusal rate

CON = Contact rate

I = Complete interview (1.1)

P = Partial interview (1.2)

R = Refusal and break-off (2.10)

NC = Non-contact (2.20)

O = Other (2.30)

UH = Unknown if household/occupied HU (3.10)

UO = Unknown, other (3.20)

e = Estimated proportion of cases of unknown eligibility that are eligible

Response Rates

$$RR1 = \frac{I}{(I + P) + (R + NC + O) + (UH + UO)}$$

Response Rate 1 (RR1), or the minimum response rate, is the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-off plus non-contacts plus others) plus all cases of unknown eligibility (unknown if housing unit, plus unknown, other).

$$RR2 = \frac{(I + P)}{(I + P) + (R + NC + O) + (UH + UO)}$$

Response Rate 2 (RR2) counts partial interviews as respondents.

$$RR3 = \frac{I}{(I + P) + (R + NC + O) + e(UH + UO)}$$

Response Rate 3 (RR3) estimates what proportion of cases of unknown eligibility is actually eligible. In estimating e, one must be guided by the best available scientific information on what share eligible cases make up among the unknown cases and one must not select a proportion in order to boost the response rate.[1] The basis for the estimate must be explicitly stated and detailed. It may consist of separate estimates (Estimate 1, Estimate 2) for the sub-components of unknowns (3.10 and 3.20) and/or a range of estimators based of differing procedures. In each case, the basis of all estimates must be indicated.[2]

[1] For example, different values of e would be appropriate in a survey requiring screening for eligibility (e.g., sampling adults 18-29 years old). Two different e's might be used for confirmed households that refused to complete the screener (for which we need an estimate of the likelihood of one or more household members being 18-29) and units that were never contacted (for which we need an estimate of the proportion that are households and an estimate of those with someone 18-29).

[2] For a summary of the main methods for estimating e in surveys (1) minimum and maximum allocation, 2) proportional allocation, 3) allocation based on disposition codes, 4) survival methods, 5) calculations of number of telephone households, 6) contacting telephone business offices, 7) linking to other records, and 8) continued calling), see Smith, 2009.

$$RR4 = \frac{(I + P)}{(I + P) + (R + NC + O) + e(UH + UO)}$$

Response Rate 4 (RR4) allocates cases of unknown eligibility as in RR3, but also includes partial interviews as respondents as in RR2.

$$RR5 = \frac{I}{(I + P) + (R + NC + O)}$$

$$RR6 = \frac{(I+ P)}{(I + P) + (R + NC + O)}$$

Response Rate 5 (RR5) is either a special case of RR3 in that it assumes that e = 0 (i.e., that there are no eligible cases among the cases of unknown eligibility) or the rare case in which there are no cases of unknown eligibility. Response Rate 6 (RR6) makes that same assumption and also includes partial interviews as respondents. RR5 and RR6 are only appropriate when it is valid to assume that none of the unknown cases are eligible ones, or when there are no unknown cases. RR6 represents the maximum response rate.

Appendix E

Template to Be Used When Planning Activities for the Design of a New Survey, Major Revision of an Existing Survey, or Update and Improvement of a Continuing Survey

Instructions – Prior to finalizing statement of work

- Complete questions 1–8 of the template
- Discuss with program director, chief statistician or project statistician, and other project and senior staff
- Complete question 8 and return to program director

1. Name of survey: ____

2. Survey manager: ____

3. Check one: _____Design of a new survey

   _____ Major revision of an existing survey

   _____ Update and improvement of a continuing survey

4. Provide statement of survey goals and concepts of interest.

5. Define target population (population of interest).

6. Define purpose of work for project.

7. For each activity 1 through 13 described in "Part A: Concepts and Question Development "in these NCSES supplemental guidelines, provide the following:

   a. Activity description (e.g., literature review)

   b. Included or not included in plans and statement of work (e.g., included)

   c. Details of included activity or justification for non-inclusion

8. For each assessment 1 through 4 described in "Part B: Statistical Aspects" in these NCSES supplemental guidelines, provide the following:

   a. Assessment description

   b. Included or not included in plans and statement of work

   c. Details of assessment activity OR justification for non-inclusion

9. For each assessment 1 through 5 described in "Part C: Data Collection and Processing Aspects" in these NCSES supplemental guidelines, provide the following:

    a. Assessment description

    b. Included or not included in plans and statement of work

    c. Details of assessment activity OR justification for non-inclusion

10. For each activity and assessment, document decisions and rationale for decisions from discussions with program director, chief statistician or project statistician, and other project and senior staff.