Development of a Quality Framework and Quality Indicators at the Bureau of Labor Statistics

Scott Fricker, Michael Horrigan, Polly Phipps, Lucilla Tan

U.S. Bureau of Labor Statistics

**Abstract**

In the last two years, the U.S. Bureau of Labor Statistics has undertaken a project to develop a quality framework for its four price programs. This initiative has included extensive review of quality concepts and frameworks, and systematic assessment of the metrics BLS price program managers use to evaluate quality. We summarize findings from that effort, distinguish between product and process quality components, and present the resulting framework and associated quality metrics. We then illustrate the potential application of this framework to efforts to monitor and report on quality in the Consumer Expenditure Survey. The implications of this work for evaluating both product and process quality in statistical organizations are discussed.

1. **Introduction**

The U.S. Bureau of Labor Statistics (BLS) is the principal fact-finding agency in the Federal Government for the broad fields of labor economics and labor conditions. Its key outputs can be broadly categorized into four divisions – prices; employment and unemployment; compensation and working conditions; and productivity – with multiple programs within each division. There has been considerable work going back decades on developing quality metrics for the various BLS programs, and periodic efforts to develop Bureau-wide models and tools to assess quality. However, there currently exists no comprehensive or systematic quality-assessment framework within the Bureau.

In this paper, we describe recent efforts to develop a quality framework for the four BLS price programs: Consumer Price Index (CPI); Producer Price Index (PPI); International Price Program (IPP); and the Consumer Expenditure Survey (CE). This work was directed by the Associate Commissioner of the BLS Office of Prices and Living Conditions (OPLC), Dr. Michael Horrigan, in an effort to consolidate, optimize, and document quality procedures in these programs. In 2011, Dr. Horrigan initiated a series of meetings with OPLC senior managers to begin examining their quality practices and metrics, and asked two senior BLS survey methodologists to review and summarize quality frameworks used by U.S. and international survey organizations. The work described in this paper reflects iterative collaborations between these parties, and with technical experts in the CE branch tasked with developing a system for evaluating the impact of a redesigned CE on CE data quality.

In the sections that follow, we briefly summarize our review of quality frameworks, and present the framework adopted by the BLS price programs along with key quality metrics for their survey outputs and survey processes. We then describe a potential application of this framework for monitoring and reporting on data quality for the CE, and discuss broader implications for survey organizations considering implementing similar approaches.

1. **Quality Frameworks**

The BLS and other U.S. federal statistical agencies are strongly committed to producing quality statistics. But, what do we mean by “quality”? In this section we consider different conceptualizations of statistical and survey quality, reviewing two common quality frameworks adopted by statistical organizations. We also describe the standards and guidance issued by the U.S. Office of Management Budget (OMB), the agency responsible for overseeing and measuring the quality of federal agency programs, policies and procedures.

*2.1 Review of Quality Frameworks*

There are two major conceptual frameworks involving statistical and survey quality. The first is the total survey error (TSE) framework, which originates from the traditional statistical literature, and focuses specifically on the accuracy of survey data. The second framework includes multiple dimensions of quality that cover a statistical product’s fitness for use by clients and users and originates from total quality management (TQM) frameworks. The major difference between the two frameworks is multiple dimensions of quality and the focus on users that is found in quality frameworks.

Total survey error is a concept that is intended to describe the statistical error properties of survey estimates by incorporating all possible sources of error that may arise in the survey process. The total survey error framework focuses on survey data quality measured by accuracy or mean square error (bias and variance) of an estimate, with the objective of reducing survey errors critical to data quality and minimizing survey costs. There are slight differences in how the term ‘total survey error’ has been defined, but there is broad agreement on its major constituent elements. For example, errors often are grouped into two major divisions - sampling and nonsampling error.

In contrast to TSE approaches to quality, TQM frameworks focus on the “fitness of use” of statistical products by different groups of users when defining quality and identifying quality measures. This type of framework broadens the concept of quality into multiple dimensions. Accuracy is the most well-defined and quantified quality dimension. Other dimensions that users tend to prioritize include: relevance, timeliness, accessibility, interpretability, and coherence. Relevance is defined as producing information on the right concepts and utilizing appropriate measurement concepts within topic. Whether or not the information is timely and accessible to users are two additional quality dimensions. Interpretability focuses on the availability of concepts, variables, classifications, collection methods, processing and estimation to users so they can make their own assessments. Coherence involves how the information fits into broad frameworks; the use of standard concepts, variables, and classification methods; and if the information can be validated with related data sets.

TQM frameworks have been adopted in many national statistical offices (e.g., see Statistics Canada, 2002, for an early example), though there is variation across agencies and organizations. For example, in the U.S., the Interagency Council of Statistical Policy set out relevance, accuracy, timeliness, and dissemination/accessibility (statistical output measures), but also adds cost and mission achievement as conceptual dimensions of performance standards for U.S. federal statistical agencies. International frameworks include cost as well as respondent burden as statistical processes and consider the capacity to measure cost and burden important, in that it allows one to evaluate the trade off of costs to be balanced against the benefit of the output quality data.

Eurostat has developed an extensive quality framework based on the European Statistical System (ESS) standards. Documents set out specifications for assessing quality and performance, including what should be included in reports, specific product quality indicators, and measurement of process quality variables. In addition, user surveys, a self–assessment tool, auditing tools, as well as labeling and certification are addressed in the Eurostat Handbook (Ehling and Korner, 2007). The International Monetary Fund (2012) has a data quality assessment frame that is set out for major indexes, including the CPI and PPI. The dimensions include: integrity, methodological soundness, accuracy and reliability, serviceability, and accessibility. The dimensions are focused on the index as the statistical measure, but have been mapped to the ESS framework (Laliberte, Grunewald, and Probst, 2004).

Many agencies have gone into great detail to identify indicators and items that measure quality. Often, the steps of the survey process are used as part of the framework, and the survey steps can be very broad or detailed. Given the level of detail that agencies have set out, prioritizing quality measures is often useful; for example, the ONS identifies a short list of key product quality measures. Statistics Canada provides guidelines and quality indicators for each of 17 steps in a survey (2009). The Office of National Statistics (ONS) in the UK focuses on eight major categories (2013). In addition, Eurostat has set out a small number of specific quality measures for economic indicators, including the harmonized consumer price index and industrial product index (Mazzi et al, 2005).

*2.2 Quality Programs in Other U.S. Statistical Agencies*

The U.S. statistical system is a decentralized network of statistical agencies, but it operates under quality standards set by the Office of Management and Budget (OMB). In 2002, OMB issued final guidelines for ensuring and maximizing the quality of information disseminated by federal agencies”[[1]](#footnote--1). It used quality as an encompassing term to include utility, objectivity, and integrity. Utility referred to the usefulness of information to the intended users. Objectivity included whether the disseminated information was presented in an accurate, clear, complete, and unbiased manner, and the substance matter was accurate. Integrity involved security or protection of information from unauthorized access or revision, or information compromise through corruption or falsification. Agencies were to develop their own information quality guidelines; the resulting guideline documents from agencies did not include any type of performance measures. BLS addressed guidelines on a web page with discussion of the various guidelines and how they are met, as well as a section with specific guidelines on data integrity. The data integrity section discusses confidentiality, safety and security procedures, data collection, and dissemination[[2]](#footnote-0). All other federal statistical agencies developed guidelines, as well.

In 2006, OMB set out 20 standards and guidelines for statistical surveys covering the survey process: development of concepts, methods, and design; collection of data; processing and editing of data, production of estimates and projections, data analysis, review procedures, and dissemination of information products.[[3]](#footnote-1) Specific performance measures associated with these guidelines include nonresponse bias analysis when unit nonresponse is below 80 percent or item response is below 70 for any items used in a report, or coverage bias studies when coverage rates fall below 85 percent.

The U.S. Census Bureau was the only agency we identified that developed documentation on these standards to provide additional guidance on their programs and activities and cover their unique methodological and operational issues.[[4]](#footnote-2) They align the standards with the utility, objectivity, and integrity dimensions set out by OMB. The document includes detailed and thorough guidelines, definitions, and requirements for all of the activities, techniques, procedures, and systems associated with each stage of the survey process. In general, the document does not include performance measures.

1. OPLC Quality Framework

Based on this review of quality frameworks, a TQM-based approach seemed most appropriate for the BLS price programs. It encompasses all of the major quality dimensions of survey products, and there are a number of well-developed quality frameworks and metrics in use by other statistical organizations (e.g., Stats Canada, Eurostat) that we leveraged for OPLC effort. Discussions with the price program managers revealed, however, that many of the metrics that they regularly monitor are focused less on the quality of the survey products or outcomes itself, and more on their business processes. Recognizing that product quality is determined in large part by the quality of procedures that generate it, and wanting to incorporate the subject-matter expertise of production managers and staff, we adopted an approach that extended the basic TQM framework by integrating quality metrics for survey products as well as survey lifecycle processes, and then tied both back to the TQM-quality dimensions. We developed an initial framework and proposed quality metrics for each dimension, and then solicited feedback from program managers and technical staff (e.g., suggestions for other metrics, indications of areas where they currently do not have well-developed metrics, etc.). Figure 1 illustrates the resulting quality framework. It lists the key quality dimensions, their definitions, and some associated quality metrics for both product and process quality. (Due to the space limitations of this paper, however, we have omitted a number of features and quality metrics.)

1. **Applications in the CE**

As we were developing the OPLC quality framework, CE was in the planning phases of a major redesign, and searching for a metrics to assess the impact of design changes on data quality. Several projects were undertaken to evaluate existing quality metrics and reports in CE, to propose new metrics for monitoring quality, and to establish new procedures for developing, implementing, and reporting those metrics. These efforts were guided by early versions of the OPLC quality framework and a survey process model (similar to the Generic Statistical Business Process Model or GSBPM) that helped identify and organize quality metrics associated with the different dimensions of quality and relevant to the routine survey activities within CE production.

Several important lessons were learned from the work undertaken to accomplish these tasks. First, the use of a single, comprehensive, and flexible framework to create, organize, and maintain all quality-related information for the survey promotes transparency and efficiency. Second, before proposing any quality metrics, it is important to first identify and understand the issues/risks associated with it; once these issues/risks have been identified and understood, the metrics or monitoring methods generally tend to “suggest themselves.” Often this requires the involvement of staff working on field and data processing procedures because they best understand and can foresee potential risks, but this can add time to the development process. Third, it is useful to adopt a standardized format, or template, to describe metrics at the point of their proposal. This not only ensures that all the necessary information to define, produce and interpret a metric is thought through and documented – i.e. effectively capturing monitoring metadata “upstream” - but it also promotes transparency in the metric’s definition and will assist in its consistent interpretation over time (e.g., see Figure 2). Having this type of documentation about a metric minimizes dependency on individual staff, helping to mitigate the impact of staff turnover. Finally, any monitoring method or metric implemented must demonstrate its usefulness - i.e., tied to a specific concern(s) and not just be “nice to know”.

The CE applications were grounded in a multidimensional OPLC quality framework that was built around CE’s survey operations, and which leveraged the expertise of staff who perform these specific survey activities. The motivation for this approach stemmed from the recognition that identifying appropriate quality monitoring methods or metrics requires familiarity with a survey activity to understand the issues associated it. In addition, there is a high cost in terms of resources and commitment to establishing the information base necessary to produce reports on quality. In order for this effort to be sustainable, the benefits from it must be relevant and useful to survey operations, apart from providing external data users information about quality to help them assess fitness of use of the CE survey products for their applications.

**5.0 Conclusion**

The OPLC quality framework and its applications in CE are very much in the spirit of the approach taken by Statistics Sweden to develop quality indicators (Biemer and Trewin, 2012), and serve as an initial step to building the information bases necessary to perform quality evaluations for all BLS price programs. The objectives such a framework attempts to achieve are to: (1) identify (potential) sources of error; (2) develop awareness of the relative risks of these errors, and quantify them where possible; (3) identify gaps in methodology; (4) promote the use and integration of monitoring information; and (5) help prioritize areas that need improvement in survey operations.

Broader implementation of this framework will necessarily be an iterative process (starting out with one survey stage, then scaling up to include more survey stages) as we will need to learn how to operationalize the general principles and guidelines proposed, understand what level of detail of quality monitoring is useful, and demonstrate if and how this approach is useful to the managing and monitoring quality for survey operations.

References

Biemer, P., and Trewin, D. (2012). Development of Quality Indicators at Statistics Sweden. Report dated January 9, 2012 submitted to Statistics Sweden.

Ehling, M. and Korner, T. (2007). Eurostat Handbook on Data Quality Assessment Methods and Tools.

Eurostat (2009). European Statistical System (ESS) Standard for Quality Reports.

Gonzalez, J., Hackett, C., To, N., and Tan, L. (2009). “Definition of Data Quality for the Consumer Expenditure Survey: A Proposal” (internal BLS memorandum)

International Monetary Fund (2012). IMF Data Quality Assessment Framework.

Laliberte, L., Grunewald, W., and Probst, L. (2004). “Data Quality: A Comparison of IMF’s Data Quality Assessment Framework (DQAF) and Eurostat’s Quality Definition.

Mazzi, G., Museux, J., and Savio, G. (2005). “Quality measures for economic indicators” Eurostat Working Papers and Studies.

Office for National Statistics (2013). Guidelines for Measuring Statistical Output Quality, Version 4.1.

Statistics Canada (2002). Statistics Canada’s Quality Assurance Framework.

Statistics Canada (2009). Statistics Canada Quality Guidelines (5th ed.), Ottawa, Canada.

**Figure 1. OPLC Quality Framework and Select Quality Metrics**

**Figure 2. Example Quality Metric Documentation Template**



1. <http://www.whitehouse.gov/omb/fedreg_final_information_quality_guidelines>. [↑](#footnote-ref--1)
2. <http://www.bls.gov/bls/quality.htm>, <http://www.bls.gov/bls/data_integrity.htm> [↑](#footnote-ref-0)
3. <http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf> [↑](#footnote-ref-1)
4. <http://www.census.gov/quality/standards/index.html> [↑](#footnote-ref-2)