

USDA Information Quality Guidelines for Statistical Information

These guidelines pertain to statistical information disseminated by USDA agencies and offices that is obtained from original data collections, administrative records, or compilations of data from primary sources, as well as estimates and forecasts derived from statistical models, expert analyses, or a combination of the two. The guidelines also apply to analyses and interpretation of statistical information.

Guidelines

To ensure the objectivity of statistical information disseminated by USDA, its agencies and offices should:

- **Clearly identify the objectives and uses of the statistical information that will be collected, compiled, or created.**

Clearly describe the objectives of the information collection, compilation, or creation activity based on the main uses of the data. Develop the objectives in consultation with key users.

- **Develop a design for conducting the information collection, compilation, or creation activity that is based on sound statistical methods or generally accepted professional and industry standards.**

Clearly articulate the study approach or data collection techniques. Define the concepts to be studied and the data to be collected using broadly understood concepts and definitions. Design research studies and data collections based on sound statistical methods or generally accepted professional and industry standards. Design the study or data collection to minimize respondent burden balanced against the need and value of the information to be obtained.

- **Conduct sample surveys and other data collections using sound statistical, survey, and data collection methodologies that are consistent with generally accepted professional standards for all aspects of survey or data collection design and implementation.**

As applicable, follow generally accepted statistical standards for sample frame development, sample design, questionnaire design and testing, data collection, analysis of sampling and coverage error, nonresponse analysis, imputation of missing data, weighting, and variance estimation. Follow the provisions, guidelines, and policies set forth in the Paperwork Reduction Act and other regulations related to the conduct of government surveys.

- **When compiling and using information from administrative data files and records or original sources, use reliable data from reliable data sources and maintain the integrity of confidential information.**

Data should be validated against other information where practicable. Assess sources

of error (e.g. bias in reporting, missing data) and correct as practicable before using. Maintain liaison with the provider of the data to understand concepts, definitions, and sources of error. Document the universe, definitions, and procedures underlying the collection of the data to facilitate editing or adjustment procedures required and to judge fitness for use. Implement continuous or periodic assessment of incoming data quality by using such techniques as edit rules, comparisons with other sources, and sample follow-up studies. When appropriate, address nonresponse using edit and imputation or weighting adjustment procedures. Use data that are timely in the context of the information collection activity.

- **When creating statistical estimates or forecasts that are derived from existing data sources using models or other techniques, use sound statistical methods.**

Use statistical methods that are appropriate to the data and analysis being conducted. Review the statistical techniques for conformity with accepted standards within the discipline. Test and review models to verify that the computer programs that were developed to implement the models conform to the stated objectives. Where appropriate, develop simulations to evaluate the success of the model in producing reasonable results. Document models to describe the goals and objectives of the models, the data sources used, and the methodologies and assumptions employed.

- **Include quality assurance processes as an integral part of all phases of data collection or compilation, processing, and analysis.**

Institute systematic quality control assessments of collection activities and processing to ensure quality standards are being achieved. Clearly define and document quality processes and their results. As practicable, document decisions to depart from planned design processes to inform users of potential impacts on the resulting data.

- **In analyzing and reporting results of a statistical activity, use sound analytical techniques, maintain objectivity and professionalism, and present the data impartially in ways that are easy to understand.**

Provide a clear explanation of methodologies used and assumptions made. Results should be validated against other information where practicable. Tables and graphs should present information clearly and be understandable with minimal explanation in the text. Where appropriate, subject the analysis to formal, independent, external peer review to ensure its accuracy and objectivity.

- **Conduct an appropriate quality evaluation of the statistical information before it is disseminated, considering the timeliness of the information.**

Evaluate the statistical information in light of the original objectives of the statistical activity, its accuracy, reliability, and timeliness. Review the data to minimize the chance that erroneous data are released and identify data of marginal quality (e.g., limitations). Check the data for consistency with previous occasions of the same data set or with other similar data. Conduct internal consistency checks to ensure the data values lie within reasonable bounds. Have knowledgeable subject-matter experts conduct reasonableness checks. Where appropriate, subject the statistical information to formal, independent, external peer review to ensure its accuracy and objectivity.

- **Provide transparent documentation of data sources, methods, and sources of error.**

Identify the sources of the data, while preserving confidentiality of individual persons, firms, or entities as required. Identify the methods used to obtain or derive the data, the definitions of the data items, limitations of the data, and measures of known errors and biases in the data. For estimates or forecasts derived from statistical models, describe the variables and parameters of the model or method, their relationships, and provide the mathematical specifications or equations of the model if possible. Describe the major assumptions underlying the model or method and how the parameters of the model were estimated or derived. Provide the results of formal evaluations of the model or method (if any) and the testing of its predictions against data of known accuracy (if any).

- **Disseminate influential statistical information with a high degree of transparency about data and methods to facilitate its reproducibility.**

Statistical information is considered "influential" if the agency determines that dissemination of the information will have or does have a clear and substantial impact on important public policies or important private sector decisions. Subject to ethical, feasibility, and confidentiality constraints, analytic results using influential statistical information must be "capable of being substantially reproduced," meaning that independent analysis of the original or supporting data using identical methods would generate similar analytic results, subject to an acceptable degree of imprecision or error. To meet the standard of reproducibility for influential statistical information, agencies must provide, subject to ethical, feasibility, and confidentiality constraints, transparency about how analytic results were generated: the specific data used, the various assumptions employed, the specific analytical methods applied, and the statistical procedures employed.

To ensure the utility of statistical information disseminated by USDA, its agencies and offices should:

- **Ensure that USDA statistical information products meet user needs.**

Identify the users of the information, stay informed of their information needs, and develop new statistical information and/or data products to meet those needs. Review statistical information products on an ongoing basis to ensure that they remain relevant and address current information needs. Discontinue statistical information products that are no longer needed.

- **Make statistical information disseminated by USDA widely available and easily accessible.**

Maximize the availability of USDA reports, data products, and other statistical information on agency websites and in electronic media. Provide data products in formats for downloading. Provides users with schedules for the release of statistical products.

- **Ensure that the statistical information disseminated by USDA is understandable.**

Tables and charts in statistical reports should be clearly titled and labeled so as to be largely understood without having to refer to accompanying text. Text statements must be supported by the data. Provide a comprehensive explanation of data sources, definitions and concepts, and methodology. Provide contact information for a knowledgeable person with each publication and each data set to allow feedback and questions from data users.

To ensure the integrity of statistical information disseminated by USDA, its agencies and offices should:

- **Ensure statistical information maintained by USDA is secure from unauthorized internal access or revision.**

Provide internal protection against unauthorized data access or revision by use of appropriate local area network (LAN), database management system (DBMS), application server, and web server security permissions. Manage retrieval of sensitive data and any data updates or modifications using additional ID/password technology on a restricted use basis, managed by central information technology (IT) staff. Limit physical access to servers to technical staff.

Data and information made available to outside customers through USDA websites should be protected by web server and web product technologies. No sensitive data should be made available through USDA websites. Data tables derived from sensitive or confidential data should be preprocessed within the agency and checked for quality and disclosure issues before release. Data and information made available through agency extranets should be protected by a combination of firewall, extranet server security, and extranet application software ID/password and security layers.

Ensure the confidentiality of individually identifiable information that was collected under a pledge or requirement of confidentiality. USDA is subject to various general and program-specific statutory requirements regarding confidentiality and integrity of information, including but not limited to:

- Privacy Act of 1974
- Freedom of Information Act
- Computer Security Act of 1987
- OMB Circulars A-123, A-127, and A-130
- Government Information Security Reform Act
- Paperwork Reduction Act

[Link to the Information Quality Guidelines main page](#)

We welcome your comments and suggestions about these pages. For information related to the USDA Quality Information Guidelines, please contact [Dr. Bette Fugitt](#), USDA Records Officer. For comments regarding the site itself: ociowebmaster@ociofc.usda.gov
http://www.ocio.usda.gov/irm/qi_guide/stats.html

Last Updated: 05.08.2002