

Session # 313 – Improving the Quality of Data Collected by Government Organizations
Section on Government Statistics
Wednesday, August 11, 2004: 8:30 am – 10:20 am

The Quality Framework: A Guide for Measuring Quality at the National Center for Health Statistics

Kenneth W. Harris, National Center for Health Statistics

Introduction

As a member of the Interagency Council on Statistical Policy (ICSP), The National Center for Health Statistics (NCHS) is one of about a dozen agencies classified as statistical Federal agencies.

NCHS is the Federal agency responsible for the collection and dissemination of the nation's vital and health statistics. To carry out its mission, NCHS conducts a wide range of annual, periodic, and longitudinal sample surveys and administers the national vital statistics registration systems.

Many of the statistical agencies work collaboratively on a number of projects and activities. In addition, they often have similar modes of survey operations, similar procedural guidelines, and similar concerns regarding quality measurement and assurance. Not surprisingly, these collaborative efforts and other parallel systems have led to numerous instances where some Federal agencies exchange and freely use procedures shown to be successful by their sister agencies.

The NCHS Quality Framework, then, is comprised of procedures developed inhouse by NCHS staff as well as those developed by other agencies that could be adapted for use by NCHS. This paper will describe aspects of the NCHS quality framework and some of the procedures and quality

enhancing activities conducted at NCHS. A detailed description of the framework and activities is under development.

Coverage

The quality control procedures and quality enhancing activities covered in this paper include the following major areas:

- 1) Development of Concepts and Methods
- 2) Planning and design of surveys and other means of collecting data
- 3) Collection of data
- 4) Production of estimates and projections
- 5) Processing and editing of data
- 6) Coding verification
- 7) Analysis of data
- 8) Establishment of review procedures
- 9) Dissemination of data

Development of Concepts and Methods

The statistical information provided by NCHS is critical in guiding actions and policies aimed at improving the health of the American people. NCHS ensures the utility of these data and its analytic reports by establishing standards for the development of concepts and for collecting and reporting data and information, and by using established statistical methods or developing new methodologies to meet the policy and

program needs of its sponsors, including the needs of federal legislation directed toward health improvement. For example, NCHS will regularly consult with data users and survey sponsors to ensure the relevance and appropriateness of collected data and public information, as well as to keep abreast of changes in data needs and to identify outdated concepts.

Concepts and definitions are clearly stated and related to their intended use. Data collection and statistical methods are fully described, statistically defensible and appropriate for the intended survey purpose.

Planning and Design of Surveys and Other Means of Collecting Data

As part of its commitment to quality and professional standards of practice, NCHS includes the use of modern statistical theory and practice in all of its technical work.

In the area of planning and design this encompasses the development and determination of survey requirements, general objectives, sponsors, major variables of interest, available budget, precision desired, survey schedule, geographic scope, collection mode, type of respondent, quality control and evaluation plans, the sampling frame, sample design estimation specifications, variance estimation specifications, and other quality measurement specifications.

Collection of Data

NCHS assumes responsibility for determining methods of data collection, while minimizing respondent burden and ensuring quality, for all NCHS operations, regardless of data collection mode, medium, respondent universe, or subject. In so doing, the NCHS seeks advice from data users and

from other professional and technical subject-matter and methodological experts.

Data collection is broadly defined as those activities and processes that obtain statistical data about the elements of a population, either directly through the mechanism of obtaining the needed information by contacting respondents representing the population elements or indirectly by using administrative records or other data sources. Examples of some of these activities include:

- Implementation of quality and performance measurement and process control systems.
- Minimizing unit and item nonresponse.
- Documenting all components of the data collection process to assure repeatability.

Production of Estimates and Projections

Survey estimation involves assigning values to unknown population parameters using information collected in the survey, and possibly from other sources. A direct survey estimator uses data collected on the variable of interest only from the time period of interest and only from sample units in the domain of interest. An indirect small-area estimator uses values of the variable of interest from a domain and/or time period other than the domain and time period of interest, and it may also incorporate information from other data sources, such as administrative records.

When using data from sample-based surveys to calculate direct survey estimates, NCHS will use sample weights appropriate for the sample design. Sampling weights may be

adjusted to reduce sampling or nonsampling errors.

Also, NCHS sometimes produces estimates for geographic areas or subgroups of the population for which direct sample-based estimates would have inadequate statistical reliability, necessitating the development of indirect small area estimates.

Processing and Editing of Data

The NCHS commitment to quality and professional standards of practice includes the goal of processing survey data to prepare the data as collected to meet the needs of end users.

The term "processing" encompasses a variety of activities, including administrative functions (such as parameter development and management information systems); post-collection processes (such as editing, imputation, data review and correction, data query, weighting, estimation and variance estimation, and application of disclosure avoidance rules); and support functions for collection technologies (such as mailout, check-in, data capture, and followup).

Coding Verification

Types of Quality Control

There are two types of quality control: process control and acceptance sampling. Process control is designed to measure the quality of a process and to determine when the process is "out of control" and needs changing in order to get back in control. Defective work resulting from the "out of control" process is seldom corrected. The process is brought back in control through retraining.

The Center's coding operations are evaluated by the second type of quality control, i.e., acceptance sampling. Coding assignments are made on the basis of work lots or batches. Each batch is then assigned to a production coder, who codes all of the records in the batch. The batch is then verified to determine whether it is "acceptable," that is, if it meets a previously established quality standard, or "unacceptable," in which case it is rejected and recoded.

Types of Verification

Virtually all of the Center's coding operations are verified under an independent verification system, as opposed to a dependent verification system. The difference in the two systems is that the verifier in an independent system codes a batch (usually a sample) without having access to the codes of the production coder; a dependent verifier knows what the production coder has entered. Although the independent system is usually more expensive, the additional cost is often justified by its greater accuracy in measuring quality. A number of studies have shown that dependent verifiers tend to be biased toward the work of the production coder. In some cases, the dependent verifier's type II error rate, i.e., agreeing with an incorrect code, was as high as 70 percent. On the other hand, the type I error rate, i.e., changing a correct code to an incorrect code, was almost non-existent.

Analysis of Data

NCHS uses sound analytical techniques to ensure objectivity in its statistical information products. This includes analysis directly used to produce data and estimates for NCHS products and analysis that indirectly affects the quality of data and

information, such as analysis performed during research to develop improved methodologies for sampling, data collection, data capture, processing, estimation; and analysis performed to evaluate the quality of NCHS data, methodologies, and processes.

Establishment of Review Procedures

The purpose of the review standard is to ensure that all reports, abstracts, papers, articles, and presentations prepared by NCHS staff are reviewed for technical content and are of high quality. This standard applies to all paper and electronic documents that will be released to the public, as well as documents put on list servers prepared by NCHS staff. It also includes work due by other organizations under contract as well as work that is done for NCHS.

Dissemination of Data

The principal functions of NCHS include the collection, compilation, analysis, and the dissemination of information. NCHS seeks to maintain its independence in the production of statistics; to preserve the confidentiality of information provided by respondents; and to maintain high quality standards for the information it disseminates to the public. These features foster credibility among data users and trust among data providers. High quality publicly available statistics on the health of the United States population produced by NCHS are essential for the nation to advance the health and the quality of life of its people.

Statistical information products disseminated to the public by NCHS must be accompanied with a statement describing their accuracy, their timeliness, their relevance, and their accessibility.

- Accuracy is the degree of closeness to the targeted value, using statistical measures of error.
- Timeliness encompasses frequency of data dissemination, as well as the closeness of the release to the data's reference period.
- Relevance is the degree to which the information products provide useful information for both current needs and anticipated future needs.
- Accessibility is the ease of access or effort needed for customers to acquire statistical data, products or services.

Summary

This paper has provided a general description of the Quality control and Quality enhancing activities conducted at NCHS. They cover a wide range of surveys, data systems, and other programmatic components for which NCHS is responsible. While not exhaustive, these procedures and activities provide a framework by which NCHS Quality can be assured.