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

Michael Horrigan, Polly Phipps, Scott Fricker
Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Washington, DC 20212

Abstract

In the last two years, the Office of Prices and Living Conditions (OPLC) of the U.S. Bureau of Labor Statistics has undertaken a project to define a quality framework for its four price programs. This initiative began with an extensive review of quality concepts and frameworks; we summarize findings from that effort, distinguish between product and process quality, and review associated quality measures. Subsequently, OPLC program managers reported measures and variables they track as part of their effort to evaluate quality. What emerged was a telling divide between the relatively greater emphasis in the quality literature on output quality and the reduction of total survey error, and the more balanced tracking of both output quality measures and detailed operational measures of the quality of survey business processes – the inputs to the development of high quality outputs. Based upon this realization, OPLC developed a consensus framework for its four programs that combines both input (process) and output (product) quality measures. What emerged from this effort is a framework that adopts a hybrid of statistical process control approaches to quality with the more traditional quality management (QM) frameworks used across many statistical agencies.

Key Words: total survey error, quality management, statistical process control

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 (e.g., Dippo, 1997). However, there currently exists no comprehensive or systematic quality-assessment framework within the Bureau.

This paper summarizes an effort undertaken by the Office of Prices and Living Conditions of the Bureau of Labor Statistics to create a quality framework for its four survey programs: The Consumer Expenditure (CE) Survey, the Consumer Price Index (CPI), the Producer Price Index (PPI), and the International Price Program (IPP). This project was directed by the Associate Commissioner of the BLS OPLC, Michael Horrigan, in an effort to consolidate, optimize, and document quality procedures in these programs. The work was motivated by a recognition that each program had adopted numerous quality indicators for both their internal processes and for their respective outputs but there were inconsistences in both coverage and depth of coverage across the programs. Prior to developing a common

framework for both internal production processes and for the quality of outputs, OPLC contacted the BLS Office of Survey Methods Research (OSMR) and asked for a briefing on the types of quality frameworks that exist across the U.S. statistical system, the international statistical community, and the private sector organizations that are engaged in data collection and dissemination. Following the extensive review of extant processes provided by my OSMR co-authors to this paper, Scott Fricker and Polly Phipps, OPLC management developed a consensus framework of quality measures to guide the evaluation of both internal survey business processes and the statistical outputs produced by each program. This paper reports on this effort. Section 2 provides a summary of the OSMR review of the quality literature. Section 3 reports on the deliberations of the four OPLC programs in reaching consensus and Section 4 summarizes that consensus framework and key quality metrics that was the result of those deliberations.

2. 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 has its origins in the Total Quality Management (TQM) movement and other quality management (QM) frameworks. The major differences between the two frameworks are in their specifications of quality dimensions and the greater emphasis on data user needs that is found in QM approaches. (See Lyberg, 2012 for a more thorough treatment of TSE and QM principles, and historical perspective on the evolution of notions of survey quality.)

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, QM 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.

QM 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 tradeoff of costs to be balanced against the benefit of the output quality data.

Eurostat has developed an extensive quality framework and Code of Practice 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 Office of National Statistics (ONS) in the UK 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 ONS 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). It also has issued guidelines for improving the quality of survey production processes, through the development of process flow charts, and the identification and monitoring of critical process variables within the different stages of the survey lifecycle.[1] These guidelines extend earlier work by Biemer and Caspar (1994) and Morganstein and Marker (1997), and provide techniques for analyzing variations in operational processes over time in order to improve the capability of those processes.

[1] http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/HANDBOOK%20ON%20IMPROVING%20QUALITY.pdf

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".[2] 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.[3] 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.[4] 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.[5] 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.

3. Deliberations on a Quality Framework for OPLC

Having benefitted from the literature review and briefings provided by OSMR, OPLC management set about the task of developing a consensus framework. The first step was to develop a detailed accounting of the quality measures adopted by each program to determine areas of commonality and difference. This was accomplished through a series of detailed memos and briefings given to OPLC management on the quality measures adopted by each program.

As the briefings indicated, there are a number of quality measures that are in common and a number that are distinct across the four OPLC programs. As well, there are some very unique measures in the three price programs owing to the nature of price statistics. For example, as described next there are measures that examine the options exercised when a

[2] http://www.whitehouse.gov/omb/fedreg_final_information_quality_guidelines.
[3] http://www.bls.gov/bls/quality.htm, http://www.bls.gov/bls/data_integrity.htm
[4] http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf
[5] http://www.census.gov/quality/standards/index.html

good or service being 're-priced' at an establishment has changed in quality or is no longer available for pricing. A new model year car, discontinued clothing lines, and imported items that change country of origin are just three such examples of the unique challenges faced by price programs in attempting to reprice the exact same good or service from one month to the next. In addition, the fourth program, the CE survey, is a household survey, which introduces several unique elements related to item and survey non-response that are not present in the establishment based price programs.

One of the unexpected benefits of this sharing of information was the identification of data capture practices by each survey that were of significant interest to other programs. Nowhere was this truer than in the presentation by CPI of their 'substitution monitor'. One of the critical features of collecting price data are the options that are exercised when the item in the market basket selected for pricing each month has changed in content or is no longer available for pricing. The options selected by data collectors are governed by detailed data collection rules and procedures and may involve determining that the change in the item's content is not price determining (such as a change in color), estimating the value of the quality change in cases where the change is substantial (and deciding on which of several quality change methods is most appropriate to use), substituting to another similar item (and quality adjusting between the two items), deciding that the item is temporarily unavailable (and using methods of pulling the previously recorded price forward), or discontinuing the item and starting a new price series for a new item. The CPI records every option taken for every item and examines trends in these options by type of item. The PPI and IPP programs have since adopted versions of this substitution monitor.

Another unexpected aspect of this sharing of quality measure approaches, one that had a significant aspect on our final consensus quality framework, was the clear divide that existed between PPI and IPP on the one hand, with their greater relative emphasis on business process measures, and the CPI, with its greater relative emphasis on output related measures. All three surveys had aspects of both, it was the relative emphasis and the number of measures in each area that was unexpected. IPP produces an extremely detailed monthly report on process measures governing every aspect of their business process. As an illustrative example, one of the quality measures is "100% of all TSUSA/B-to-SCG mappings and TSUSA/B-to-secondary mappings are correct prior to running frame edits during the third week of February and September each year." PPI has a similar emphasis on process measures, although not quite as detailed. While CPI has several process measures, they are also not as detailed as the ones tracked by IPP or PPI. In developing a consensus framework, a decision was made to include a separate section on business process measures, but as will be seen in the next section, we settled on a set of measures that have greater commonality across the three price programs, while acknowledging that some programs, especially IPP, may choose to track process measures that are at a much finer level of detail than other programs.

In contrast to IPP and PPI, CPI also has a relatively greater number of measures related to the quality of their outputs. One significant reason for this difference between CPI and PPI/IPP is that the former produces annual estimates of standard error for each index, while at that point, neither PPI nor IPP constructed any measures of variance for its indexes. Indeed, part of the path toward achieving a consensus framework was the very active work both PPI and IPP were doing to construct such variances in the future, and the types of quality measures related to survey error from CPI were of substantial interest.

The CE survey, as a household survey, is distinct from the three price programs in many respects. One aspect that became immediately apparent, was the fact that CE's guiding principle in its development of a quality framework was the explicit adoption of the total survey error measurement approach. There are a number of factors contributing to CE's adoption of the goal of minimizing total survey error. Key among these is the fact that underreporting of consumer expenditures for certain items (especially items related to personal habits such as expenditures on tobacco or liquor) are known to be underreported relative to estimates derived from independent business sales data. As well, the CE is undertaking a major survey redesign and the verifiable reduction in survey error is a guiding principle of the redesign effort (along with the reduction in respondent burden). This emphasis on survey error and the concern over underreporting has led CE to emphasis particular aspects of quality that do not have the same emphasis or priority in the three price programs. This is especially true in the efforts made by CE to examine the ***coherence*** of their estimates relative to other surveys and administrative data sources, both in terms of the levels of consumer expenditures and the distribution of expenditures by item type.

As with any quality measure, drawing such conclusions on relative emphasis of certain quality measures over others is often a matter of degree. In the price programs, without the existence of significant alternative measures of producer or consumer prices, relatively more attention is paid to the behavior or price indexes in relationship to other price indexes (for example, does a volatile price behavior of an input to production get reflected partly in the volatility of products for which that input has a significant value relative to the total cost of production) or to the behavior of price indexes over time. That being said, CPI has recently embarked on an interesting set of projects to compare the distribution of items selected for sampling with the distributions implied by scanner data for the same items, a coherence check of its internal sample selection process.

4. Toward a Common Quality Framework

To arrive at a consensus framework, we relied on asking programs to react to straw man versions of a quality framework that attempted to include measures that were consistent with what we learned from both the literature review and the sharing of practices across the programs (and subsequent discussion). Early on we decided to develop a two part system, one emphasizing the quality of outputs, and a second emphasizing the business processes that give rise to these outputs. The output quality measures were divided into 14 separates categories: (1) Relevance; (2) Accuracy and reliability – sampling errors; (3) Accuracy and reliability- non-sampling errors; (4) Coverage errors; (5) Measurement errors; (6) Processing errors; (7) Revision errors; (8) Modelling errors; (9) Timeliness; (10) Accessibility; (11) Interpretability; (12) Coherence; (13) Cost; and (14) Credibility, Integrity, and Confidentiality.

The business process framework was based on the following major business processes that are relevant for both the three price programs and CE: (1) Design/concepts; (2) sampling; (3) data collection; (4) estimation; (5) weight updates; and (6) dissemination.
In many respects, the specification of quality output measures was immediately more accessible given the relative emphasis on quality in the academic literature and the greater commonality across the programs either in the measures they adopt, or in the case of variances, the measures they wanted to adopt in the future. As mentioned above, the business process measures varied greatly in their relative degree of detail, so the consensus

you will see below aimed to find as much common ground as possible, allowing for program specific more detailed measures to exist.

As we began specifying our quality measures, one aspect of the Fricker/Phipps literature review that became more readily apparent was the need to develop measures that relate to user needs, measures that were not explicitly mentioned in any of the program specific quality measures. This is not to say that the individual programs were not attempting to measure how well they met user needs (although practices do vary), it is more that they had not included such practices in their own program frameworks. This latter point was particularly in evidence with measures of timeliness and accessibility.

Finally, as the two exhibits given below indicate, there is some natural overlap between the specific output and business process measures that we adopted. We had a number of discussions trying to develop mutually exclusive measures across the two paradigms but in the end we decided to allow for such duplication for the simple reason that each paradigm has unique quality measures that are not found in the other paradigm as well as having measures that are in common. These approaches were found to each have value added in terms of the management of our programs and offer a different perspective as to the nature of quality in our survey processes and outputs.

Table 1 displays the output quality measures from the OPLC consensus framework; Table 2 presents the business process quality measures. These tables list the key quality dimensions, their definitions, and the associated quality metrics for both product and process quality.

Table 1. OPLC Output Quality Measures

Quality Dimension	**Definition**	**OPLC Quality Measures**
1.Relevance	Do the data satisfy user needs? Is the program producing information on the right concepts	Program provides a narrative statement on the relevance of their data. This includes a description of the measurement objective of the program, data produced, classification systems used, and the coverage of the data product updates Program gathers feedback from data users on their satisfaction and their data needs through a variety of feedback mechanisms on a regular basis Program has a written description identifying known gaps in meeting user needs and priorities for closing those gaps. This document is updated on a regular basis.

Table 1 continued

Quality Dimension	Definition	OPLC Quality Measures
2a. Accuracy and Reliability – Sampling Errors	Is the program minimizing total random survey error?	Standard errors Coefficient of variation Relative standard errors Confidence Intervals
2b. Coverage Errors	Coverage errors arise due to omissions, duplications, erroneous inclusions, and content errors. Does the program attempt to measure the closeness between the covered and targeted population?	Out of scope rate Misclassification rate Under and over-coverage rate
2c. Nonresponse Errors	Nonresponse errors result from a failure to collect complete information on all units in a selected sample. Does the program calculate both unit and item response rates?	Unit response rates (weighted and unweighted) Key item response rates % of final weight from non-imputed cells for key estimation items Imputation rates
2d. Measurement Errors	Measurement errors occur when the response provided differs from the real value, and may be attributable to the respondent, interviewer, questionnaire, or collection method. Does program attempt to compare results to other sources?	Benchmark Comparisons Substitution process The scope of this include: - Substitution rates - Percent of items linked out - Percent of items quality adjusted - Percent of items where no replacement is found
2e. Processing Errors	A processing error is the error in final survey results arising from the faulty implementation of correctly planned implementation methods. For example, does the program track the number of coding or data entry errors discovered as a result of a reinterview?	edit failure rates coding error rates data entry error rates

Table 1 continued

Quality Dimension	Definition	OPLC Quality Measures
2f. Revision Errors	Does the program track the number and relative size of revisions between initial and final publication?	Relative size of revisions in key estimation concepts between first and final publication
2g. Modeling Errors	Model errors occur with the use of methods, such as calibration, generalized regression, seasonal adjustment and other models not included in the preceding accuracy components, in order to calculate statistics or indexes. Does the program attempt to understand the effect of errors associated with their models?	Goodness of fit statistics for seasonally adjusted series
3. Timeliness	Are survey estimates reported in time to maximize their usefulness? Are data and related documentation released on schedule?	Program provides a narrative describing their product lines, their periodicity, and an overall assessment of timeliness. Number of news releases released on time each year, according to a pre announced schedule Lag between the end of data collection and the publication of estimates
4. Accessibility	Does the program provide user-friendly data extraction tools? Is access to the data affordable?	Program provides a narrative on the accessibility of their data and products. User assessments of accessibility Number of subscribers to program publications Number of information requests recorded monthly Number of web hits overall and by product, monthly

Table 1 continued

Quality Dimension	Definition	OPLC Quality Measures
5. Interpretability	What is the quality and coverage of survey documentation explaining survey concepts and methods?	Program provides a narrative describing their documentation that is available on concepts and methodology User assessments of interpretability
6. Coherence	What is the degree to which data derived from different sources but measuring the same phenomena, are similar to the estimates generated by the program?	Program conducts on-going studies comparing its key estimates to estimates derived from alternative sources.
7. Cost	What is the cost of collecting, processing and disseminating data?	Direct and fully loaded costs per initiated unit (PPI/IPP/CPI C&S/CPI housing/CE Diary/CE Interview) stratified by mode of collection Direct and fully loaded costs per reinterviewed unit (PPI/IPP/CPI C&S/CPI housing/CE Diary/CE Interview) stratified by mode of collection Number and distribution of respondents by data collection method (PPI/IPP/CE Diary/CE Interview)
8.Credibility, Integrity, Confidentiality	To what extent are the estimates produced by the survey program viewed as being credible? To what extent is the program and the underlying institution viewed as having integrity? Does the program safeguard the confidentiality of its respondents?	# data breaches reported each quarter or year Number of mistakes found in published numbers after release

Table 2. OPLC Business Process Quality Measures

Quality Dimension	Definition	OPLC Quality Measures
1. Design/Concepts	This business process specifies the principal estimation objectives of the survey program and how these objectives measured through the data collection instruments and processes.	Program provides a narrative on the estimation objective(s) of its survey program and the conceptual / theoretical basis for that objective.
2. Sampling	This business process identifies the universe frame(s) used for selecting the sample for data collection, the process of selecting the sample, and processes and periodicity for sample reselection (as appropriate) and methods for adjusting for sample attrition over time.	Program provides a narrative statement describing the relevant universe frames, the sample selection process, and methods for adjusting for sample attrition over time Average age of sample overall and by relevant characteristics such as industry and products, demographics, etc <u>IPP/PPI Only</u> Comparison of targeted versus actual sample sizes
3 Data Collection -	<u>CPI, PPI, IPP</u> In the BLS Price Programs, this business process describes the collection of the first set of prices collected at an establishment, including the disaggregation steps conducted by the Field Economist to get to the exact goods and services being priced, and the repricing of those items. In the CE program, this business process includes the collection of data for the 5 quarterly interviews in the CE Household Survey and the collection of data for the 2 week CE Diary.	<u>All</u> Unit response rates (weighted and unweighted) Estimates of the average length of an interview by data collection method Number and distribution of respondents by data collection method <u>CPI/PPI/IPP Specific</u> Unit response rates (weighted and unweighted) for initiation and repricing Substitution Rates Quality adjustment rates by method

Table 2 continued

Quality Dimension	Definition	OPLC Quality Measures
4. Estimation	This business process includes the variety of steps needed to translate the data captured through the data collection and develop estimates for publication. These steps include editing, imputation, weight adjustments, seasonal adjustments, outlier analysis, and preparation of estimation files and review and final sign off	Program provides a narrative describing its estimation methodology Imputation rates % of final weight from non-imputed cells for key estimation items Program conducts on-going studies comparing its key estimates to estimates derived from alternative sources. <u>IPP/PPI Only</u> Substitution rates Relative size of revisions in key estimation concepts between first and final publication
5. Weight Updates	This business process describes the incorporation of new universe weights that have been updated (such as through the administration of a new Decennial Census of the Population or new Quinquinniel Census of Establishments, among others).	Program provides a narrative describing its weighting plan and how weights are updated
6. Dissemination	This business process describes the posting of survey press releases, data bases, analysis of data, survey methodology documentation, among others, and the ways in which the data user community can access and use our published materials.	Program provides a narrative on the products and services it provides to the public. User assessment of the effectiveness of dissemination Summary of on-time performance for release of data series and micro-data files # of data series published Number of web hits overall and by product, monthly

5. Discussion

The quality framework developed for the BLS price programs is very much in the spirit of the approach taken by Eurostat, Statistics Sweden, and other statistical agencies outside the U.S. to develop quality indicators for their surveys. It adopts a multidimensional model of quality in line with QM-based approaches, and defines quality measures for both product outputs and survey business processes. The objectives of this framework 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.

In developing this framework, OPLC incorporated many features found in existing approaches to survey quality assessment and leveraged the expertise of their staff who manage and carry out specific survey activities. We offer several observations about the development process and the resulting framework. First, the diversity of OPLC programs – which include both household and establishment surveys, different types of business processes, and program-specific quality measures that vary in number and level of detail – made it challenging to create a single, comprehensive framework. Iterative refinements were necessary to build adequate flexibility into a consensus approach suitable across programs, and additional work will be needed to operationalize a fuller set of quality measures and to develop program-specific guidelines for those metrics. Second, as noted above, there is considerable overlap between the product and process quality measures listed in Tables 1 and 2, and relatively few process measures that are available in real-time (or near real time), which limits the types of monitoring and intervention activities one associates with typical statistical process control approaches.

This presentation of the consensus framework, which emphasizes common quality dimensions and measures across the four programs, however, masks more detailed data and procedures that are being used to assess quality in two OPLC programs. For example, the IPP program regularly publishes internal quality reports that track progress on numerous process quality targets (e.g., percent of items repriced, percent of published strata, timeliness of revisions and publications, etc.). Similarly, the CE program has developed a set of metrics to assess the impact of design changes on product and process quality (based on contact history information and survey estimates for sample units), and proposed new procedures for developing, implementing, and reporting those metrics. Thus, a third observation is that the programs' existing quality-assessment infrastructure and activities will affect the speed with which it can integrate and advance the overall quality framework presented here. Broader implementation of this framework within OPLC will be an ongoing process as its programs continue to develop their quality initiatives. We also hope that other BLS programs will adopt and help refine this quality framework. Although there are challenges to such an effort (e.g., organizational decentralization, the added complexity of accommodating diverse program needs, resources, and operational considerations), there also are real opportunities to leverage existing quality assessment activities in each program to build a systematic yet flexible framework for managing quality in BLS survey products and operations.

References

Biemer P and Caspar R (1994) Continuous Quality Improvement for Survey Operations: Some General Principles and Applications, *Journal of Official Statistics*, vol 10, no 3, pp 307-326.

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

Dippo, C. (1997). "Survey Measurement and Process Improvement: Concepts and Integration," in Lyberg L, Biemer P, Collins M, De Leeuw E, Dippo C, Schwarz N, and Trewin D (eds.*), Survey Measurement and Process Quality*, New York: Wiley, pp. 457-474.

Ehling, M. and Korner, T. (2007). Eurostat Handbook on Data Quality Assessment Methods and Tools. Available at http://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK%20ON%20DATA%20QUALITY%20ASSESSMENT%20METHODS%20AND%20TOOLS%20%20I.pdf.

Eurostat (2009). European Statistical System (ESS) Standard for Quality Reports. Available at http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf.

Eurostat (2011). European Statistics Code of Practice. Eurostat. Available at http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-32-11-955/EN/KS-32-11-955-EN.PDF.

Gonzalez, J., Hackett, C., To, N., and Tan, L. (2009). "Definition of Data Quality for the Consumer Expenditure Survey: A Proposal" (internal BLS memorandum). Available at http://www.bls.gov/cex/ovrvwdataqualityrpt.pdf.

International Monetary Fund (2012). IMF Data Quality Assessment Framework. Available at http://dsbb.imf.org/pages/dqrs/dqaf.aspx.

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. Available at http://www.imf.org/external/pubs/ft/bop/2003/dataq.pdf.

Mazzi, G., Museux, J., and Savio, G. (2005). "Quality measures for economic indicators" Eurostat Working Papers and Studies. Available at http://unstats.un.org/unsd/EconStatKB/Attachment288.aspx.

Morganstein D and Marker D A (1997). "Continuous Quality Improvement in Statistical Agencies," in Lyberg L, Biemer P, Collins M, De Leeuw E, Dippo C, Schwarz N, and Trewin D (eds.*), Survey Measurement and Process Quality*, New York: Wiley, pp. 475-500.

Office for National Statistics (2013). Guidelines for Measuring Statistical Output Quality, Version 4.1. Available at http://www.ons.gov.uk/ons/guide-method/method-quality/quality/guidelines-for-measuring-statistical-quality/index.html.

Statistics Canada (2002). Statistics Canada's Quality Assurance Framework. Available at http://www.statcan.gc.ca/pub/12-586-x/12-586-x2002001-eng.pdf.

Statistics Canada (2009). Statistics Canada Quality Guidelines (5th ed.), Ottawa, Canada. Available at http://www.statcan.gc.ca/pub/12-539-x/12-539-x2009001-eng.htm.