



U.S. Department
of Transportation

**Research and
Innovative Technology
Administration**



Bureau of
Transportation Statistics

BTS Statistical Standards Manual October 2005

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

CHAPTER 2: PLANNING AND DESIGN OF DATA COLLECTION SYSTEMS

- 2.1 Objectives and Requirements
- 2.2 Target Population and Sample Design
- 2.3 Data Collection Methods
- 2.4 Documents and Documentation

CHAPTER 3: COLLECTION OF DATA

- 3.1 Acquiring Data from External Sources
- 3.2 Frame Maintenance and Updates
- 3.3 Data Collection Operations
- 3.4 Documentation of Data Collection Procedures

CHAPTER 4: PROCESSING OF DATA

- 4.1 Data Protection
- 4.2 Data Editing
- 4.3 Missing Data
- 4.4 Data Coding
- 4.5 Monitoring and Evaluation
- 4.6 Documentation of Data Processing Procedures

CHAPTER 5: DATA ANALYSIS

- 5.1 Data Analysis Planning
- 5.2 Statistical Estimation and Inference
- 5.3 Data Analysis Documentation

CHAPTER 6: DISSEMINATION OF INFORMATION

- 6.1 Releasing Information
- 6.2 Tables, Graphs, and Maps
- 6.3 Text Discussion
- 6.4 Micro-data Releases
- 6.5 Data Protection Prior to Dissemination
- 6.6 Rounding
- 6.7 Information Revisions
- 6.8 Public Documentation

CHAPTER 7: EVALUATING INFORMATION QUALITY

- 7.1 Continuing Activities
- 7.2 Data Quality Reviews
- 7.3 Data Evaluation Projects

APPENDIX A: DEFINITIONS OF KEY TERMS

APPENDIX B: REFERENCES

CHAPTER 1

INTRODUCTION

OVERVIEW

The Bureau of Transportation Statistics (BTS), like other federal statistical agencies, establishes professional standards to guide the methods and procedures for the collection, processing, storage, and presentation of statistical data. Standards and guidelines define the professional basis and the level of quality and effort expected in all statistical activities, including those of contractors. The standards ensure consistency among studies conducted by BTS and provide users clear documentation of the methods and principles employed in the development, collection, processing, analysis, and dissemination of BTS statistical information.

The standards and guidelines in this *Manual* apply to BTS data collections or surveys whose purposes include the description, estimation, or analysis of the characteristics of groups. This includes the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support those purposes. Certain standards and guidelines also apply to the compilation of data from external sources and to the dissemination of BTS information products.

BACKGROUND

BTS issues statistical standards and guidelines in response to various legal and OMB requirements:

- BTS is responsible for: “issuing guidelines for the collection of information by the Department [of Transportation] required for statistics ... in order to ensure that such information is accurate, reliable, relevant, and in a form that permits systematic analysis.”¹

¹ 49 U.S.C. 111 (c)(7)

- The Data Quality Act² requires that each federal agency issue guidelines ensuring the quality of disseminated information.

In 2002, the U.S. Office of Management and Budget (OMB) issued government-wide guidelines³ that "provide policy and procedural guidance to Federal agencies for ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by Federal agencies."

In addition to guidelines, agencies were also required to develop a process for pre-dissemination review of information, an administrative mechanism to allow the public to request correction of information not complying with the guidelines, and an annual report to OMB describing the outcome of these requests.

ORGANIZATION OF THESE STATISTICAL STANDARDS

In October 2002, DOT issued its Information Dissemination Quality Guidelines, which included statistical guidelines as Appendix A. This *Manual*, the *BTS Statistical Standards Manual*, does not intend to replace the DOT guidelines. Rather, the *Manual* provides more specific statistical standards and guidelines that BTS needs as a statistical agency.

The content of the *BTS Statistical Standards Manual* follows the outline of the DOT statistical guidelines (DOT 2002, Appendix A), except that the *BTS Manual* divides the DOT chapter on "Processing Data" into two chapters, "Processing of Data" and "Data Analysis."

In addition, OMB, through the Federal Committee on Statistical Methodology, is currently revising its *Standards for Statistical Surveys*. OMB intends these standards to serve as general standards for all Federal statistical agencies. BTS reviewed draft versions of the OMB standards during the development of the *Manual*, which refers to the draft OMB standards in many places. BTS's standards extend the OMB standards to deal more explicitly with BTS data issues, including the use of non-survey data and data from external sources.

This *Manual* contains 28 standards for BTS statistical practice. Each standard is accompanied by guidelines that represent best practices in meeting the standard. Each section also provides a list of key terms (defined in Appendix A of the *Manual*) and a list of related materials.

OMB defines quality in terms of utility (usefulness of information to intended users), objectivity in presentation and in substance, and integrity (protection of information from unauthorized access or revision). BTS addresses the OMB quality criteria in the following fashion:

² Section 515 (PL 106-554).

³ 67 FR 8452, February 22, 2002.

- Utility – the planning standards and guidelines (Chapter 2 and Section 5.1) stress user involvement, while the dissemination standards and guidelines (Chapter 6) emphasize accessibility and transparency to users.
- Objectivity – objectivity in substance, through sound statistical methods, is the focus of the standards and guidelines. Chapter 6 deals with objectivity in presentation.
- Integrity – the standards and guidelines (Chapters 2 through 6) incorporate compliance with existing BTS policies for maintaining data security and protecting confidentiality.

Finally, Chapter 7 contains standards and guidelines to ensure the quality of BTS statistical processes and products by monitoring compliance with these standards.

IMPLEMENTING STANDARDS

BTS project and product managers should adhere to all standards for every statistical activity. Sponsoring offices should evaluate the compliance of their data collection systems and information products with applicable standards. The offices should establish goals for compliance with any applicable standards that are not met. In those rare instances where the strict application of a standard is impractical or infeasible, consider alternative methods of achieving the standard's purpose. Document the reasons why the standard cannot be met and what actions have been taken or will be taken to address any resulting issues.

In applying the standards, BTS managers should consider the importance of the uses of the information as well as the fitness of the information to those uses. At times, BTS must evaluate the potential improvement in data quality that would arise from adherence to the standard if resource constraints or other contingencies make it impossible to meet all standards. BTS must consider these standards and guidelines and apply them efficiently and effectively to achieve the goal of information quality.

REFERENCES

49 U.S.C. 111 as amended by the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users. P.L. 109-59.

Consolidated Appropriations Act of 2001. Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001. P.L. 106-554.

Office of Management and Budget (OMB). 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*. Federal Register, Vol. 67, No. 36, pp. 8450-8460. Washington, DC. February 22.

_____. 2005. *Standards for Statistical Surveys (Proposed)*. Washington, DC. July 14.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*. Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of August 22, 2005.

Approval Date: September 21, 2005

CHAPTER 2

PLANNING AND DESIGN OF DATA COLLECTION SYSTEMS

2.1 OBJECTIVES AND REQUIREMENTS

2.2 TARGET POPULATION AND SAMPLE DESIGN

2.3 DATA COLLECTION METHODS

2.4 DOCUMENTS AND DOCUMENTATION

BTS data collection systems must be designed to meet both internal and external user needs and the agency's legislative mandates.

This chapter covers the planning and design of data collection systems, including:

- Establishing data needs and data collection system objectives (Section 2.1),
- Identifying the data providers (Section 2.2),
- Planning and designing data collection methods to meet data needs and objectives (Section 2.3), and
- Documenting data collection plans and designs (Section 2.4).

2.1 OBJECTIVES AND REQUIREMENTS

Standard 2.1: Planning for a data collection system, whether it is a new system or a revision of an established system, must include:

- Consultation with data users and providers,
- Definition of data needs and objectives, and
- Choice of how to meet data requirements.

Key Terms: major data users, precision

Guideline 2.1.1: Consultation with Data Users and Providers. Develop and update the data system objectives in partnership with major data users and data providers. Establish a process to consult regularly with major data users regarding changes in data needs and possible updates to the data collection system.

- OMB requires publication of a Federal Register notice requesting public comments for all proposed information collections, administered by a federal agency, that would collect data from ten or more persons outside the federal government within a year,
- Consultations with data users and providers should be expanded to include other means for collecting comments and suggestions, such as individual meetings, focus groups, presentations at conferences and workshops, cognitive testing, and pretests/pilot tests.
- When revising an established data collection system, review any previous evaluation studies for information relating user needs to current system performance.

Guideline 2.1.2: Definition of Data Needs and Objectives. Establish system objectives in clear, specific terms that identify data user needs and data analysis goals before initiating data system development. Modifications required later are often difficult and expensive to implement. The definition of data needs should include:

- What data items are needed and how they will be used,
- The precision level required for estimates,
- The format, level of detail, and types of tabulations and outputs, and
- When and how frequently users need the data.

The final data collection choices will be made in the design phase (Section 2.3), taking into account constraining factors (e.g., cost, time, legal factors), and quality of available data.

Guideline 2.1.3: Choice of How to Meet Data Requirements. Before beginning detailed planning for the collection of specific data items, review related studies and data collection systems. Determine whether all or part of the required data are already available, or could be more easily obtained by adding or modifying questions in existing federal data collections.

- If the required information is not directly available, determine whether it can be derived or estimated using existing data sources.

- If existing federal data collection systems meet some but not all of the data requirements, determine whether the existing data systems can be altered to meet the data requirements through, for example, an inter-agency agreement.
-

Related Information

Office of Management and Budget (OMB). 2005. *Standards for Statistical Surveys (Proposed)*, Section 1.1 (Survey Planning). Washington, DC. July 14.

Stopher, P. and Jones, P., eds. 2003. *Transport Survey Quality and Innovation*. Oxford, UK: Pergamon.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Chapter 2 (Planning Data Systems). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: August 15, 2005

2.2 TARGET POPULATION AND SAMPLE DESIGN

Standard 2.2: Planning and design must specify the proposed target population, source for lists of the target population, and (where applicable) sample design and sample size, accuracy requirements, and response rate goals.

Key Terms: accuracy, coverage, frame, response rate, target population

Guideline 2.2.1: Target Population and Frames. Lists of the units in target population are required to obtain information from the target population. Availability of such lists (also known as frames) is often a restriction to the method used in data collection. When a new frame is needed for a data program, develop and implement a plan for constructing the frame. The plan should cover:

- Choice of the target population and the rationale,
- Any exclusions that have been applied to target and/or frame populations by design,
- Sources of lists of target population units,
- Identification and description of other frame files which exist and whether portions of other frame files will be used to construct a new file,
- When applicable, a description of any multistage sampling, such as geographic area sampling, that will be undertaken prior to development of lists of units and the stages in which the final lists will be developed,
- Methods for matching and merging population lists, if applicable,
- Data items needed for units in the frame,
- Anticipated coverage of the target population by the frame,
 - Coverage rates in excess of 95 percent overall and for each major target population subgroup are desirable.
 - Consider using frame enhancements, such as frame supplementation or dual frame estimation, to increase coverage.
 - If the anticipated coverage falls below 85 percent, evaluate and document the potential for bias (OMB 2005).
- Any estimation techniques used to improve the coverage of estimates, such as post-stratification procedures,
- Other limitations of the frame including the timeliness of the frame, and
- Projected frequency of frame updates.

Guideline 2.2.2: Sample Design. A 100 percent data collection may be required by law, necessitated by accuracy requirements, or relatively inexpensive (e.g., data readily available). Otherwise, the sample design should include appropriate sampling methods. Any sample design chosen should ensure the sample will yield the data required to meet the objectives of the data collection.

- Use probability sampling so that sampling error can be estimated. Any use of nonprobability sampling methods (e.g., cut-off or model-based samples) must be justified statistically and be able to measure estimation error.

- The sample design should include:
 - Identification of the sampling frame and the adequacy of the frame,
 - The sampling unit used (at each stage if multistage design),
 - Criteria for stratifying or clustering,
 - Sampling strata,
 - Sample size by stratum,
 - Expected yield by stratum,
 - Sample selection procedures,
 - The known probability (or probabilities) of selection,
 - Estimated efficiency of sample design,
 - Power analyses to determine sample sizes and effective sample size for key variables by reporting domains (where appropriate),
 - Response rate goals (Guideline 4.5.3),
 - Estimation and weighting plan,
 - Variance estimation techniques appropriate to the sample design,
 - Expected precision of estimates for key variables, and
 - References for the sampling methods used.
 - For nonprobability sample designs, include a detailed selection process and demonstrate that units not in the sample are impartially excluded on objective grounds.
 - Discuss potential nonsampling errors, including reporting errors, response variance, measurement bias, nonresponse, imputation error, and errors in processing the data. Indicate steps to be taken to minimize the effect of these problems on the data.
-

Related Information

Bureau of Transportation Statistics. 2004. *Confidentiality Procedures Manual*. Washington, DC.

_____. 2005. *BTS Statistical Standards Manual*, Section 3.2 (Frame Maintenance and Updates). Washington, DC. Available at http://www.bts.gov/programs/statistical_policy_and_research/bts_statistical_standards_manual/index.html as of July 29, 2005.

Cochran, W.G. 1977. *Sampling Techniques*, 3rd ed. New York: Wiley.

Office of Management and Budget (OMB). 2004. Questions and Answers When Designing Surveys for Information Collection. Washington, DC. December 6.

_____. 2005. *Standards for Statistical Surveys (Proposed)*, Section 1.2 (Survey Design) and Section 2.1 (Developing Sampling Frames). Washington, DC. July 14.

Särndal, C.-E., Swensson, B., and Wretman, J. 1991. *Model Assisted Survey Sampling*. New York: Springer Verlag.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Chapter 2 (Planning Data Systems). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Wolter, K.M. 1985. *Introduction to Variance Estimation*. New York: Springer Verlag.

Approval Date: August 15, 2005

2.3 DATA COLLECTION METHODS

Standard 2.3: The design and planning for data collection must include:

- The detailed methods to be used to collect data,
- The data collection instruments and associated instructions,
- A pretest for new data collection systems, or existing systems with major revisions, and
- Plans for the dissemination of major resulting information products to the public.

Key Terms: bias, bridge study, collection instrument, confidentiality, crosswalk, key variable, measurement error, response rate

Guideline 2.3.1: Methods of Obtaining Data. The data collection method should be appropriate to the nature, amount, and complexity of the data requested, the number of data providers, available resources, and the amount of time available.

- Determine the method, or combination of methods, of data collection (e.g., mail, telephone, Internet, etc.) that is appropriate for the target population and the objectives of the data program. The determination should include consideration of the likely effect of method choice on response rates.
- Establish a data collection period that allows sufficient response time for data providers to supply reliable data, including time to follow up on missing data, and meets the required dissemination schedule.
- Develop a plan for confidentiality protection (BTS 2004) during sampling, data collection, processing, data analysis, and dissemination.
- Develop plans for data processing, including data editing and imputation (BTS 2005, Chapter 4).
- Plan for quality assurance during each phase of the data collection process to permit monitoring and assessing the performance during implementation. Include contingencies to modify the procedures if critical requirements (e.g., for the response rate) are not met.
- Establish a formal training process for persons involved in interviewing, observing, or reporting data to ensure that the intended procedures are followed.
- If redesigning an existing data system, analyze and document the potential impact of changes in key variables or data collection procedures.
- Plan for evaluating data collection and processing procedures, results, and potential biases.
- Develop general specifications for an internal project management system for the complete data collection cycle that identifies critical activities and key milestones that will be monitored, and the time relationships among them.

Guideline 2.3.2: Instruments and Instructions. Design the data collection instrument in a manner that maximizes data quality, while minimizing respondent burden:

- Do not use instrument formats that are inappropriate for the method of data collection. For example, if using a self-administered collection instrument, limit skip patterns to ease navigation.
- Develop clearly written instructions to help reporters minimize missing data and measurement error.
- Require that data items are clearly defined in terms the reporters understand, with entries in a logical sequence and with reasonable visual cues and instrument formatting (if applicable). Pretest to identify problems with interpretability.
- Structure the order and presentation of data items such that responses do not unduly influence responses to subsequent items.
- Minimize the number of data calculations and conversions the reporter must make.
- For computer-assisted and other forms of electronic data collection (using GPS devices, sensors, etc.):
 - Test for validity and reliability under conditions similar to those of the planned data collection.
 - Develop protocols for the backup and recovery of data.
 - If possible, have alternate methods of data collection available in case of equipment failure. Otherwise, develop plans to impute or adjust for faulty or missing observations.
- Establish protocols that minimize measurement error, such as conducting response analysis surveys that ensure records exist for data elements requested for business data collections, establishing recall periods that are reasonable for personal data collections, and developing computer systems that ensure internet data collections function properly.

Guideline 2.3.3: Standard Codes and Classifications. To allow data comparisons across databases, use standard names, variables, numerical units, codes, and definitions. Use codes and classifications consistent with the federal coding standards listed below, if applicable. If a federal coding standard does not exist, consult with subject area experts to determine if applicable non-federal standards exist. Provide crosswalk tables to the federal standard codes for any legacy coding that does not meet the federal standards. These codes are updated periodically. Current federal standard codes include:

- FIPS Codes. The National Institute of Standards and Technology (NIST n.d.) maintains Federal Information Processing Standards (FIPS) required for use in federal information processing in accordance with OMB Circular A-130. The following FIPS should be used for coding:
 - 5-2, Codes for the Identification of the States, the District of Columbia and the Outlying Areas of the United States, and Associated Areas.
 - 6-4, Counties and Equivalent Entities of the U.S., Its Possessions, and Associated Areas.
 - 10-4, Countries, Dependencies, Areas of Special Sovereignty and Their Principal Administrative Divisions.
- Statistical Areas. OMB (2005b) defines Metropolitan Statistical Areas, Micropolitan Statistical Areas, Combined Statistical Areas, and New England

- City and Town Areas for use in Federal statistical activities. These areas, as well as principal cities, are updated annually to reflect changes in population estimates.
- **NAICS Codes.** The North American Industry Classification System (NAICS) should be used to classify establishments (U.S. Census Bureau n.d.). NAICS was developed jointly by the United States, Canada, and Mexico to provide new comparability in statistics about business activity across North America. (NAICS coding replaced the U.S. Standard Industrial Classification (SIC) system.)
 - **SOC Codes.** The Standard Occupational Classification (SOC) system (BLS 2000) should be used to classify workers into occupational categories for the purpose of collecting, calculating, or disseminating data.
 - **Race and Ethnicity.** Classification of race and ethnicity, as well as methods of collection, should comply with OMB's Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity (OMB 2000).
 - **Aviation.** The International Air Transport Association, an airline industry association, establishes standard codes for airlines and airport locations (IATA n.d.). The BTS Office of Airline Information also develops and maintains Aviation Support Tables (BTS n.d.) that provide standard codes and other information for air carriers (U.S. and foreign), worldwide airport locations, and for aircraft types and models. The BTS codes do not always agree with IATA coding.
 - **Standard Classification of Transported Goods (SCTG) Reporting System Codes.** The SCTG coding system (Statistics Canada n.d.) was created by the U.S. and Canadian governments, and is used to address statistical needs regarding the transportation of products.
 - **United Nations (UN) Numbers and North American (NA) Numbers.** UN numbers are four digit numbers used worldwide to identify different hazardous materials. The UN numbers are developed through the framework of the United Nations Model Regulations on the Transport of Dangerous Goods. NA numbers are assigned by the U.S. and Canada to hazardous materials that have not been assigned a UN number. The PHMSA Office of Hazardous Materials Safety (PHMSA n.d.) maintains a consolidated table of hazardous materials codes and information.
 - **Injury Codes.** "The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)" (NCHS n.d.) is the official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States. The E-codes in this manual are for injuries. Transportation related injuries span from E800 to E848.
 - **Human Factors Codes.** The FAA Office of Aviation Medicine (FAA 2000) uses "The Human Factors Analysis and Classification System—HFACS."

Guideline 2.3.4: Pretesting. For new data collections or major revisions of ongoing collections, all components must be pretested so that they minimize measurement error and function as intended prior to full implementation.

- One component of pretesting is a pilot test in which some components of a data collection can be pretested prior to a field test of the data collection (for example, using focus groups, cognitive laboratory work, and or calibration studies).

- Another component of pretesting is a field test. Components of a data collection that cannot be successfully demonstrated through previous work should be field tested prior to implementation of the full-scale data collection. The design of a field test should reflect realistic conditions, including those likely to pose difficulties for the data collection.

Guideline 2.3.5: Proposed Data Analysis and Information Products. Develop a dissemination agenda that identifies proposed major information products, timing of release, and their target audiences.

- Proposed data analysis should identify issues, objectives, and key variables, and be linked to the questions the data collection was intended to answer.
- Develop adjustment methods, such as crosswalks and bridge studies that will be used to preserve trend analyses and inform users about the impact of changes.

Related Information

Bureau of Labor Statistics (BLS). 2000. Standard Occupational Classification (SOC) System. Available at <http://www.bls.gov/soc/> as of November 15, 2004.

Bureau of Transportation Statistics (BTS). n.d. Aviation Support Tables. Office of Airline Information: Washington, DC. Available at http://www.transtats.bts.gov/Tables.asp?DB_ID=595&DB_Name=Aviation%20Support%20Tables&DB_Short_Name=Aviation%20Support%20Tables as of July 20, 2005.

_____. 2004. *Confidentiality Procedures Manual*. Washington, DC.

_____. 2005. *BTS Statistical Standards Manual*, Chapters 3-6. Washington, DC.

Available at

http://www.bts.gov/programs/statistical_policy_and_research/bts_statistical_standards_manual/index.html as of July 29, 2005.

Energy Information Administration (EIA). 2002. *EIA Standards Manual*, Standard EIA 2002-5 (Frames Development and Maintenance) and Standard 2002-4 Supplementary Materials, Forms Design Checklist. Washington, DC. Available at <http://www.eia.doe.gov/smg/Standard.pdf> as of January 25, 2005.

Federal Aviation Administration (FAA). 2000. The Human Factors Analysis and Classification System—HFACS. DOT/FAA/AM-00/7. Office of Aviation Medicine: Washington, DC. Available at

<http://www.hf.faa.gov/Portal/ShowProduct.aspx?ProductID=54> as of June 15, 2005.

International Air Transportation Association (IATA). n.d. Airline Coding Directory. London, UK. Available at <http://www.iata.org/ps/publications/9095.htm> as of July 26, 2005.

National Center for Health Statistics (NCHS). n.d. "The International Classification of Diseases, Ninth Revision, Clinical Modification" (ICD-9-CM). Available at <http://www.cdc.gov/nchs/about/otheract/icd9/abtcd9.htm> as of June 14, 2005

National Institute of Standards and Technology (NIST). n.d. Federal Information Processing Standards Publications. Available at <http://www.itl.nist.gov/fipspubs/index.htm> as of November 15, 2004.

Office of Management and Budget (OMB). 2000. Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity. Available at <http://www.whitehouse.gov/omb/inforeg/statpolicy.html#dr> as of November 15, 2004.

_____. 2004. Questions and Answers When Designing Surveys for Information Collection. Washington, DC. December 6.

_____. 2005a. *Standards for Statistical Surveys (Proposed)*, Section 3.3 (Coding). Washington, DC. May 19.

_____. 2005b. *Update of Statistical Area Definitions and Guidance on Their Uses*. Available at <http://www.whitehouse.gov/omb/inforeg/statpolicy.html#ms> as of July 15, 2005.

Pipeline and Hazardous Materials Safety Administration (PHMSA). n.d. "Hazmat Table." Office of Hazardous Material Safety: Washington, DC. Available at <http://www.myregs.com/dotrspa/> as of July 20, 2005.

Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, M., Martin, J., and Eleanor Singer. 2004. *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.

Statistics Canada. n.d. Standard Classification of Transported Goods (SCTG). Ottawa, Canada. Available at <http://www.statcan.ca/english/Subjects/Standard/sctg/sctg-intro.htm> as of June 14, 2005.

Stopher, P. and Jones, P., eds. 2003. *Transport Survey Quality and Innovation*. Oxford, UK: Pergamon.

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. Census Bureau. n.d. The North American Industry Classification System (NAICS). Washington, DC. Available at <http://www.census.gov/epcd/www/naics.html> as of November 15, 2004.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Chapter 2 (Planning Data Systems). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: August 15, 2005

2.4 DOCUMENTS AND DOCUMENTATION

Standard 2.4: Planning activities must include the documentation of user needs and design decisions as well as the preparation of required administrative documents.

Key Terms: coverage, frame, target population

Guideline 2.4.1: Documentation of Data Needs. After establishing the data needs and requirements, prepare a detailed technical document that describes the goals and objectives of the data collection, including:

- A summary of the consultations with major data users and data providers, plus any other sources consulted,
- The information needs that will be met, including the desired accuracy, timeliness, and dissemination format(s) for the data, and
- The choices made for meeting data needs and their relationship to the requirements.

Guideline 2.4.2: Target Population and Frames Documentation. Describe the target populations and associated frames (lists of population units) in detail. Include a discussion of coverage issues (Guideline 2.2.1).

Guideline 2.4.3: Sample Design Documentation. If sampling is part of the data collection design, prepare a detailed description of the sample design (Guideline 2.2.2) and how it will yield the data required to meet the objectives of the data collection. When a nonprobabilistic sampling method is employed, the survey design documentation should include:

- A discussion of what options were considered and why the final design was selected,
- An estimate of the potential bias in the estimates, and
- The methodology to be used to measure estimation error.

Guideline 2.4.4: Collection and Processing Methodology Documentation. Document the collection design and its connection to the data requirements (Section 2.3). The documentation should include the methods of obtaining data, copies of the data collection instrument and instructions, pretest design and findings, and plans for disseminating the results of the data collection to the public.

Guideline 2.4.5: Administrative Documents. Comply with the following requirements as part of the data collection planning and design:

- When planning and design is in its initial stages, prepare a project plan specifying schedules and resource requirements in the format specified by BTS management.
- Data collections (and related activities such as focus groups, cognitive interviews, pilot studies, field tests, etc.) are all collections of information subject to the requirements of the Paperwork Reduction Act of 1995 (P.L. 104-13, 44 U.S.C. 3501 et seq.) and OMB's regulations (5 CFR Part 1320, Controlling Paperwork

Burdens on the Public). OMB approval is required before the agency may collect information from ten or more persons outside the Federal government in a twelve-month period. The documentation specified in this section can all be used in Part B of the submission to OMB (OMB 2004a)

- Projects that require a new IT investment or significant modification of an existing IT investment must go through the Capital Planning and Investment Control process.
 - Contracts should include language stating that the contractor shall comply with all standards and guidelines contained in the *BTS Statistical Standards Manual* and the *BTS Confidentiality Procedures Manual*.
-

Related Information

Bureau of Transportation Statistics. 2004. *Confidentiality Procedures Manual*. Washington, DC.

_____. 2005. *BTS Statistical Standards Manual*. Washington, DC. Available at http://www.bts.gov/programs/statistical_policy_and_research/bts_statistical_standards_manual/index.html as of July 29, 2005.

Office of Management and Budget (OMB). 2004a. Paperwork Reduction Act Submission (Form OMB 83-I). Washington, DC. February. Available at <http://www.whitehouse.gov/omb/inforeg/83i-fill.pdf> as of June 15, 2005.

_____. 2004b. Questions and Answers When Designing Surveys for Information Collection. Washington, DC. December 6.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Chapter 2 (Planning Data Systems). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: August 15, 2005

CHAPTER 3

COLLECTION OF DATA

3.1 ACQUIRING DATA FROM EXTERNAL SOURCES

3.2 FRAME MAINTENANCE AND UPDATES

3.3 DATA COLLECTION OPERATIONS

3.4 DOCUMENTATION OF DATA COLLECTION PROCEDURES

Data collection includes all the processes involved in implementing a planned design to acquire data. Common types of data collection include:

- Regulatory data collections (e.g., the airline traffic data required by 14 CFR 241),
- Administrative data collections (e.g., the border crossing data), and
- Surveys (e.g., the Commodity Flow Survey).

In cases where BTS conducts or sponsors the data collections, BTS has control over the collection process. BTS also uses data from external sources. In these cases, BTS has little or no control over the data collection. External-source data vary in importance for BTS use. Some data, such as the border crossing data, BTS both disseminates and uses in further analyses. BTS uses other external-source data only incidentally in analysis reports.

This chapter contains standards for acquiring data from external sources (Section 3.1), maintaining the frame (list of the target population) (Section 3.2), conducting data collection operations (Section 3.3), and documenting the data collection process (Section 3.4). Except for the guideline on confidentiality protection (Guideline 3.3.4), only Section 3.1 is required for incidentally used external data.

3.1 ACQUIRING DATA FROM EXTERNAL SOURCES

Standard 3.1: Data that BTS acquires from external sources must be evaluated and understood in order to assess the quality for the intended BTS use.

Key Terms: confidentiality, external source

Guideline 3.1.1: Obtaining External Data. Obtain the highest quality version of the external data that is available from the source.

- Verify that the data set is the latest version, and that no corrected or revised data are available for the current or previous time periods. Keep a backup copy of the data.
- Obtain the most complete data documentation available for the corresponding time periods. Acquire any available documentation that can be used to assess data quality.
- Evaluate data from external sources for data quality before deciding whether the data are appropriate for the intended BTS use. The level of BTS's evaluation effort should depend on the thoroughness of the external source's quality control and on the importance of the data for the intended use.

Guideline 3.1.2: Confidential External Data. If the external data are confidential or proprietary, written agreements to acquire the data must stipulate the confidentiality requirements for protecting it.

Related Information

Bureau of Transportation Statistics. 2005. *BTS Statistical Standards Manual*, Chapter 2 (Data Collection Planning and Design) and Chapter 4 (Processing of Data). Washington, DC.

Office of Management and Budget. 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*, Public Comments and OMB Response (Applicability of Guidelines). Federal Register, Vol. 67, No. 36, pp. 8453-8454. Washington, DC. February 22.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Appendix A, Section 1.3 (Applicability). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: April 20, 2005

3.2 FRAME MAINTENANCE AND UPDATES

Standard 3.2: Frames (lists of potential data providers) must be maintained, updated, evaluated, and archived to ensure that coverage is as complete and current as possible.

Key Terms: administrative data collection, bias, coverage, frame, regulatory data collection, target population

Guideline 3.2.1: Maintaining Coverage. Frames must be maintained and updated.

- Maintenance is the continuous revision of the frame based on new information that becomes available during data collection. For regulatory or administrative data collections, frame maintenance requires that changes related to reporting eligibility are promptly reflected in the data collection system.
- Updates are systematic, comprehensive searches for frame changes that canvass all available information. Updates can also include re-examination of reporting categories using more recent information, such as reclassifying airlines based on annual operating revenues.

Maintenance and updating actions include:

- Additions of new potential data providers,
- Revisions due to changes in ownership, name, or address.
- Changes in how data providers are classified (for reporting or sampling purposes), and
- Deletions of data providers no longer in the target population.

Guideline 3.2.2: Coverage Evaluation. In addition to routine maintenance and updates, periodically evaluate target population coverage of frames that are used for recurring data collections.

- The frequency of coverage evaluations depends on the relative stability of the target population and on the frequency of data collection.
- Evaluate coverage of administrative or regulatory data collections at least annually.
- If the frame is properly maintained and updated, problems in coverage for regulatory based systems can be avoided.
- Conduct an evaluation of the potential bias if the frame's coverage of the target population falls below 85 percent (OMB 2005).

Guideline 3.2.3: Archiving. Frames are a critical component of data collection and documentation. A backup copy of the current frame must be created and archived prior to each major frame update (or periodically, for continuously maintained frames).

- All active and inactive data providers must be included on the archive file.
- Inactive records may be periodically deleted from the current file, after the prior file has been archived.

- During a frame update, information on potential data providers should not be deleted from the frame. Instead, a status indicator field in the frame should designate whether the entry is active/inactive or in-scope/out-of-scope.
 - Whenever the information contained in a frame is modified, record the effective date of the change.
 - Provide a way of tracking changes in frame record identifiers over time.
-

Related Information

Bureau of Transportation Statistics. 2005. *BTS Statistical Standards Manual*, Section 2.2 (Target Population and Sample Design). Washington, DC.

Federal Committee on Statistical Methodology. 1990. *Survey Coverage*, Statistical Policy Working Paper 17, Washington, DC: Office of Management and Budget. Available at <http://www.fcsm.gov/working-papers/wp17.html> as of November 5, 2004.

_____. 2001. *Measuring and Reporting Sources of Error in Surveys*, Chapter 5 (Coverage Error), Statistical Policy Working Paper 31, Washington DC: Office of Management and Budget. Available at http://www.fcsm.gov/01papers/SPWP31_final.pdf as of December 20, 2004.

Office of Management and Budget (OMB). 2005. *Standards for Statistical Surveys (Proposed)*, Section 2.1 (Developing Sampling Frames). Washington, DC. July 14.

Approval Date: April 20, 2005

3.3 DATA COLLECTION OPERATIONS

Standard 3.3: Design and administer data collection methods and instruments to balance among:

- The maximization of data quality,
- The control of measurement error and bias due to missing data, and
- The minimization of respondent burden and cost.

In addition, if BTS promises confidentiality of respondent's data, then BTS must protect the privacy rights of the respondents and data providers, and protect their data from unauthorized disclosure.

Key Terms: confidentiality, data collection, Information Collection Request (ICR), key variable, measurement error, nonresponse bias, response rates

Guideline 3.3.1: Quality Assurance. Develop protocols to monitor data collection activities, with strategies to identify and correct problems to ensure quality during data collection:

- Implement a process control system during data collection to monitor data quality. The quality control system should be integrated into the data collection process, and enable staff to identify and resolve problems. The control system should also provide data quality measurements for use as indicators of data collection performance and data quality. Use a data tracking process to ensure that data are not lost when transferred to BTS.
- Use a verification process in data entry to ensure entry errors remain below a set limit based on data accuracy requirements. Include data verification rules in online or other electronic data collection systems.
- Conduct refresher training periodically for persons involved in interviewing, observing, or providing data to maintain proper procedures and standards.
- Track on-going response rates and item nonresponse for key variables. Conduct an evaluation of potential item nonresponse bias if response rates (defined in Section 4.3.1) fall below 70 percent for core items (OMB 2005).
- Determine the core items to obtain when a respondent is unwilling to complete the whole information collection instrument. Target the core items to meet the minimum standard for unit response and to analyze nonresponse bias (Section 4.4).

Guideline 3.3.2: Encouraging Cooperation. To encourage data providers and respondents to participate, train data collection staff on obtaining cooperation, building rapport, and converting refusals, even for mandatory data collections. Response rates and data quality can also be improved through means such as the use of prenotification letters, multiple contacts, and reminder notices.

Guideline 3.3.3: Information Collection Request. Provide respondents with an Information Collection Request (ICR) when collecting information. The ICR is usually

placed on the information collection instrument. Follow the requirements for ICRs given in the “Information Collection” section of the *BTS Confidentiality Procedures Manual*.

Guideline 3.3.4: Protecting Confidential Data. In all phases of data collection, confidential data must be protected from unauthorized access or release:

- Protect identifying information of respondents as collected or on the sample frame from unauthorized release or access.
 - Ensure that controls are in place to prevent unauthorized access to electronic information collections (computer assisted interviewing, web based collections, or other electronic filing methods).
 - Ensure that all data collection staff have received confidentiality training and signed a non-disclosure form prior to collecting data.
 - Use secure means when handling and storing the data during collection to protect against disclosure.
 - Use other means to protect confidential information as outlined in the *Confidentiality Procedures Manual*.
-

Related Information

49 U.S.C. 111, as amended by the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users. P.L. 109-59.

Bureau of Transportation Statistics. 2004. *Confidentiality Procedures Manual*. Washington, DC.

_____. 2005. *BTS Statistical Standards Manual*, Section 2.3 (Data Collection Methods) and Chapter 4 (Processing of Data). Washington, DC.

Groves, R. 1989. *Survey Errors and Survey Costs*. New York, NY: Wiley, Chapters 10 and 11.

Office of Management and Budget (OMB). 2005. *Standards for Statistical Surveys (Proposed)*, Section 2.3 (Data Collection Methodology). Washington, DC. July 14.

Privacy Act of 1974.

Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002, P.L. 107-347, Title V.

Approval Date: April 20, 2005

3.4 DOCUMENTATION OF DATA COLLECTION PROCEDURES

Standard 3.4: The data collection procedures should be documented both for internal staff reference and for the public. Documentation should be thorough enough to allow reproduction of the steps leading to the results.

Key Terms: external source, frame

Guideline 3.4.1: Documentation of External Data Sources. All data that BTS acquires from external sources must have adequate levels of documentation.

Documentation for external sources should include:

- The organization providing data,
- The exact name of the data source,
- If the data were obtained from a publication, the full publication information and source for the data within the publication,
- If the data were acquired as a data file, how the file was obtained, the date obtained, and the cost (if any),
- If the data were obtained from the web, the web address and the date acquired,
- The best documentation available from the external data source on the data collection design (including sampling, if used), the data collection and processing procedures, any analysis or modeling performed, and any evaluations of the data quality,
- Information on whether the external data are confidential or proprietary, and if so, a copy of the written agreement used to obtain the data,
- Any additional notes on the interpretation and use of the data,
- Any personal communications required to obtain the data, information about the data source, or information about data quality, and
- Contact information for further questions.

Guideline 3.4.2: Frame Maintenance Documentation. Documentation for maintaining and updating frames must be written and revised as necessary. The documentation must include:

- The frequency of routine maintenance and major updates,
- Sources of information used for maintenance and updates,
- Procedures for incorporating the results of the updates on all appropriate files, mailing lists, and other data collection control forms or listings
- Summary of results of the frame maintenance and updates, and
- The results of periodic coverage studies.

Guideline 3.4.3: Documentation of Data Collection Operations. The data collection operations documentation should include:

- The method of data collection (e.g., mail, telephone, Internet, etc.), including methods used to track and follow up delinquent reports,

- The data collection period, response achieved by the end of the period, and final response achieved,
 - Copies of materials used in the data collection, including instructions given to data providers,
 - Copies of materials used in training data collection and data provider staff,
 - Schedule of data collection operations,
 - Any response analysis or other validation surveys conducted for new data collection efforts,
 - Quantification of response errors to the extent possible.
-

Related Information

Office of Management and Budget. 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*. Federal Register, Vol. 67, No. 36, pp. 8450-8460. Washington, DC. February 22.

Approval Date: April 20, 2005

CHAPTER 4

PROCESSING OF DATA

4.1 DATA PROTECTION

4.2 DATA EDITING

4.3 MISSING DATA

4.4 DATA CODING

4.5 MONITORING AND EVALUATION

4.6 DOCUMENTATION OF DATA PROCESSING PROCEDURES

Once the data have been collected or acquired from an external source, some processing is usually necessary to make the data ready for conversion into information products.

This chapter contains standards for securing the data during processing (Section 4.1), checking data for potential errors (Section 4.2), dealing with missing data issues (Section 4.3), and adding information to the data (Section 4.4). This chapter also contains standards for monitoring and evaluating data operations, including nonresponse analysis, (Section 4.5) and for documenting (Section 4.6) the data processing operations.

4.1 DATA PROTECTION

Standard 4.1: Safeguards must be taken throughout data processing to protect the data from disclosure, theft, or loss.

Key Terms: confidentiality, information security, storage

Guideline 4.1.1: Confidentiality Procedures. Implement the confidentiality procedures given in the BTS *Confidentiality Procedures Manual* sections on “Physical Security Procedures” and “Security of Information Systems” to protect the data from unauthorized disclosure or release during data production, use, storage, transmittal, and disposition (e.g., completed data collection forms, electronic files and hard copy printouts).

Guideline 4.1.2: Security of Information Systems. Follow the information system security procedures in the BTS *Confidentiality Procedures Manual*, and periodically monitor and update them. Ensure that:

- Data files, networks, servers, and desktop PCs are secure from malicious software, unauthorized access, or theft.
- Access to confidential data is controlled so that only authorized staff can read and/or write to the data. The project manager responsible for the data should periodically review staff access rights to guard against unauthorized release or alteration.

Guideline 4.1.3: Data Storage. Develop and implement routine data backups. Secure backup data from unauthorized access or release.

Related Information

Bureau of Transportation Statistics. 2004. *Confidentiality Procedures Manual*. Washington, DC.

Federal Committee on Statistical Methodology. 1994. *Report on Statistical Disclosure Limitation Methodology*, Statistical Policy Working Paper 22. Washington, DC: Office of Management and Budget. Available at <http://www.fcsm.gov/working-papers/spwp22.html> as of November 15, 2004.

Office of Management and Budget. 2005. *Standards for Statistical Surveys (Proposed)*, Section 3.4 (Data Protection [*during data collection*]) and Section 6.5 (Data Protection [*during information dissemination*]). Washington, DC. July 14.

Approval Date: April 20, 2005

4.2 DATA EDITING

Standard 4.2: As part of standard data processing, mitigate errors by checking and editing both data BTS collects and data it acquires from external sources.

Key Terms: edit, imputation, outliers, skip pattern

Guideline 4.2.1: Types of Edits. At a minimum, the editing process must include checking for the items below, and appropriate editing if errors are detected.

- Omission or duplication of records/units,
- Data that fall outside a pre-specified range, or for categorical data, data that are not equal to specified categories,
- Data that contradict other data within an individual record/unit,
- Data inconsistent with past data or with data from outside sources,
- Missing data that can be directly filled from other portions of the same record or through follow-up with the data provider,
- Incorrect flow through prescribed skip patterns, and
- Selections in excess of the allowable number, such as multiple selections for a “mark one” data item.

Guideline 4.2.2: Editing Process. In a data editing system:

- Develop editing rules in advance of any data processing. Rules may be modified during data processing (Section 4.5.1).
- Minimize manual intervention, since it will result in inconsistent applications of the edit rules and may introduce human error.
- Set the acceptable data ranges for outlier checks at broad enough levels so that legitimate special effects, trend shifts, or industry changes are not erroneously removed.

Guideline 4.2.3: Edit Resolution. Several actions are possible when a data value fails an edit check. Recommended procedures are:

- Verify with the original source or respondent and correct as appropriate, or
 - Change the data value to the most likely value based upon other information collected, or impute a substitute value (Guideline 4.3.4).
 - For administrative or regulatory data, any changed value needs the data provider’s acceptance.
 - Notify the source if a change is made to data provided by an external source.
 - Replacing the failed value with a missing value indicator (Guideline 4.4.2), and
 - Accepting the data value as reported. Provide reasons for overriding edits.
-

Related Information

Federal Committee on Statistical Methodology. 1990. *Data Editing in Federal Statistical Agencies*, Statistical Policy Working Paper 18. Washington, DC: Office of Management and Budget. Available at <http://www.fcsm.gov/working-papers/wp18.html> as of November 15, 2004.

_____. 1996. *Data Editing Workshop and Exposition*, Statistical Policy Working Paper 25, Washington, DC: Office of Management and Budget. Available at <http://www.fcsm.gov/working-papers/wp25a.html> as of November 15, 2004.

_____. 2001. *Measuring and Reporting Sources of Error in Surveys*, Statistical Policy Working Paper 31, Section 7.2.3 (Editing Errors), Washington, DC: Office of Management and Budget. Available at http://www.fcsm.gov/01papers/spwp31_final.pdf as of November 15, 2004.

Hawkins, D.M. 1980. *Identification of Outliers*. New York: Chapman and Hall.

Office of Management and Budget. 2005. *Standards for Statistical Surveys (Proposed)*, Section 3.1 (Data Editing). Washington, DC. July 14.

Approval Date: April 20, 2005

4.3 MISSING DATA

Standard 4.3: Unit and item nonresponse must be appropriately measured, adjusted for, and reported. Response rates must be computed using standard formulas to measure the proportion of the eligible respondents represented by the responding units.

Key Terms: bias, eligible unit, imputation, item, item nonresponse, multivariate analysis, nonresponse bias, overall unit nonresponse, probability of selection, response rates, sample substitution, unit, unit nonresponse, weight

Guideline 4.3.1: Basis for Rates. Calculate unit and item response rates based either on the probability of selection (for household or personal data collections) or on the unit's measure of size for industry or establishment data collections.

- Base proportions of the total industry on a measure of size available for all eligible units (e.g., annual operating revenue, total employment).
- For sample surveys, use the inverse of the probability of selection (base weights) in response rate calculation. For 100 percent (universe) data collections, the base weight for each unit is one.
- For sample designs using unequal probabilities, such as stratified designs with optimal allocation, report weighted missing data rates along with unweighted missing data rates.
- If sample substitutions were made, calculate response rates without the substituted cases.

Guideline 4.3.2: Unit Response Rates. Calculate unit response rates (*RRU*) as the ratio of the number of completed data collection cases (*CC*) to the number of in-scope sample cases (AAPOR 2000). A number of different categories of cases comprise the total number of in-scope cases:

- CC* = number of completed cases;
- R* = number of cases that refused to provide any data;
- O* = number of eligible units not responding for reasons other than refusal;
- NC* = number of noncontacted units known to be eligible;
- U* = number of units of unknown eligibility; and
- e* = estimated proportion of units of unknown eligibility that are eligible.

The unit response rate (OMB 2005) represents a composite of these components:

$$RRU = \frac{CC}{CC + R + O + NC + e(U)}$$

- The numerator includes all cases that have submitted sufficient information to be considered complete responses for the data collection period.

- Complete cases may contain some missing data items. Data collection staff and principal data users should jointly determine the criteria for considering a case to be complete.
- The denominator includes all original survey units that were identified as being eligible, including units with pending responses with no data received, new eligible units added to the data collection effort, and an estimate of the number of eligible units among the units of unknown eligibility. The denominator does not include units deemed out-of-business, out-of-scope, or duplicates.
- An unweighted version of the unit response rate can be used for tracking and analyzing data collection operations.
- A simple way to calculate $e(U)$ is to compute the weighted ratio of eligible to ineligible in completed cases or eligibility-known cases and assume the same ratio will apply to the U cases.
- If a data collection has special circumstances that justify a formula other than the one above, such as longitudinal or partial response considerations, a more appropriate formula can be used if accompanied by a full explanation of the calculation method.
- When a data collection has multiple stages, calculate the overall unit response rates (RRO^C) as the product of two or more unit level response rates.

Guideline 4.3.3: Item Response Rates. Calculate item response rates (RRI) as the ratio of the number of respondents for whom an in-scope response was obtained (CC^x for item x) to the number of respondents who were requested to provide information for that item. The number requested to provide information for an item is the number of unit level respondents (CC) minus the number of respondents with a valid skip for item x (V^x). When an abbreviated questionnaire is used to convert refusals, the eliminated questions are treated as item nonresponse.

$$RRI^x = \frac{CC^x}{CC - V^x}$$

- Calculate the total item response rates (RRT^x) for specific items as the product of the overall unit response rate (RRO) and the item response rate for item x (RRI^x).

$$RRT^x = RRO * RRI^x$$

Guideline 4.3.4: Imputation. Decisions regarding whether or not to adjust data, adjust weights, and impute for missing data should be based on how the data will be used and the assessment of the bias due to missing data that is likely to be encountered.

- To avoid biased estimates, include imputed data in any reported totals.
- When used, imputation procedures should be internally consistent, be based on theoretical and empirical considerations, be appropriate for the analysis, and make use of the most relevant data available.

- Since most data sets are subject to analysis by users to detect relationships between variables, implement imputation methods that preserve multivariate relationships.
- To ensure data integrity, re-edit data after imputation.

Guideline 4.3.5: Weight Adjustments. For data collections involving sampling, adjust weights for unit nonresponse, unless unit imputation is warranted. Adjust weights for missing units within classes of sub-populations to reduce bias.

Related Information

American Association for Public Opinion Research. 2000. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Lenexa, Kansas: AAPOR.

Kalton, G. 1983. *Compensating for Missing Survey Data*. Institute for Social Research, University of Michigan.

_____ and Flores-Cervantes, I. 2003. Weighting Methods, *Journal of Official Statistics* Vol.19, No.2.

_____ and Kasprzyk, D. 1982. Imputing for missing survey responses. *Proceedings of the Section on Survey Research Methods American Statistical Association, 1982, 22-31.*

_____ and Kasprzyk, D. 1986. The treatment of missing survey data. *Survey Methodology*, Vol. 12, No. 1, 1-16.

Little, R.J.A. and Rubin, D. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.

Office of Management and Budget. 2005. *Standards for Statistical Surveys (Proposed)*, Section 3.2 (Missing Data). Washington, DC. July 14.

Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Schafer, J.L. 1997. *Analysis of Incomplete Multivariate Data*. London, UK: Chapman and Hall.

Approval Date: April 20, 2005

4.4 DATA CODING

Standard 4.4: To allow appropriate analysis, use codes to identify missing, edited, and imputed items. Codes added to convert collected text information into a form that facilitates analysis must use standardized codes, when available, to enhance comparability with other data sources.

Key Terms: coding, editing, external source, imputation, skip pattern

Guideline 4.4.1: Codes for Missing and Inapplicable Data. Use codes on the file that clearly distinguish between cases where an item is missing and cases where an item does not apply, such as when skipped over by a skip pattern.

- Distinguish between data missing initially from the source, unreadable data, and data deleted in the editing process.
- If the data collection instrument contains skip patterns, distinguish between items skipped and items not ascertained (such as refusals).
- Do not use blanks and zeros to identify missing data, as they tend to be confused with actual data. Similarly, do not use numeric codes like a series of nines or eights for missing numeric items if these could be legitimate reported values.
- If a data file acquired from an external source was not previously coded, the level of coding effort should depend on how BTS plans to use the file and on whether BTS plans to further disseminate the file.
- For data in tabular form, the *BTS Guide to Style and Publishing Procedures* contains a number of symbols and abbreviations to place in cells with various types of missing or inapplicable data.

Guideline 4.4.2: Indicating Edit Actions and Imputations. Code the data set to indicate edit actions and imputed values.

- Indicate whether cases passed or failed each edit. If a case fails an edit, indicate the edit disposition (Guideline 4.2.3).
- If more than one method could be used to impute a missing data item, indicate the imputation method used.

Guideline 4.4.3: Coding Text Information. Although it is preferable to pre-code responses, it may be necessary to code open-ended text fields for further use.

- To code text data for easier analysis, use standardized codes if they exist (Guideline 2.3.3). Develop other types of codes by using existing DOT or other federal agency practice, or by using standard codes from industry or international organizations, when they exist.
 - When manually coding text, create a quality assurance process that verifies at least a sample of the coding to determine if a specific level of coding accuracy and reliability is being maintained.
-

Related Information

American Association for Public Opinion Research. 1998. "Standard Definitions – Final Dispositions of Case Codes and Outcome Codes for RDD Telephone Surveys and In-Person Household Surveys," <http://www.aapor.org/ethics/stddef.html>.

Bureau of Transportation Statistics (BTS). 2003. *BTS Guide to Style and Publishing Procedures*. Washington, DC.

_____. 2005. *BTS Statistical Standards Manual*, Chapter 2 (Data Collection Planning and Design). Washington, DC.

Office of Management and Budget. 2005. *Standards for Statistical Surveys (Proposed)*, Section 3.3 (Coding). Washington, DC. July 14.

Approval Date: April 20, 2005

4.5 MONITORING AND EVALUATION

Standard 4.5: Monitor and evaluate each data processing activity, both to assess the impact on data quality and to inform data users.

Key Terms: frame, imputation, item nonresponse, incident data, longitudinal, missing at random, multivariate modeling, nonresponse bias, overall unit nonresponse, population, response rates, unit nonresponse, weight

Guideline 4.5.1: Quality Control. Establish quality control procedures to monitor and report on the operation of data processing procedures.

- Incorporate quality control into the processing procedures to automatically produce outputs useable by data system managers. Outputs produced during data processing should be used to adjust procedures for higher quality results and greater efficiency.
- Monitor failure rates for each edit and by case. Analyze the pattern of edit failures graphically to pinpoint problems more easily and prioritize items for follow-up.
- When applicable, automate the process of referring data problems to data providers for quicker resolution.
- Maintain information on the amount of missing data, actions taken, and problems encountered during imputation for inclusion in the data processing (Guidelines 4.6.2 and 4.6.3) and user documentation (Guideline 6.8.1).

Guideline 4.5.2: Unit Response Analysis Requirement. Conduct an analysis of nonresponse for any data collection with an overall unit response rate (Guideline 4.3.2) less than 80 percent. The objective is to measure the impact of the nonresponse and to determine whether the data are missing at random.

- Compare respondents and nonrespondents across subgroups using external or frame data, if available, or through a nonresponse follow-back survey.
- Compare respondents' characteristics to known characteristics of the population from an external source. This comparison can indicate possible bias, especially if the characteristics in question are related to the data collection effort's key variables.
- Consider multivariate modeling of response using respondent and nonrespondent external data to determine if nonresponse bias exists.
- For a multi-stage data collection effort, focus the response analysis on the stages with the higher missing data rates.
- Evaluate the impact of weighting adjustments on nonresponse bias.

Guideline 4.5.3: Item Response Analysis Requirement. If the item response rate (Guideline 4.3.3) is less than 70 percent, conduct an item nonresponse analysis to determine if the data are missing at random at the item level, in a similar fashion to Guideline 4.5.2.

- Analyze missing data rates at the item level and compare the characteristics of the reporters and the non-reporters.

- For some data collections, such as incident data collections, missing data rates may not be known. In such cases, provide estimates or qualitative information on what is known.

Guideline 4.5.4: Timing of Nonresponse Bias Analyses. Conduct unit and item nonresponse bias analyses prior to the release of any information products

- Analyze the missing data effect at least annually if the data collection occurs more than once a year or is continuous.
- Analyze the missing data effect each time data are collected if the collection occurs annually or less often.
- For data collections from longitudinal panels, analyze the effect of missing data after each collection due to attrition of respondents over time.

Guideline 4.5.5: Publishable Items. In those cases where the analysis indicates that the data are not missing at random, the decision to publish individual items should be based on the amount of potential bias due to missing data.

- If the missing data bias analysis shows that the data are not missing at random and the total item missing data rate (Section 4.3.3) is less than 70 percent, the estimate should be regarded as unreliable.
- Suppress or flag estimates that are unreliable due to missing data.

Related Information

Bureau of Transportation Statistics. 2005. *BTS Statistical Standards Manual*, Chapter 6 (Dissemination of Information). Washington, DC.

Groves, R. 1989. *Survey Errors and Survey Costs*. New York, NY: Wiley, Chapters 10 and 11.

Interagency Household Survey Nonresponse Group. Information available at <http://www.fcs.gov/committees/ihsng/ihsng.htm> as of April 18, 2005.

Office of Management and Budget. 2005. *Standards for Statistical Surveys (Proposed)*, Section 3.2 (Nonresponse Analysis and Response Rate Calculation). Washington, DC. July 14.

Approval Date: April 20, 2005

4.6 DOCUMENTATION OF DATA PROCESSING PROCEDURES

Standard 4.6: The data processing procedures must be documented for both BTS and public use. For external source data, the documentation must include procedures used by the external source as well as procedures that were implemented on the data at BTS. Documentation must allow reproduction of the steps leading to the results.

Key Terms: coding, derived data, edit, external source, imputation, item response, response rates, unit response, weight

Guideline 4.6.1: Edit Procedures. Documentation must describe:

- The edit rules and their purpose,
- Procedures for handling records that fail edits,
- A description of the codes used to indicate edit disposition (Guideline 4.2.3), and
- The procedures for, and the results of, any edit performance evaluations.

Guideline 4.6.2: Measures of Edit Performance. For key edits as identified by the data collection staff, maintain measures for the number of:

- Edit messages, by edit disposition (Guideline 4.2.3),
- Edit messages resulting in revisions of the original data, and
- Edit messages overridden, by reason for overriding the edit.

Guideline 4.6.3: Procedures for Handling Missing Data. Documentation of procedures for handling missing data must include:

- The unit response rate or rates,
- Item response rates for key variables as identified by the data collection staff,
- Item response rates for any items with response rates less than 70 percent,
- Formulas used to calculate unit and item response rates,
- Results of response bias analyses,
- Full documentation of the methods of imputation or weight adjustments,
- A description of the coding schemes used to identify missing and imputed values, and
- An assessment of the nature, extent, and effects of imputation or weight adjustments.

Guideline 4.6.4: Procedures for Coding Text Information. Document both the source for any coding scheme used and the coding process (whether automated or manual), and make it available to data users. Any reliability or accuracy studies of the coding process should also be documented and made available.

Guideline 4.6.5: Derived Data Items. Documentation should include all formulas, detailed descriptions on how the item was created, and the sources of any external information used to derive additional data items for the file.

Guideline 4.6.6: Information Systems Documentation. Systems for the processing of data should have documentation of all operations (both automated and manual) necessary to operate, maintain, and update the systems.

- The documentation should provide an overview of integrated manual and automated operations, workflow, interfaces, and personnel requirements.
- Documentation should be sufficiently detailed and complete that personnel unfamiliar with the systems can become knowledgeable and operate them, if necessary.
- Information systems documentation may be incorporated into existing documentation or written as a separate document.

Guideline 4.6.7: Documentation Updates. Update documentation whenever a major change to the processing system is made, but at least annually when the frequency is less than annual.

Related Information

American Association for Public Opinion Research. 1998. "Standard Definitions – Final Dispositions of Case Codes and Outcome Codes for RDD Telephone Surveys and In-Person Household Surveys." Available at <http://www.aapor.org/ethics/stddef.html> as of April 18, 2005.

Office of Management and Budget. 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*. Federal Register, Vol. 67, No. 36, pp. 8450-8460. Washington, DC. February 22.

Approval Date: April 20, 2005

CHAPTER 5

DATA ANALYSIS

5.1 DATA ANALYSIS PLANNING

5.2 STATISTICAL ESTIMATION AND INFERENCE

5.3 DATA ANALYSIS DOCUMENTATION

BTS employs a wide variety of statistical techniques in its work. However, regardless of the techniques used, there are some steps that should be included in any data analysis. This chapter provides general guidance on those steps, and then leaves the choice of analytical tools up to the data analyst performing the work.

This chapter contains standards for planning a data analysis (Section 5.1), calculating estimates and performing inferences (Section 5.2), and documenting the data analysis (Section 5.3). For quick-response projects, compliance with these standards is recommended, but not required.

5.1 DATA ANALYSIS PLANNING

Standard 5.1: Plan before starting a specific data analysis to ensure that the resulting product addresses the needs of BTS customers and that the resources are available to complete the data analysis.

Key Terms: key variable, target audience

Guideline 5.1.1: Criteria for the Conduct of Data Analysis. The data analysis should be relevant, objective, comprehensive, and add value to existing information. To meet these goals, data analysts need to:

- Conduct the data analysis in an objective and policy-neutral manner that focuses on the statistical and economic facts.
- Maintain awareness of subject matter issues so that the data analysis can address topics of interest and importance.
- Consult with subject area specialists about relevant issues, the strengths and weaknesses of data sources, and important references to key topic elements.
- If the data analysis is not comprehensive, indicate what further types of data analysis should be considered and whether BTS plans to do that work.

Guideline 5.1.2: Data Analysis Plan Requirement. Prepare a data analysis plan in the proper format (BTS 2004) prior to the start of the data analysis.

- Include the purpose of the data analysis, the research question, target audience, data sources (including a description and any limitations), key variables to be used, and the data analysis methods. Also provide target completion dates and an estimate for the amount of resources needed to complete the product.
 - Subject matter experts should review the plan to ensure that the proposed data analysis will answer relevant questions. Data analysis experts should review the plan to ensure that appropriate data and methods will be used.
 - The data analysis plan must be approved by the designated manager.
-

Related Information

Bureau of Transportation Statistics (BTS). 2004. *BTS Information Product Scoping Paper*. Washington, DC.

Approval Date: June 28, 2005

5.2 STATISTICAL ESTIMATION AND INFERENCE

Standard 5.2: Estimates and statistical inferences made regarding the data must be based on acceptable statistical practice.

Key Terms: accuracy, bias, bridge estimates, estimates, inference, reliability, robustness, time series, trend, variance

Guideline 5.2.1: Data Analysis Methods. Analyses must use theory and methods justifiable by reference to statistical literature (provided below in “Related Information”) or by mathematical derivation.

- Use appropriate analysis methods for complex sample, time series, and geospatial data, or variance estimates may be seriously biased.
- If extensive seasonality, irregularities, known special causes, or variation in trends are present in the data, take those into account in the trend analysis.
- Use robust methods if in doubt about the quality of the data (i.e., the quality of the data cleaning) or about the suitability of the data for analysis by standard parametric methods.

Guideline 5.2.2: Indicating Uncertainty. Statistical statements should be accompanied by some assessment of the limitations and uncertainty of the results.

- Estimated errors due to statistical sampling or modeling indicate the reliability of the estimate. However, these estimated errors do not account for bias, which may have a greater effect on accuracy, and does not decrease as the number of cases increases.
- Analysts must consider data quality issues related to measurement error and missing data. The purpose, design, methods, and quality of processing can all place limitations on the analysis and interpretation of the data. If possible, quantify and eliminate biasing effects. Otherwise, discuss the nature and estimated magnitude of these limitations in the report.

Guideline 5.2.3: Inference and Comparisons. Support statistical statements with proper testing and inference procedures.

- Sampling error estimates should accompany any estimates from samples.
 - For complex sample designs, the BTS office originating the data should provide guidance on estimation and variance calculation. The guidelines should cover proper use of weights and recommend a maximum coefficient of variation and a minimum cell size for usability.
- When doing multiple comparisons with the same data between subgroups, include a note with the test results indicating whether or not the significance criterion (Type I error) was adjusted and, if adjusted, the method used.
- Not every statistically significant difference is important. Given a comparison with a statistically significant difference, subject matter expertise is needed to determine whether the difference is important. In the context of the measure and its fluctuation over time, it may be regarded as insignificant.

Guideline 5.2.4: Bridge Estimates. If the scope of data collection changes or part of an historical series is revised, data for both the old and the new series should be published for a suitable overlap period.

Guideline 5.2.5: Assumptions and Diagnostics. State all statistical assumptions (such as assumptions about data distributions or structured dependence) made during the data analysis.

- Perform diagnostics to detect violations of assumptions, and provide the results of the diagnostics in the report. Plots of data and statistical output, such as residuals, are often useful in detecting violations of assumptions.
- For each assumption, include a discussion of the likelihood that the assumption will be violated by small or large amounts and the robustness of the data analysis method to each such violation.

Related Information

Agresti, A. 1990. *Categorical Data Analysis*. New York, NY: Wiley.

Anderson, T.W. 2003. *An Introduction to Multivariate Statistical Analysis, 3rd ed.* New York: Wiley.

Box, G.P., Jenkins, G.M., and Reinsel, G.C. 1994. *Time Series Analysis: Forecasting and Control, 3rd ed.* New York: Prentice Hall.

Casella, G. and Berger, R.L. 2001. *Statistical Inference, 2nd ed.* Belmont, CA: Duxbury Press.

Chatfield, C. 2003. *The Analysis of Time Series: An Introduction, 6th ed.* New York: Chapman and Hall.

Cleveland, W.S. 1993. *Visualizing Data*. Summit, NJ: Hobart Press.

Cochran, W.G. 1977. *Sampling Techniques, 3rd ed.* New York: Wiley.

Cook, R.D. and Weisberg, S. 1999. *Applied Regression Including Computing and Graphics*. New York: Wiley.

Cressie, N. 1991. *Statistics for Spatial Data*. New York: Wiley.

Daniel, C. and Wood, F.S. 1980. *Fitting Equations to Data*. New York: Wiley.

DeGroot, M.H. 1989. *Probability and Statistics*. Reading, MA: Addison-Wesley.

- Diggle, P.J., Liang, K.-Y., and Zeger, S.L. 2000. *Analysis of Longitudinal Data*. Oxford: Oxford University Press.
- Draper, N.R. and Smith, H. 1998. *Applied Regression Analysis, 3rd ed.* New York: Wiley.
- Efron, B. and Tibshirani, R.J. 1994. *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Fleiss, J.L. 1981. *Statistical Methods for Rates and Proportions, 2nd ed.* New York: Wiley.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. 2005. *Robust Statistics: The Approach Based on Influence Functions, rev. ed.* New York: Wiley.
- Harvey, A.C. 1993. *Time Series Models, 2nd ed.* Cambridge, MA: MIT Press.
- Hicks, C.R., and Turner, K.V. 1999. *Fundamental Concepts in the Design of Experiments*. Oxford, UK: Oxford University Press.
- Hogg, R.V., Craig, A., and McKean, J.W. 2004. *Introduction to Mathematical Statistics, 6th ed.* New York: Prentice Hall.
- Hosmer, D.W., and Lemeshow, S. 1989. *Applied Logistic Regression*. New York: Wiley.
- Huber, P.J. 1981. *Robust Statistics*. New York: Wiley.
- Kelsey, J.L., Whittemore, A.S., Evans, A.S., and Thompson, W.D. 1996. *Methods in Observational Epidemiology*. New York: Oxford University Press.
- Kleinbaum, D.G., Kupper, L.L., and Muller, K.E. 1988. *Applied Regression Analysis and Other Multivariable Methods*. Boston: PWS-Kent.
- Lehmann, E.L. and Romano, J.P. 2005. *Testing Statistical Hypotheses, 3rd ed.* New York: Springer Verlag.
- Lehmann, E.L. and Casella, G. 1998. *Theory of Point Estimation, 2nd ed.* New York: Springer Verlag.
- Little, R.J.A. and Rubin, D. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.
- McCulloch, C.E. and Searle, S.R. 2001. *Generalized, Linear, and Mixed Models*. New York: Wiley.

- Mood, A.M., Graybill, F.A., and Boes, D.C. 1974. *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
- Office of Management and Budget (OMB). 2005. *Standards for Statistical Surveys (Proposed)*, Sections 4.1 (Developing Estimates and Projections) and 5.2 (Inference and Comparisons). Washington, DC. July 14.
- Pankratz, A. 1983. *Forecasting with Univariate Box-Jenkins Models*. New York: Wiley.
- Rao, C.R. 1973. *Linear Statistical Inference and Its Applications, 2nd ed.* New York: Wiley.
- Rohatgi, V.K. 1976. *An Introduction to Probability Theory and Mathematical Statistics*. New York: Wiley.
- _____. 1984. *Statistical Inference*. New York: Wiley.
- Rousseeuw, P.J., and Leroy, A.M. 1987. *Robust Regression and Outlier Detection*. New York: Wiley.
- Särndal, C.-E., Swensson, B., and Wretman, J. 1991. *Model Assisted Survey Sampling*. New York: Springer Verlag.
- Scheffé, H. 1959. *Analysis of Variance*. New York: Wiley.
- Searle, S.R., Casella, G., and McCulloch, C.E. 1992. *Variance Components*. New York: Wiley.
- Seber, G.A.F., and Lee, A.J. 2003. *Linear Regression Analysis, 2nd ed.* New York: Wiley.
- Selvin, S. 1996. *Statistical Analysis of Epidemiologic Data*. Oxford, UK: Oxford University Press.
- Skinner, C., Holt, D., and Smith, T. 1989. *Analysis of Complex Surveys*. New York: Wiley.
- Snedecor, G.W. and Cochran, W.G. 1989. *Statistical Methods, 8th ed.* Ames, IA: Iowa State University Press.
- Tukey, J. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- U.S. Department of Transportation. 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Appendix A, Sections 4.3 (Production of Estimates and Projections) and 4.4 (Data Analysis and Interpretation). Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Wolter, K.M. 1985. *Introduction to Variance Estimation*. New York: Springer Verlag.

Zacks, S. 1971. *Theory of Statistical Inference*. New York: Wiley.

Approval Date: June 28, 2005

5.3 DATA ANALYSIS DOCUMENTATION

Standard 5.3: Document the methods and models used in data analysis products to help ensure objectivity, utility, transparency, and reproducibility of the estimates and projections.

Key Terms: reproducibility, transparency

Guideline 5.3.1: Documentation Content. The data analysis report must contain details of the methods used during the data analysis, including a description of software used, a discussion of the data analysis assumptions, and key information relevant to obtaining the data analysis results.

- Document all methods, assumptions, diagnostics, and robustness checks. Provide references to support the methods used in the data analysis, or a derivation of the theory supporting the method used in the report.
 - Include a statement of the limitations of the data analysis, including coverage and response limitations and statistical variation.
 - Archive the data and models used in the data analysis so the estimates can be reproduced.
 - Archive supporting technical documentation, such as standard error and significance test calculations, that help ensure transparency and reproducibility.
 - For recurring reports, consider producing a methodological report.
-

Related Information

Bureau of Transportation Statistics (BTS). 2005. *BTS Statistical Standards Manual*, Section 6.8 (Public Documentation), Washington, DC. Available at http://www.bts.gov/programs/statistical_policy_and_research/bts_statistical_standards_manual/index.html, as of June 10, 2005.

Office of Management and Budget (OMB). 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*. Federal Register, Vol. 67, No. 36, pp. 8452-8460. Washington, DC. February 22.

_____. 2005. *Standards for Statistical Surveys (Proposed)*, Section 4.1 (Developing Estimates and Projections). Washington, DC. July 14.

Approval Date: June 28, 2005

CHAPTER 6

DISSEMINATION OF INFORMATION

- 6.1 RELEASING INFORMATION
- 6.2 TABLES, GRAPHS, AND MAPS
- 6.3 TEXT DISCUSSION
- 6.4 MICRO-DATA RELEASES
- 6.5 DATA PROTECTION PRIOR TO DISSEMINATION
- 6.6 ROUNDING
- 6.7 INFORMATION REVISIONS
- 6.8 PUBLIC DOCUMENTATION

Dissemination is the distribution of information to the public, in any medium or form, including press releases, reports, data files, or web products. These standards cover releasing information (Section 6.1) and ensuring the accuracy and interpretability of different types of BTS information products: tables, graphs, and maps (Section 6.2), text (Section 6.3), and micro data (Section 6.4). The standards also cover issues affecting all information products: data protection (Section 6.5), rounding (Section 6.6), and revisions (Section 6.7). Finally, the public documentation standard (Section 6.8) provides for the transparency and reproducibility of the information disseminated by BTS.

6.1 RELEASING INFORMATION

Standard 6.1: Procedures for the release of information products to the public must receive predissemination reviews (disclosure, content matter, statistical and methodological) and must include provisions for ensuring fair access to all users.

Key Terms: information product, peer review, sensitive material

Guideline 6.1.1: Release Schedules. To provide fair access to the public, major information products should follow published release schedules.

- Provide the schedule for the release of information products to the BTS public affairs office for release.
- Protect information to be published against any unauthorized pre-release or disclosure in advance of the publication schedule.

Guideline 6.1.2: Ease of Accessibility and Understanding. Information products should be made accessible to the public.

- All information products disseminated through the Internet should comply with the requirements for Section 508 of the Rehabilitation Act of 1973, as amended.
- Codes, abbreviations, and acronyms should be used sparingly and defined in accordance with the *BTS Guide to Style and Publishing Procedures*. Provide definitions to the user in the product.
- As appropriate, information products should also include definitions of any subject-matter-specific or otherwise technical terms.

Guideline 6.1.3: Formal Pre-Dissemination Review. All information products require pre-dissemination review to ensure compliance with OMB and DOT Information Quality Guidelines, and BTS standard procedures.

- Before sending an information product outside the originating office for review, the product manager should:
 - Verify compliance with all applicable BTS standards and guidelines (BTS 2002, 2003, 2004, 2005),
 - Double-check facts,
 - Proofread text, and
 - Clearly mark the product as a draft for review only, and not for attribution or further distribution.
- All information products require a confidentiality protection review (BTS 2004).
- Verify calculations through an independent recalculation of a random selection of statistics in the information product.
- Persons not directly involved in preparing the information product should proofread the text and verify that numbers in tables, graphs, maps, and text are consistent.
- All information products require a subject-matter review by someone, preferably from within BTS, who is familiar with the topic area and with the techniques

used. The information product may require a separate review of the statistical methodology.

- If the topic may be of interest to another DOT organization, industry group, or others in general, ask formally for review by those deemed most interested.
- Publication specialists should edit text products to ensure consistency and readability.
- The appropriate office director should review and clear all information products before submitting the products to the Director or the Director's designee. The Director or designee will review the product and determine whether the product needs further review within the Research and Innovative Technology Administration (RITA) prior to final dissemination approval.
- Information products to be posted on the web require review for compliance with BTS web guidelines. Each information product should be assigned to at least one of the program web pages.

Guideline 6.1.4: External Peer Review. If an external peer review process is used:

- Select peer reviewers primarily on the basis of necessary technical expertise,
- Any non-government peer reviewers paid by BTS must disclose to DOT any prior technical/policy positions they may have taken on the issues at hand and their sources of personal and institutional funding (private or public),
- Conduct peer reviews in an open and rigorous manner, and
- Consider all relevant technical comments, although outside reviews are not binding on BTS.

Guideline 6.1.5: Contact Information. All information products must include a contact reference to BTS customer service. As part of the dissemination process, inform the information service of new products and provide background information so that the information service staff can appropriately respond to, or forward, inquiries regarding the information product and its data sources.

Related Information

Bureau of Transportation Statistics (BTS). 2002. *Section 508 Compliance Plan, Version 1.0*. Washington, DC.

_____. 2003. *BTS Guide to Style and Publishing Procedures*. Washington, DC.

_____. 2004. *Confidentiality Procedures Manual*. Washington, DC.

_____. 2005. *BTS Statistical Standards Manual*. Washington, DC.

Office of Management and Budget (OMB). 2000. *Electronic and Information Technology Accessibility Standards*, Final Rule. Federal Register, Vol. 65, No. 246, pp. 80500-80528. Washington, DC. December 21.

_____. 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*. Federal Register, Vol. 67, No. 36, pp. 8452-8460. Washington, DC. February 22.

_____. 2005. *Final Information Quality Bulletin for Peer Review*, Final Bulletin. Federal Register, Vol. 70, No. 10, pp. 2664-2677. Washington, DC. January 14.

_____. 2005. *Standards for Statistical Surveys (Proposed)*, Sections 6.1 (Review of Information Products) and 7.1 (Missing Data). Washington, DC. July 14.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*. Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: May 4, 2005

6.2 TABLES, GRAPHS, AND MAPS

Standard 6.2: Tables, graphs, and maps in BTS information products must accurately and effectively convey the information intended.

Key Terms: external source, statistical map, weighted average

Guideline 6.2.1: Identifying Content and Sources. As far as possible, tables, graphs, and maps should be interpretable as stand-alone products.

- Titles for tables, graphs, and maps should be clearly worded and identify the content. Include the timeframe and geographical limitations. The axis units may serve to identify the timeframe for graphs.
- All tables, graphs, and maps must have a complete source note (BTS 2003b). Include information not immediately evident from the main body of the presentation, such as definition of codes, acronyms and special terms, and anything else that would not be obvious to the general reader.
 - Source references should be sufficiently detailed for a reader to identify the data used. Source notes in all products must give a full citation for the actual source from which the data were taken, even if that source merely collected data from other sources.
 - Since databases and documents may be updated, the “as of” date for the source should also be noted. Web links should include the URL and date accessed. Even a report featuring results entirely from one source should have the source note with each table, graph, or map, in case they are separated from the report.
- When presenting estimates that are calculated using data from external sources, note each source and add a statement describing how the calculation was done. If the calculation is complex (e.g., a weighted average constructed from raw data and weights), include a description of the methods used or a reference to where they are described. Cite BTS as the source of the calculations, based on the external source.
- Use footnotes to clarify data illustrations, tables, graphs, and maps regarding particular points, abbreviation symbols, and general notes.

Guideline 6.2.2: Consistency of Presentation. To facilitate comparability, be consistent in constructing tables, graphs, and maps within an information product that cover similar material and use similar units.

- Tables, graphs, and maps within the same information product should use similar fonts, units, spacing, and line thicknesses. Symbols and codes should also be similar throughout an information product.
- For comparability across BTS products, tables and graphs must comply with BTS style and formatting guidelines (BTS nd, 2003a, 2003b).

Guideline 6.2.3: Tables. Each cell in a table must have a number, a zero indicator, or a symbol indicating the reason that data are not displayed. Numbers in tables must comply with the following criteria:

- All values in a vertical list of numbers must have the same number of decimal places. Use no decimal if all of the values in a vertical list are integers.
- Use no greater precision than is warranted by the data (see section 6.6).
- Only display zeros for values that are true zeros. If a value rounds to zero, use alternate symbols (BTS 2003b), such as "--," to indicate that the estimate rounds to zero in the units being presented.
- For sample-based zero estimates, use alternate symbols to indicate that the estimates are negligible, but possibly non-zero, in the population.
- All tables that should logically sum to either 100 percent or some other numeric total must provide a note if the summation is affected by independent rounding or missing data.

Guideline 6.2.4: Graphs. Design graphs to maximize clarity and comparability within the information product and with other BTS products.

- Design color graphs to show sufficient contrast if printed in black and white or viewed by a colorblind user. Web graphs need appropriate alternative text for use by screen readers.
- Graph titles and axis labels should be clear with no unexplained or undefined acronyms or industry jargon. In graphs with axes, indicate well-defined variable names and units for each axis. Both axes of a graph should be labeled with the names of variables, except where the axis label "years" is unnecessary because the years are shown.
- Graphs that users are likely to compare should have similar scaling to facilitate the comparison.
- Gridlines can be helpful to users if kept inconspicuous.
- Minimize non-data clutter.
- Minimize use of stacked bar or line graphs. They tend to present minimal information and are usually harder to interpret than simple tables or line graphs.
- Do not use 3D graphs to present two-dimensional data
- When using time intervals, spacing should be equidistant only if the intervals are equidistant.
- In graphs, a vertical numerical axis should normally include zero or a break indicator (two slashes). If adding such a break is not reasonable due to software restrictions, add a note that the vertical axis is not zero-based.
 - For graphs showing relative quantities such as an index, zero is not a meaningful reference point. In such graphs, use the natural basis (such as 100) as a reference line in the graph.

Guideline 6.2.5: Statistical Maps. Statistical maps must comply with graph standards where applicable.

- Use shadings for a statistical map that can be easily distinguished, even if reproduced in black and white.

- Design category intervals to minimize the differences within classes and maximize the differences between classes. Limit the number of intervals to show better contrasts of shades. Three to five intervals should suffice.
 - Take care that the statistical map displays the data intended.
 - Use a scale consistent with the statistical information displayed.
 - Note that statistical area maps tend to emphasize geographic area versus other factors.
 - Provide an accuracy statement when appropriate. Note when data displayed in statistical maps have been collected locally and reflect varying methods of data collection.
 - Provide a distance scale and a legend that defines symbols and other graphic devices used in the map.
 - Map symbols and categories should be consistent throughout an information product and throughout series of related maps.
-

Related Information

Bureau of Transportation Statistics (BTS). nd. *BTS Web Software Guidelines*. Washington, DC.

_____. 2002. *Section 508 Compliance Plan, Version 1.0*. Washington, DC

_____. 2003a. *BTS Excel Table Standards*. Washington, DC.

_____. 2003b. *BTS Guide to Style and Publishing Procedures*. Washington, DC.

Energy Information Administration (EIA). 1998. *EIA Guidelines for Statistical Graphs*. Washington, DC. Available at <http://www.eia.doe.gov/neic/graphs/preface.htm> as of April 19, 2005.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Section 5.1 (Publications and Disseminated Summaries of Data). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: May 4, 2005

6.3 TEXT DISCUSSION

Standard 6.3: Information should be presented clearly and objectively to the public, including a full disclosure of source(s).

Key Terms: confidential, objectivity, reliability, significant, time series, variance

Guideline 6.3.1: Presentation. Documents should be well organized with language that clearly conveys the message to the intended audience. Text discussion in information products must be consistent with accompanying tables, graphs, and maps, whether they are adjacent to the text or in other areas of the product. Include tables with text wherever possible.

Guideline 6.3.2: Sources. Data presented in the text that do not refer directly to the tables, graphs, or maps in the text must have a source reference (see Section 6.2.1).

- Information used in BTS information products should come from known reliable sources.
- Sources for which methodological information is unavailable (such as proprietary data) must include advisories indicating the lack of source and accuracy information.

Guideline 6.3.3: Data Discussions. Discussions of data should be objective and make statistically appropriate statements.

- Fundamental changes within time series should be fully discussed. These changes may include, but are not limited to, changes to how the data were collected, changes in definitions, changes to the population, or changes in processing methods.
 - Statistical interpretations should indicate the amount of uncertainty. Only discuss differences or changes if the appropriate statistical tests verify their statistical significance. Terms such as “confidential,” “reliability,” “significant,” and “variance” should only be used in the statistical sense.
 - Avoid statements that imply a specific cause and effect relationship where one has not been established. Speculative statements about possible causes are acceptable if worded as speculation and not fact, and if supported by legitimate research citations.
 - No policy recommendations may be made regarding solutions to problems except with regard to data requirements.
-

Related Information

Bureau of Transportation Statistics (BTS). 2005. *BTS Statistical Standards Manual*, Chapter 5 (Data Analysis). Washington, DC.

Office of Management and Budget (OMB). 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies, Public Comments and OMB Response (Applicability of Guidelines)*. Federal Register, Vol. 67, No. 36, pp. 8453-8454. Washington, DC. February 22.

Plain Language Action & Information Network. nd. Writing User-Friendly Documents. Available at <http://www.plainlanguage.gov/handbook/index.htm> as of February 9, 2005.

U.S. Department of Transportation (DOT). nd. *Plain Language Resource Page*. Available at <http://www.dot.gov/ost/ogc/plain.htm> as of February 9, 2005.
_____. 2002. *The Department of Transportation Information Dissemination Quality Guidelines*. Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: May 4, 2005

6.4 MICRO-DATA RELEASES

Standard 6.4: Where confidentiality protections permit their release, release micro data (unit-level data) in a manner that facilitates its usefulness to the public. Documentation must be readily accessible to customers, provide the metadata necessary for users to access and manipulate the data, and clearly describe how the information is constructed.

Key Terms: metadata, micro data, record layout, standard error, variance, weight

Guideline 6.4.1: Software Accessibility. If micro data are released as an information product, all micro-data products and documentation should be made accessible without requiring the use of any one particular commercial product. Open source formats (ASCII text, space delimited, comma delimited, etc.) must be made available in addition to any others.

Guideline 6.4.2: File Description. Provide complete documentation for all data files.

- Data producers should determine what metadata standards are current at the time data files are prepared and produce associated metadata for their files that comply with applicable standards.
- Documentation must include a description of the data files including the title, data collection sources, tables that make up the set, inter-relation among tables (e.g., keys), and record layouts for data files.
- Documentation must also include descriptions for each variable in the data set that includes the variable name, description, type (categorical, numerical, date/time, etc.), format, entry restrictions (e.g., categories, range), and missing value codes.
- Indicate changes made to previously released data and the “as of” date of the data file.

Guideline 6.4.3: Information Quality Discussion. Micro-data files must include a discussion of how the data were collected and the limitations of the data (see Section 6.8).

Guideline 6.4.4: Items Needed for Variance Estimation. Datasets containing sample data must contain appropriate weights and associated variables for accurate variance estimation. A dataset that requires weights and additional variables for the computation of estimates and standard errors should not be released before these items become available.

Related Information

American Association for Public Opinion Research (AAPOR). nd. *Best Practices for Survey and Public Opinion Research*. Available at http://www.aapor.org/default.asp?page=survey_methods/standards_and_best_practices/best_practices_for_survey_and_public_opinion_research#best12 as of April 29, 2005.

International Organization for Standardization. 2002-2003. *ISO/IEC 11179, Information Technology -- Metadata Registries (MDR)*, (multipart standard). Available at <http://metadata-standards.org/11179/> as of January 25, 2005.

Office of Management and Budget (OMB). 2005. *Standards for Statistical Surveys (Proposed)*, Section 7.4 (Documentation and Release of Public Use Micro Data). Washington, DC. July 14.

Approval Date: May 4, 2005

6.5 DATA PROTECTION PRIOR TO DISSEMINATION

Standard 6.5: All information products must be released in accordance with applicable Federal law and regulations in conjunction with any confidentiality pledges given to data providers.

Key Terms: confidentiality, disclosure limitation

Guideline 6.5.1: Non-disclosure of Confidential Data. For information collected under a confidentiality pledge, employ statistical disclosure limitation procedures and methods to protect any identifiable or other confidential data from disclosure prior to public dissemination. BTS staff must follow the established confidentiality procedures outlined in BTS' *Confidentiality Procedures Manual* (2004).

Guideline 6.5.2: Security of Disclosure Limitation Methods. The BTS confidentiality officer must review and approve any descriptions of disclosure limitation methods prior to their public dissemination.

- Do not publish the details about how disclosure limitation methods were used to protect the data, if publication could jeopardize data confidentiality. For example, do not reveal information on how noise may have been added to the data, what variables were used to implement record swapping, or the parameter values used to protect tabular data.

Guideline 6.5.3: Disclosure Review Requirements. All information products must be reviewed for compliance with the disclosure protection procedures stated under the section, "Disclosure Review Board," in the BTS *Confidentiality Procedures Manual* (2004).

Related Information

Bureau of Transportation Statistics (BTS). 2004. *Confidentiality Procedures Manual*. Washington, DC.

Bureau of Transportation Statistics Confidentiality Statute, 49 U.S.C. 111(k) as amended by the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users. P.L. 109-59.

Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002. P.L. 107-347, Title V.

Privacy Act of 1974, as Amended, 5 U.S.C. 552a.

Approval Date: May 4, 2005

6.6 ROUNDING

Standard 6.6: Use consistent practices for rounding and displaying numbers in text, tables, and figures.

Key Terms: precision, significant digit

Guideline 6.6.1: Using Rounded Numbers. All calculations should be made before rounding. In particular, tabulations to produce summary data and computations performed for purposes of estimating standard errors should be done on data as collected. No rounding should take place before completing these kinds of tabulations.

- The sum of the rounded numbers may not equal the rounded sum. In such a case, add a note indicating that totals may not equal the sum of their individual components due to independent rounding.
- To allow users to make further calculations accurately, do not further round estimates disseminated in a spreadsheet.

Guideline 6.6.2: Degree of Rounding in Text and Graphs. The degree of rounding for text discussion and graphs should depend on the type of data (actual measure vs. sample), the known or suspected accuracy of the data, and the differences being discussed.

- Round percentages appearing in text to whole numbers unless smaller differences being discussed require decimal places and the accuracy supports it.
- Perform rounding consistently for similar subjects throughout the information product.
- In multiplying or dividing numbers, the resulting precision cannot be more precise than the least precise of the component numbers.

Guideline 6.6.3: General Rounding Rule. Consistent with BTS standard software (BTS nd), the general rules for rounding are:

- If the first digit to be dropped is less than 5, then do not change the last retained digit (e.g., round 6.1273 to 6.127).
 - If the first digit to be dropped is 5 or greater, then increase the last retained digit by 1 (e.g., round 6.6888 to 6.69).
-

Related Information

Bureau of Transportation Statistics (BTS). nd. *BTS Web Software Guidelines*. Washington, DC.

Energy Information Administration (EIA). 2002. *EIA Standards Manual*, Standard 2002-15 (Rounding) and Standard 2002-15 Supplementary Materials (Guidelines on the Standard for Rounding). Washington, DC. Available at <http://www.eia.doe.gov/smg/Standard.pdf> as of January 25, 2005.

National Center for Educational Statistics (NCES). 2002. *NCES Statistical Standards*, Standard 5.3 (Rounding). Washington, DC. Available at http://nces.ed.gov/statprog/2002/std5_3.asp as of January 25, 2005.

Approval Date: May 4, 2005

6.7 INFORMATION REVISIONS

Standard 6.7: A standard process for handling possible post-dissemination data changes should exist and be documented.

Key Terms: external source, revision

Guideline 6.7.1: Scheduled Revisions. When appropriate, establish a schedule for anticipated revisions and make it available to users.

- Identify the first dissemination of a data value in an information product as "preliminary" if revisions are anticipated in a subsequent dissemination.
- Designate scheduled revisions to data values as "revised" (or "final") when disseminating the changes.

Guideline 6.7.2: Errors in Previously Disseminated Information. Actions taken when data errors are discovered, or an external data source makes changes, are dependent on the impact that the potential revision would have on previously disseminated estimates.

- Establish threshold criteria for making revisions. For example, the threshold criteria might be to revise for changes exceeding five percent in smaller values or exceeding one percent in larger values.
- If the change does not exceed the threshold criteria, or threshold criteria do not exist, then management will determine whether the error is serious enough to warrant a revision.
- Document the error discovery and correction process.

Guideline 6.7.3: Documentation of Error Corrections. Document the nature of the changes, any corrective action needed to fix an error, and provide this information to data users.

- Identify data values changed due to unscheduled revisions and explain the reasons for these changes to data users.
- Document problems regardless of the scope of the error or the decision whether or not to revise the data.
- Provide error documentation to data users. Ensure timely and wide dissemination of information product revisions.

Guideline 6.7.4: Monitoring Revisions to Disseminated Data. Track the differences between an initial release of estimates and the corresponding final disseminated estimates for key data series.

- Examine the effect of revisions (number of times data are revised and the magnitude of the revisions). Revision error information can help users better understand the variability between initial estimates and final estimates. For data systems that are continuously updated, compare the initial estimates with estimates obtained after a suitable period has elapsed.

- Some ways to present revision error information include the average revision error, the maximum revision error, or the distribution of revision errors during a specified time period.
 - If revision error for a key data series shows an initial release is an unreliable indicator of the final estimate, consider whether publishing the estimate with a measure of revision error or withholding the initial estimate is the best way to serve data users.
-

Related Information

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Section 6.4 (Data Error Correction). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: May 4, 2005

6.8 PUBLIC DOCUMENTATION

Standard 6.8: Documentation for the public must include the materials and tools (if applicable) necessary to properly interpret and evaluate disseminated information.

Key Terms: archive, reproducibility, transparency

Guideline 6.8.1: Source and Accuracy Information. Source and accuracy information should provide summary information suitable for posting on the web, and should be regularly updated to include methodological changes and the results of any quality assessment studies. Source and accuracy statements should summarize:

- Data system objectives and frequency of information release,
- Target population and coverage, geographic or other characteristic distribution and, where applicable, sample selection methodology and sample size,
- Data collection methodology and content of forms,
- Data adjustments for missing data, nonresponse, coverage error, measurement error, seasonality, and (if applicable) confidentiality protection,
- Estimation methodology, including variance estimation methodologies for statistical samples,
- Description of major sources of error, including coverage of the target population, missing data effects, and measurement error, and
- A BTS point of contact for further questions and comments.

Guideline 6.8.2: Availability of Additional Documentation. To ensure the transparency of BTS information products, additional documentation (as specified in Chapter 2, Section 3.3, Section 4.6, Chapter 5, and Guideline 7.1.4) should be made available to customers upon request, unless such release would jeopardize confidentiality or disclose the actual methods used to protect the data.

Guideline 6.8.3: Reproducibility. Data users should be able to reproduce any publicly released information product to a reasonable degree of closeness. Information products that have been revised should clearly indicate the “as of” date.

Guideline 6.8.4: Archive Requirements. To ensure reproducibility within BTS, the product manager should establish criteria for retaining and archiving:

- All electronic product files,
 - Complete information products, whether paper or electronic, representing a specific continuing publication product or one-time report,
 - The data files and/or databases (at the most disaggregated level), which are used to generate publicly released information products, and
 - System and model documentation and computer software/programs used to generate any information product.
-

Related Information

Bureau of Transportation Statistics (BTS). 2005. *BTS Statistical Standards Manual*. Washington, DC.

Office of Management and Budget (OMB). 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*. Federal Register, Vol. 67, No. 36, pp. 8452-8460. Washington, DC. February 22.

U.S. Department of Transportation (DOT). 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Section 5.3 (Source and Accuracy Statements). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: May 4, 2005

CHAPTER 7

EVALUATING INFORMATION QUALITY

7.1 CONTINUING ACTIVITIES

7.2 DATA QUALITY REVIEWS

7.3 DATA EVALUATION PROJECTS

Ensuring data quality requires regular assessments of all aspects of data collection and processing and the implementation of corrective actions as appropriate. This can be accomplished by incorporating data quality checks within routine data collection processes and information product releases (Section 7.1), independently reviewing data products and data collection systems for standards compliance (Section 7.2), and targeting evaluations to diagnose and resolve serious data problems (Section 7.3).

7.1 CONTINUING ACTIVITIES

Standard 7.1: BTS information products and the processes that BTS uses to create them must routinely include an evaluation component.

Key Terms: information product

Guideline 7.1.1: Quality Assurance. All BTS information products and the processes that BTS uses to create them must routinely include:

- Process checks throughout data collection (Guideline 3.3.1), data processing (Section 4.5), and information dissemination (Section 6.7),
- Pre-dissemination review of information products (Guideline 6.1.3), and
- Measurement of performance (Guideline 4.5.1) and of information quality (Guideline 5.2.2).

Guideline 7.1.2: Periodic Quality and Performance Self-Assessment. Product and project managers should periodically conduct a self-assessment of the quality and performance of their products and processes.

- Assess quality and performance on an annual basis for products and processes that occur at least once a year. Assess less frequently for products and processes that occur less than once a year.
 - The assessment should highlight significant events that occurred during the past period, any events anticipated to occur during the next period, and identify strengths, weaknesses, and improvement opportunities.
 - Submit a summary of the self-assessment findings to the Director or the Director's designated manager.
 - Use assessment results to update internal and public documentation (Section 6.8) and to improve data quality.
-

Related Information

Bureau of Transportation Statistics. 2005. *BTS Statistical Standards Manual*, Chapters 3-6. Washington, DC. Available at http://www.bts.gov/programs/statistical_policy_and_research/bts_statistical_standards_manual/index.html as of July 29, 2005.

U.S. Department of Transportation. 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Chapter 6 (Evaluating Information Quality). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: September 23, 2005

7.2 DATA QUALITY REVIEWS

Standard 7.2: Independent statistical reviews are required of all BTS data collection systems (including those handling external-source data) that BTS uses to produce information products.

Key Terms: external source, information product

Guideline 7.2.1: Independent Review Team. An independent data quality review team should include:

- The Director or the Director's designee responsible for statistical methods and standards, who should establish the team,
- At least one person knowledgeable about BTS statistical standards but not involved in the data collection process, and
- At least one person familiar with the data collection process.

Guideline 7.2.2: Review Areas. The independent data quality review should focus on compliance with statistical standards (BTS 2005, OMB 2005) and with design specifications (Section 2.4). The review should include:

- The most recent self-assessment report (Section 7.1),
- The data collection design specifications (Section 2.4),
- An historical review of problems with identified by staff in collecting the data, the primary data users in applying the data to their needs, and data providers in reporting the data (Chapters 2 and 3),
- If sampling is used, a review of the sample design and the sample selection and maintenance processes (Sections 2.2 and 3.2),
- A review of data processing problems, such as problems in converting raw data files to databases, problems with lack of editing or edit resolution, and problems with missing data (Chapter 4),
- Review of the procedures for dissemination of data through various media, and of the source and accuracy information provided to users (Chapter 6),
- Verification that users can independently reproduce estimates, including sampling error estimates where applicable, contained in the information products coming out of the system (Sections 5.3 and 6.8), and
- Verification that documentation is accurate, complete (Sections 2.4, 3.4, 4.6, 5.3, and 6.8), and current.

Guideline 7.2.3: Review Outputs. Outputs from the data quality review should include:

- A report to the Director on review findings,
- A reply from the office responsible for the data collection system, and
- A quality improvement plan, prepared by the office responsible for the data collection system.

Guideline 7.2.4: Follow-up Review. A follow-up review should verify that the improvements have been implemented.

Related Information

Bureau of Transportation Statistics. 2005. *BTS Statistical Standards Manual*. Washington, DC. Available at http://www.bts.gov/programs/statistical_policy_and_research/bts_statistical_standards_manual/index.html as of July 29, 2005.

Office of Management and Budget (OMB). 2005. *Standards for Statistical Surveys (Proposed)*. Washington, DC. July 14.

U.S. Department of Transportation. 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Chapter 6 (Evaluating Information Quality). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: September 23, 2005

7.3 DATA EVALUATION PROJECTS

Standard 7.3: BTS should undertake a data evaluation project if analysis of the data reveals that key data elements fail to meet data quality requirements.

Key Terms: primary data user, secondary data user

Guideline 7.3.1: Evaluation Project Teams. An evaluation project team should report to the Director or to a designee with authority to allocate resources in support of the team's mission. The team members should include:

- A team leader who is not involved in the data collection process,
- Personnel selected for their expertise but not involved in the data collection process, possibly including non-BTS staff, and
- Personnel who are directly involved in the data collection process.

Guideline 7.3.2: Evaluation Plan. The project team should plan the evaluation as a type of data analysis (Guideline 5.1.2) that targets specific problems or issues in BTS data products. Solicit input from the following sources, with the greatest weight given to the primary users:

- Primary users for whom BTS designs information products, and normally include analysts in DOT, Congress, and BTS,
- Secondary users, including commercial interests and the general public,
- BTS data collection experts who can identify additional process quality issues, and
- Sponsored independent expert reviews.

Guideline 7.3.3: Conduct of an Evaluation Study. The major tasks in an evaluation study are:

- Specifying the processes contributing to the observed data quality problem,
- Identifying the root problems,
- Ascertaining solutions to the problems, and
- Drafting an improvement plan to address the problems identified. In some cases, the recommendations could lead to a redesign of the data collection system (Chapter 2).

Guideline 7.3.4: Implementation of the Study Recommendations. The project team should report the evaluation results and recommendations to the Director, or to the Director's designee (7.3.1). Upon concurrence, the office conducting the data collection is responsible for the implementation of the recommendations. A follow-up data evaluation should verify that the recommendations have been implemented and that the data meet quality requirements.

Related Information

U.S. Department of Transportation. 2002. *The Department of Transportation Information Dissemination Quality Guidelines*, Chapter 6 (Evaluating Information Quality). Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of January 19, 2005.

Approval Date: September 23, 2005

APPENDIX A

KEY TERMS

Note: Sections referencing each key term are indicated by square brackets.

accuracy [2.2, 5.2]: Accuracy refers to the closeness of an estimate to the value of the population parameter.

administrative data collection [3.2]: Administrative data are records produced in conjunction with the administration of a program, such as motor vehicle registrations. In addition to providing a source for the data itself, administrative records may also provide information helpful in the design of the data collection process (e.g., sampling lists, stratification information).

archive [6.8]: Archiving is the preservation of records or documents in long-term storage.

bias [2.3, 3.2, 4.3, 5.2]: Bias refers to a systematic deviation of an estimate from the value of the population parameter. In statistical estimation, bias exists when the expected value of an estimator does not equal the parameter that it is intended to estimate.

bridge estimates, bridge study [2.3, 5.2]: A bridge study defines the relationship between an existing methodology and a new methodology for the purpose of reconciling the estimates from both methods.

coding [4.4, 4.6]: Coding is the process of adding alphanumeric values to a data file either to convert text information to categories that can be more easily counted, tabulated, or analyzed, or to indicate case-level operational information such as missing data information.

collection instrument [2.3]: Collection instruments are devices, such as forms, survey questionnaires, file layouts, online computer entry screens, traffic sensors, etc., used to collect data.

confidential [6.3]: “Confidential” is a status accorded to information identified as sensitive by the authority (law) under which the information was collected. (The information is not classified “confidential” in a national security sense.) Confidential information must be protected and access to it controlled. *See also* **confidentiality, disclosure limitation, sensitive material**.

confidentiality [2.3, 3.1, 3.3, 4.1, 6.5]: The term “confidentiality” implies both a pledge used during data collection, which guarantees that the uses of the data will be limited to those purposes specified in an Information Collection Request (ICR), and the active implementation of administrative procedures and security protocols to protect confidential data from unauthorized disclosure. *See also* **confidential, disclosure limitation, Information Collection Request (ICR), sensitive material**.

coverage [2.2, 2.4, 3.2]: Coverage refers to the relationship between the elements on a list used as a frame and the target population units. Undercoverage errors occur when target population units are missed during frame construction, and overcoverage errors occur when units are duplicated or enumerated in error. *See also* **frame**.

crosswalk [2.3]: A crosswalk relates categories from one classification system to categories in another classification system.

data collection [3.3]: Data collection refers to all processes involved in acquiring data from a target population, including cases where previously gathered data are obtained from an external source.

derived data [4.6]: Derived data are additional unit-level data that are either directly calculated from other collected data or added from a separate data source. Assumptions may be used in deriving data. For example, flight distances can be derived from reported origin and destination airports by assuming that planes fly the most direct routes.

disclosure limitation [6.5]: Disclosure limitation involves techniques that are used to prevent the public release of individually identifiable data that were obtained under a pledge of confidentiality. *See also* **confidential, confidentiality, sensitive material**.

edit, editing [4.2, 4.4, 4.6]: Editing is the application of checks that identify missing, invalid, duplicate, or inconsistent entries, or otherwise point to data records that are potentially in error.

eligible unit [4.3]: An eligible unit is a unit that is in the target population. An eligible sample unit is a unit selected for a sample that is confirmed to be a member of the target population.

estimates [5.2]: Estimates are numerical values for population parameters based on data collected from a survey or other sources.

external source [3.1, 3.4, 4.4, 4.6, 6.2, 6.7, 7.2]: An external source is a data source over which BTS has little or no control during design, planning, and implementation of the data collection.

frame [2.2, 2.4, 3.2, 3.4, 4.5]: A frame consists of one or more lists and/or procedures that allow the target population to be enumerated.

imputation [4.2, 4.3, 4.4, 4.5, 4.6]: Imputation is a statistical procedure that uses available information and some assumptions to derive substitute values for missing values in a data file.

incident data [4.5]: Incident data consist of reports submitted only when a certain kind of event occurs, such as the release of a hazardous material during shipment.

inference [5.2]: Inference is the statistical derivation of information from data.

Information Collection Request (ICR) [3.3]: An ICR is a set of information, required by the Privacy Act, which is given to a data provider prior to the collection of any information. *See also* **confidential, confidentiality, disclosure limitation, sensitive material**.

information product [6.1, 7.1, 7.2]: An information product is any agency release of information to the public, regardless of physical form or characteristic. Printed reports, micro-data files, press releases, and tables posted on the web are all information products.

information security [4.1]: Information security refers to the safeguards, whether administrative or physical, in information systems or the building space that protect information against unauthorized disclosure and limit access to only authorized users in accordance with established procedures.

item [4.3]: An item is the smallest piece of information that can be obtained from a data collection instrument.

item nonresponse [4.3, 4.5]: Item nonresponse occurs when data are missing for one or more items in an otherwise complete report.

item response [4.6]: *See* **item nonresponse**.

key variable [2.3, 3.3, 5.1]: Key variables are data collection items for which aggregate estimates are commonly published. Key variables may include important analytic composites and other policy-relevant variables that are essential elements of the data collection.

longitudinal [4.5]: A longitudinal data collection is a series of repeated data collections on the same units over time. The data from a single unit on a single variable over time

constitute a time series. The analysis of interrelations within the time series is longitudinal analysis.

major data user [2.1]: *See primary data user.*

measurement error [2.3, 3.3]: Measurement error is the difference between observed values of a variable recorded under similar conditions and its actual value (e.g., errors in reporting, reading, calculating, or recording a numerical value).

metadata [6.4]: Metadata is descriptive information about a data file.

micro data [6.4]: Micro data are sets of unit-level records. A micro-data file includes the detailed responses for individual respondents.

missing at random [4.5]: A variable is missing at random if the probability that an item is missing does not depend on its value, but may depend on the values of other observed variables. A variable is missing completely at random if the probability that an item is missing does not depend on the values of any items, missing or not.

multivariate analysis [4.3]: Multivariate analysis is a generic term for many methods of analysis that are used to investigate relationships among two or more variables.

multivariate modeling [4.5]: Multivariate modeling is a method of analyzing the relationships between two or more variables by assuming some form of mathematical model, fitting the model, and statistically testing the model fit.

nonresponse bias [3.3, 4.3, 4.5]: Nonresponse bias is the impact on the observed value of an estimate due to differences between respondents and nonrespondents. The impact of nonresponse on a given estimate is affected by both the degree of nonresponse (missing data rates) and the degree that the respondents' reported values differ from what the nonrespondents would have reported (usually unknown).

objectivity [6.3]: Objectivity is the accurate, clear, complete, and unbiased presentation of information developed using sound statistical and research methods.

outliers [4.2]: An outlier is an isolated extreme high or low value, not necessarily erroneous, in a statistical distribution.

overall unit nonresponse [4.3, 4.5]: Overall unit nonresponse combines unit nonresponse across two or more levels of data collection, where participation at the second stage of data collection is conditional upon participation in the first stage of data collection.

peer review [6.1]: A peer review is an evaluation conducted by one or more technical experts independent of an information product's development.

population [4.5]: *See target population.*

precision [2.1, 6.6]: Numerical precision refers to the number of significant digits of numerical values. Precision of sample estimate refers to its reliability. *See also reliability.*

primary data user [7.3]: Primary data users are people or organizations who use information products, in either raw or aggregate form, and are identified in strategic plans and legislation that support the creation and maintenance of a data system. *See also secondary data user.*

probability of selection [4.3]: The probability of selection is the probability that a given population unit will be selected by a sampling process, based on the probabilistic methods used in sampling.

record layout [6.4]: A record layout is a description of the data elements in a file (variable names, data types, and length of space on the file) and their physical locations.

regulatory data collection [3.2]: A regulatory data collection is mandated by a regulation to provide information for regulatory purposes. In addition to providing a source for the data itself, regulatory data may also provide information helpful in the design of the data collection process (e.g., sampling lists, stratification information).

reliability [5.2, 6.3]: Reliability refers to the degree of consistency of an estimate, such as measured by its relative standard error.

reproducibility [5..3, 6.8]: Reproducibility is the ability to substantially replicate the disseminated information.

response rates [2.2, 2.3, 3.3, 4.3, 4.5, 4.6]: A response rate is to the proportion of the eligible units that is represented by the responding units.

revision [6.7]: A revision is a change made to a previously disseminated information product.

robustness [5.2]: The robustness of an estimator or analysis method is the degree to which the required calculations are insensitive to violations of their assumptions.

sample substitution [4.3]: Sample substitution refers to the practice of sampling matched pairs (in which the members of the pair do not have an independent probability of selection), and obtaining data from the second member only if the first member does not respond.

secondary data user [7.3]: Secondary data users are people or organizations who use information products, in either raw or aggregate form, but who are not identified in

strategic plans and legislation that support the creation and maintenance of a data system. *See also* **primary data user**.

sensitive material [6.1]: Sensitive material is information whose release would jeopardize confidentiality, privacy, or other guarantees given to data providers. *See also* **confidential, confidentiality, disclosure limitation**.

significant [6.3]: A result is statistically significant if a statistical test indicates, at a pre-specified probability level, that the result is unlikely to have occurred by chance.

significant digit [6.6]: A significant digit is a digit needed to express a number to within the uncertainty of measurement.

skip pattern [4.2, 4.4]: A skip pattern in a data collection instrument is the process of skipping over non-applicable questions depending upon the answer to a prior question.

standard error [6.4]: The standard error is the standard deviation (or square root of the variance) of an estimator. *See also* **variance**.

statistical map [6.2]: Statistical maps are depictions of geographically related data on maps using colors and symbols. Statistical maps are also known as thematic maps.

storage [4.1]: Storage refers to warehousing of project documents and/or data in a secure location.

target audience [5.1]: The target audience is the set of the data users that a particular information product is intended to serve.

target population [2.2, 2.4, 3.2]: The target population is the set of all people, businesses, objects, or events about which information is required.

time series [5.2, 6.3]: A time series is a series of values of a variable at successive times.

transparency [5.3, 6.8]: Transparency means possessing sufficient detail and clarity about data and methods to facilitate reproducibility. *See also* **reproducibility**.

trend [5.2]: A trend is a long-term change in the mean of a time series.

unit [4.3]: A unit in a data collection is the entity that provides the lowest level raw data, such as a household in a household survey.

unit nonresponse [4.3, 4.5]: Unit nonresponse occurs when a report that should have been received is completely missing or is received and cannot be used (e.g., garbled data, missing key variables).

unit response [4.6]: Unit response occurs when a report is received and contains usable data. *See* **unit nonresponse**.

variance [5.2, 6.3, 6.4]: The variance of a sample-based estimator is a measure of the degree to which estimates would vary about its mean if it were recomputed on successive, identically designed, samples of the population. The variance of a population is a measure of the degree to which individual values vary about the population mean. Technically, it is the expected value of the square of the difference between a random variable and its expected mean. *See also* **standard error**.

weight [4.3, 4.5, 4.6, 6.4]: Weights are relative values associated with each sample unit that are intended to correct for unequal probabilities of selection for each unit due to sample design. Weights most frequently represent the number of units in the population that the sampled unit represents. Weights may be adjusted for nonresponse.

weighted average [6.2]: A weighted average is a mean in which the components have non-negative weights that sum to one but are not necessarily equal.

Approval Date: October 5, 2005

APPENDIX B

BIBLIOGRAPHY

49 U.S.C. 111 as amended by the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users. P.L. 109-59.

Agresti, A. 1990. *Categorical Data Analysis*. New York, NY: Wiley.

American Association for Public Opinion Research (AAPOR). n.d. *Best Practices for Survey and Public Opinion Research*. Available at http://www.aapor.org/default.asp?page=survey_methods/standards_and_best_practices/best_practices_for_survey_and_public_opinion_research#best12 as of April 29, 2005.

_____. 1998. "Standard Definitions – Final Dispositions of Case Codes and Outcome Codes for RDD Telephone Surveys and In-Person Household Surveys." Available at <http://www.aapor.org/ethics/stddef.html> as of April 18, 2005.

_____. 2000. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Lenexa, Kansas: AAPOR.

Anderson, T.W. 2003. *An Introduction to Multivariate Statistical Analysis, 3rd ed.* New York: Wiley.

Box, G.P., Jenkins, G.M., and Reinsel, G.C. 1994. *Time Series Analysis: Forecasting and Control, 3rd ed.* New York: Prentice Hall.

Bureau of Labor Statistics (BLS). 2000. Standard Occupational Classification (SOC) System. Available at <http://www.bls.gov/soc/> as of November 15, 2004.

Bureau of Transportation Statistics (BTS). n.d. *BTS Web Software Guidelines*. Washington, DC.

_____. n.d. Aviation Support Tables. Office of Airline Information: Washington, DC. Available at http://www.transtats.bts.gov/Tables.asp?DB_ID=595&DB_Name=Aviation%20Support%20Tables&DB_Short_Name=Aviation%20Support%20Tables as of July 20, 2005.

_____. 2002. *Section 508 Compliance Plan, Version 1.0*. Washington, DC.

- _____. 2003. *BTS Excel Table Standards*. Washington, DC.
- _____. 2003. *BTS Guide to Style and Publishing Procedures*. Washington, DC.
- _____. 2004. *Confidentiality Procedures Manual*. Washington, DC.
- _____. *BTS Information Product Scoping Paper*. Washington, DC.

Bureau of Transportation Statistics Confidentiality Statute, 49 U.S.C. 111(k).

Casella, G. and Berger, R.L. 2001. *Statistical Inference, 2nd ed.* Belmont, CA: Duxbury Press.

Chatfield, C. 2003. *The Analysis of Time Series: An Introduction, 6th ed.* New York: Chapman and Hall.

Cleveland, W.S. 1993. *Visualizing Data*. Summit, NJ: Hobart Press.

Cochran, W.G. 1977. *Sampling Techniques, 3rd ed.* New York: Wiley.

Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002, P.L. 107-347, Title V.

Consolidated Appropriations Act of 2001. Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001. P.L. 106-554.

Cook, R.D. and Weisberg, S. 1999. *Applied Regression Including Computing and Graphics*. New York: Wiley.

Cressie, N. 1991. *Statistics for Spatial Data*. New York: Wiley.

Daniel, C. and Wood, F.S. 1980. *Fitting Equations to Data*. New York: Wiley.

DeGroot, M.H. 1989. *Probability and Statistics*. Reading, MA: Addison-Wesley.

Diggle, P.J., Liang, K.-Y., and Zeger, S.L. 2000. *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

Draper, N.R. and Smith, H. 1998. *Applied Regression Analysis, 3rd ed.* New York: Wiley.

Efron, B. and Tibshirani, R.J. 1994. *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Energy Information Administration (EIA). 1998. *EIA Guidelines for Statistical Graphs*. Washington, DC. Available at <http://www.eia.doe.gov/neic/graphs/preface.htm> as of April 19, 2005.

_____. 2002. *EIA Standards Manual*. Washington, DC. Available at <http://www.eia.doe.gov/msg/Standard.pdf> as of January 25, 2005.

Federal Aviation Administration (FAA). 2000. The Human Factors Analysis and Classification System—HFACS. DOT/FAA/AM-00/7. Office of Aviation Medicine: Washington, DC. Available at <http://www.hf.faa.gov/Portal/ShowProduct.aspx?ProductID=54> as of June 15, 2005.

Federal Committee on Statistical Methodology. 1990. *Survey Coverage*, Statistical Policy Working Paper 17. Washington, DC: Office of Management and Budget. Available at <http://www.fcsm.gov/working-papers/wp17.html> as of November 5, 2004.

_____. 1990. *Data Editing in Federal Statistical Agencies*, Statistical Policy Working Paper 18, Washington, DC: Office of Management and Budget. Available at <http://www.fcsm.gov/working-papers/wp18.html> as of November 15, 2004.

_____. 1994. *Report on Statistical Disclosure Limitation Methodology*, Statistical Policy Working Paper 22, Washington, DC: Office of Management and Budget. Available at <http://www.fcsm.gov/working-papers/spwp22.html> as of November 15, 2004.

_____. 1996. *Data Editing Workshop and Exposition*, Statistical Policy Working Paper 25, Washington, DC: Office of Management and Budget. Available at <http://www.fcsm.gov/working-papers/wp25a.html> as of November 15, 2004.

_____. 2001. *Measuring and Reporting Sources of Error in Surveys*, Statistical Policy Working Paper 31. Washington DC: Office of Management and Budget. Available at http://www.fcsm.gov/01papers/SPWP31_final.pdf as of December 20, 2004.

Fleiss, J.L. 1981. *Statistical Methods for Rates and Proportions*, 2nd ed. New York: Wiley.

Groves, R. 1989. *Survey Errors and Survey Costs*. New York, NY: Wiley.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. 2005. *Robust Statistics: The Approach Based on Influence Functions*, rev. ed. New York: Wiley.

Harvey, A.C. 1993. *Time Series Models*, 2nd ed. Cambridge, MA: MIT Press.

Hawkins, D.M. 1980. *Identification of Outliers*, New York: Chapman and Hall.

Hicks, C.R., and Turner, K.V. 1999. *Fundamental Concepts in the Design of Experiments*. Oxford, UK: Oxford University Press.

Hogg, R.V., Craig, A., and McKean, J.W. 2004. *Introduction to Mathematical Statistics*, 6th ed. New York: Prentice Hall.

Hosmer, D.W., and Lemeshow, S. 1989. *Applied Logistic Regression*. New York: Wiley.

Huber, P.J. 1981. *Robust Statistics*. New York: Wiley.

International Air Transportation Association (IATA). n.d. Airline Coding Directory. London, UK. Available at <http://www.iata.org/ps/publications/9095.htm> as of July 26, 2005.

International Organization for Standardization. 2002-2003. *ISO/IEC 11179, Information Technology -- Metadata Registries (MDR)*, (multipart standard). Available at <http://metadata-standards.org/11179/> as of January 25, 2005.

Interagency Household Survey Nonresponse Group. Information available at <http://www.fcsn.gov/committees/ihsng/ihsng.htm> as of April 18, 2005.

Kalton, G. 1983. Compensating for Missing Survey Data. Institute for Social Research, University of Michigan.

_____ and Flores-Cervantes, I. 2003. Weighting Methods, *Journal of Official Statistics* Vol.19, No.2.

_____ and Kasprzyk, D. 1982. Imputing for missing survey responses. *Proceedings of the Section on Survey Research Methods American Statistical Association, 1982*, 22-31.

_____ and Kasprzyk, D. 1986. The treatment of missing survey data. *Survey Methodology*, Vol. 12, No. 1, 1-16.

Kelsey, J.L., Whittemore, A.S., Evans, A.S., and Thompson, W.D. 1996. *Methods in Observational Epidemiology*. New York: Oxford University Press.

Kleinbaum, D.G., Kupper, L.L., and Muller, K.E. 1988. *Applied Regression Analysis and Other Multivariable Methods*. Boston: PWS-Kent.

Lehmann, E.L. and Romano, J.P. 2005. *Testing Statistical Hypotheses, 3rd ed.* New York: Springer Verlag.

_____ and Casella, G. 1998. *Theory of Point Estimation, 2nd ed.* New York: Springer Verlag.

Little, R.J.A. and Rubin, D. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.

McCulloch, C.E. and Searle, S.R. 2001. *Generalized, Linear, and Mixed Models*. New York: Wiley.

Mood, A.M., Graybill, F.A., and Boes, D.C. 1974. *Introduction to the Theory of Statistics*. New York: McGraw-Hill.

National Center for Educational Statistics (NCES). 2002. *NCES Statistical Standards, Standard 5.3 (Rounding)*. Washington, DC. Available at http://nces.ed.gov/statprog/2002/std5_3.asp as of January 25, 2005.

National Center for Health Statistics (NCHS). n.d. "The International Classification of Diseases, Ninth Revision, Clinical Modification" (ICD-9-CM). Available at <http://www.cdc.gov/nchs/about/otheract/icd9/abtcd9.htm> as of June 14, 2005.

National Institute of Standards and Technology (NIST). n.d. Federal Information Processing Standards Publications. Available at <http://www.itl.nist.gov/fipspubs/index.htm> as of November 15, 2004.

Office of Management and Budget (OMB). 2000. *Electronic and Information Technology Accessibility Standards*, Final Rule. Federal Register, Vol. 65, No. 246, pp. 80500-80528. Washington, DC. December 21.

_____. 2000. Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity. Available at <http://www.whitehouse.gov/omb/inforeg/statpolicy.html#dr> as of November 15, 2004.

_____. 2002. *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies*. Federal Register, Vol. 67, No. 36, pp. 8450-8460. Washington, DC. February 22.

_____. 2004. Paperwork Reduction Act Submission (Form OMB 83-I). Washington, DC. February. Available at <http://www.whitehouse.gov/omb/inforeg/83i-fill.pdf> as of June 15, 2005.

_____. 2004. Questions and Answers When Designing Surveys for Information Collection. Washington, DC. December 6.

_____. 2005. *Final Information Quality Bulletin for Peer Review*, Final Bulletin. Federal Register, Vol. 70, No. 10, pp. 2664-2677. Washington, DC. January 14.

_____. 2005. *Standards for Statistical Surveys (Proposed)*. Washington, DC. July 14.

_____. 2005. *Update of Statistical Area Definitions and Guidance on Their Uses*. Available at <http://www.whitehouse.gov/omb/inforeg/statpolicy.html#ms> as of July 15, 2005.

Pankratz, A. 1983. *Forecasting with Univariate Box-Jenkins Models*. New York: Wiley.

Pipeline and Hazardous Materials Safety Administration (PHMSA). n.d. "Hazmat Table." Office of Hazardous Material Safety: Washington, DC. Available at <http://www.myregs.com/dotrspa/> as of July 20, 2005.

Plain Language Action & Information Network. n.d. Writing User-Friendly Documents. Available at <http://www.plainlanguage.gov/handbook/index.htm> as of February 9, 2005.

Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, M., Martin, J., and Eleanor Singer. 2004. *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.

Privacy Act of 1974.

- Rao, C.R. 1973. *Linear Statistical Inference and Its Applications, 2nd ed.* New York: Wiley.
- Rohatgi, V.K. 1976. *An Introduction to Probability Theory and Mathematical Statistics.* New York: Wiley.]
- _____. 1984. *Statistical Inference.* New York: Wiley.
- Rousseeuw, P.J., and Leroy, A.M. 1987. *Robust Regression and Outlier Detection.* New York: Wiley.
- Rubin, D.B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.
- Särndal, C.-E., Swensson, B., and Wretman, J. 1991. *Model Assisted Survey Sampling.* New York: Springer Verlag.
- Schafer, J.L. 1997. *Analysis of Incomplete Multivariate Data.* London, UK: Chapman and Hall.
- Scheffé, H. 1959. *Analysis of Variance.* New York: Wiley.
- Searle, S.R., Casella, G., and McCulloch, C.E. 1992. *Variance Components.* New York: Wiley.
- Seber, G.A.F., and Lee, A.J. 2003. *Linear Regression Analysis, 2nd ed.* New York: Wiley.
- Selvin, S. 1996. *Statistical Analysis of Epidemiologic Data.* Oxford, UK: Oxford University Press.
- Skinner, C., Holt, D., and Smith, T. 1989. *Analysis of Complex Surveys.* New York: Wiley.
- Snedecor, G.W. and Cochran, W.G. 1989. *Statistical Methods, 8th ed.* Ames, IA: Iowa State University Press.
- Statistics Canada. n.d. Standard Classification of Transported Goods (SCTG). Ottawa, Canada. Available at <http://www.statcan.ca/english/Subjects/Standard/sctg/sctg-intro.htm> as of June 14, 2005.
- Stopher, P. and Jones, P., eds. 2003. *Transport Survey Quality and Innovation.* Oxford, UK: Pergamon.
- Sudman, S., Bradburn, N., and Schwarz, N. 1996. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology.* San Francisco: Jossey-Bass.
- Tukey, J. 1977. *Exploratory Data Analysis.* Reading, MA: Addison-Wesley.

U.S. Census Bureau. n.d. The North American Industry Classification System (NAICS). Washington, DC. Available at <http://www.census.gov/epcd/www/naics.html> as of November 15, 2004.

U.S. Department of Transportation (DOT). n.d. *Plain Language Resource Page*. Available at <http://www.dot.gov/ost/ogc/plain.htm> as of February 9, 2005.

_____. 2002. *The Department of Transportation Information Dissemination Quality Guidelines*. Washington, DC. Available at <http://dms.dot.gov/ombfinal092502.pdf> as of August 22, 2005.

Wolter, K.M. 1985. *Introduction to Variance Estimation*. New York: Springer Verlag.

Zacks, S. 1971. *Theory of Statistical Inference*. New York: Wiley.

Approved: October 5, 2005