Definition of Data Quality for the Consumer Expenditure Survey: A Proposal

Jeffrey Gonzalez, Catherine Hackett, Nhien To, and Lucilla Tan

FINAL REPORT

Submitted: 2009.10.22

1. INTRODUCTION

The Gemini Project was developed in response to the findings of the Consumer Expenditure Survey Program (CE) Review conducted between March 2007 and September 2008. The primary objective of the Gemini Project is to develop a detailed research roadmap for a redesign of the CE to improve data quality through a verifiable reduction in measurement error, with a particular focus on under-reporting. The effort to reduce measurement error is to be explored in a manner consistent with combating further declines in response rates. [1]

The Gemini Project Steering Team chartered the Data Quality Definition (DQD) Team to produce an operational definition of data quality for the CE.[2] The mission of the team is stated as follows:

> "The mission of the Data Quality Definition (DQD) Team is to produce an operational definition of data quality for the Consumer Expenditure Surveys Program (CE). The definition should include a statement of the procedures or ways in which the program should measure data quality, provide a framework for assessing the overall quality of CE survey data, and address the fitness-for-use concerns of individual program stakeholders."

In developing this definition, the DQD Team should:

1. Give consideration to the final users of the definition, and how the definition should be used.

2. Use the data quality definition to identify sources of measurement error that may have a negative impact on aspects of CE data, and where possible identify the mechanisms by which those sources operate. The team will create a matrix that associates these sources of error with the needs of individual program stakeholders, and use this set of associations to create a hierarchy of data quality concerns.

In this report, we present our proposal for a definition of data quality for the CE. We also illustrate how the proposed data quality definition can assist CE to prioritize survey improvement activities in ways that are related to the survey's primary data uses. Finally, we also propose indicators and measures that can be used to routinely monitor data quality, as well as to assess the impact of survey changes on data quality.

[1] Consumer Expenditure Survey Program Gemini Project Charter (version 11), 17 February 2009.
[2] Consumer Expenditure Survey Data Quality Definition Team Charter (version 5), 7 April 2009.

2. DEFINING DATA QUALITY FOR THE CE

2.1 CRITERIA

In the development of our proposed data quality definition, we gave primacy to the following considerations:

1. The CE's mission.

   The CE's mission is stated in the 2007 Strategic Plan as follows:[3]

   "The mission of the Consumer Expenditure Survey (CE) is to produce and disseminate statistical data on consumer expenditures, demographic information, and related data needed by the Consumer Price Index and other public and private data users; to design and manage the CE survey; to provide education and assistance in the use of the data; and to conduct analytical studies. CE data must satisfy a number of criteria, including relevance, timeliness, respondent confidentiality, reliability, impartiality, and consistently high statistical quality."

2. Attention to data needs of primary data users.

   While there are many users of CE data, there are some whose needs should take precedence over others (e.g., because of funding); we refer to these as *primary data users*. Regardless of which data users CE Management identifies as primary, the proposed definition should be applicable to any data user.

3. Making the definition useful towards the current redesign effort for the CE.

   The definition should facilitate the documentation and communication of data quality concerns of a specific primary data use to all levels of CE staff involved with ongoing efforts to improve the CE, so that all staff have a common understanding of the issues. The definition should also facilitate the identification of common problems across primary data users, which could then suggest how redesign activities may be prioritized.

Given these three considerations, we believe it is appropriate that CE adopts a multidimensional definition of data quality that reflects concern for the suitability of its data products to meet the needs of primary data uses. This is often referred to as "fitness for use". While the Team's charter specifically mentions measurement error, the proposed data quality definition also gives consideration to other error sources and dimensions of data quality which are relevant, if "fitness for use" is the basis for defining data quality. The team received approval from the Gemini Steering Team to proceed with this broader approach to defining data quality.[4]

[3] Consumer Expenditure Survey Strategic Plan, 2007. Internal BLS document.
[4] Meeting with Gemini Steering Team on June 1, 2009.

2.2 BACKGROUND

Over the past two decades, the general perspective of data quality has evolved from a statistical focus on accuracy (Total Survey Error framework [TSE] in Biemer and Lyberg 2003; Groves 1989; Groves et al. 2004) to a broadening of the definition to include other dimensions that reflect concern for the suitability of the final survey products for users (Total Quality Management framework [TQM]; Brackstone 1999; Dobbs et al. 1998; Lyberg et al. 1998). In addition to accuracy, the broader definition of data quality often includes dimensions pertaining to relevance, timeliness, accessibility, interpretability and coherence.

A consequence of the emphasis placed on the "fitness" of the final survey products to data users is the recognition that there are often a diverse range of users, so that what is fit for one purpose may be inappropriate for another. Since it is impractical to meet all user needs simultaneously, data producers have made efforts not only to deliver the best quality data possible within practical constraints (e.g., time and resources), but also have emphasized the importance of providing users with sufficient description of aspects of the processes and quality so that users can make informed decisions about the suitability of the survey products for their use (e.g., Office of National Statistics 2007; Lee and Allen 2001). There is also recognition that most error sources are difficult to quantify, and qualitative measures should not be disregarded:

> "Reporting of quality measures that depend only on direct measures of quantifiable error provide only a partial quality picture to users. A much wider view of quality needs to be taken…include descriptive information, and not just indicators of output but also indicators relating to intermediate outputs and processes leading to outputs" (Dobbs et al. 1998)

Thus in the paradigm where "fitness for use" is emphasized, reporting on quality to assist users is also a major responsibility of data producers and an integral dimension of the framework.

Different agencies and analysts provide different definitions of data quality, but *all* agree that data quality is a multidimensional concept (e.g., Australian Bureau of Statistics 2009; European Statistics Code of Practice 2005; Statistics Canada 2003).

2.3 DEFINITION

We propose CE's data quality definition be based on both the TQM (Brackstone 1999) and TSE (Groves et al. 2004) paradigms, as illustrated in Figure 1. The six TQM dimensions are defined in Table 1. Examples of how particular data users might relate these dimensions to their specific uses are also provided. For the accuracy dimension, the TSE paradigm is used to identify the sources of error that could occur at each stage of the survey process. These error sources are defined in Table 2.

Figure 1. Dimensions of the Proposed Data Quality Definition for the Consumer Expenditure Survey

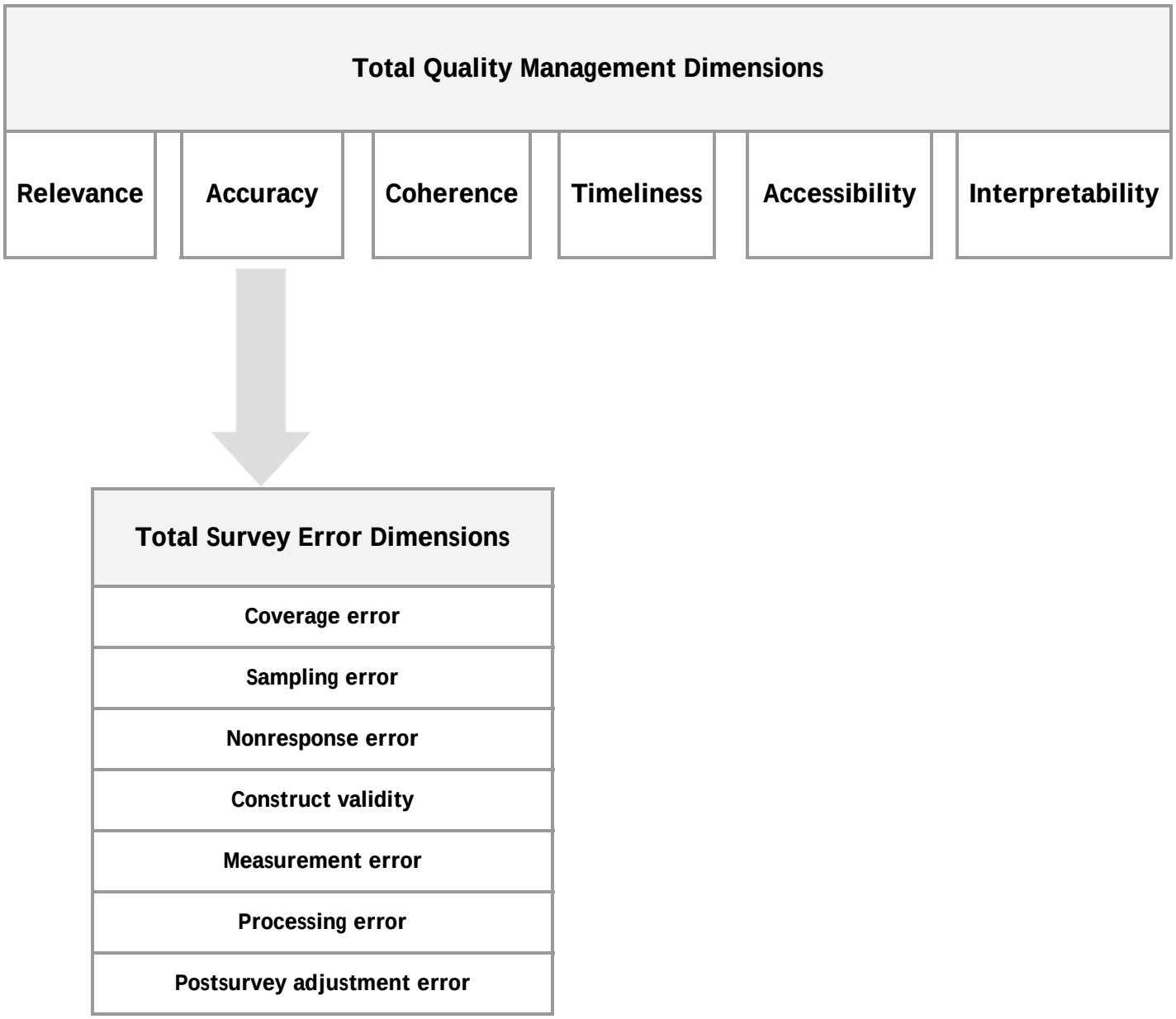


Table 1. Total Quality Management Dimensions of Quality

Dimension and Definition	Examples for the CE
1. Relevance The degree to which the survey products meets the user's specific needs in terms of both content and coverage.	Relative shares of expenditure categories are used in the formation of the Consumer Price Index (CPI) weights. Failure of the CE to capture certain types of expenditures may result in the misrepresentation of the relative shares of expenditure categories.
2. Accuracy The degree to which the estimate is similar to the true value of the population parameter. The difference between the estimate and the true value is the Total Survey Error.	Error in the measurement of actual purchases could result in the misrepresentation of relative shares.
3. Coherence Degree to which different sources or methods on the same phenomenon are similar.	If there are other data sources that describe US households' out-of-pocket spending patterns, to what degree are the key estimates from these other sources similar to those from the CE?
4. Timeliness The interval between the time data are made available to users and the event or the phenomena the data describe.	The superlative CPI, which allows for changes in consumer spending behavior in response to changes in the relative prices of CPI item categories, is based on weights constructed from the CE that lag two years.
5. Accessibility The ease with which statistical information and appropriate documentation describing that information can be obtained from the statistical organization.	Tabulated data are published on CE's public website. Microdata are available for purchase. Data users can contact CE by phone, fax, mail, and email. CE responds promptly to data inquiries.
6. Interpretability Availability of adequate information to allow users to properly use and interpret the survey products.	CE provides information on survey metadata, data quality information and indicators on CE's public website.

Table 2. Accuracy Dimension: Sources of Error from the Total Survey Error Paradigm

Source of Error	Definition
Coverage error	The error resulting from a mismatch between the sampling frame (the "list" or set of materials used to select the sample), and the target population (the finite population that the survey is attempting to characterize or draw inferences about). A mismatch can occur because: (1) part of the target population is not linked to the sampling frame; (2) there are ineligible units on the sampling frame (e.g., vacant housing units); (3) multiple eligible units can be linked to one listing on the sampling frame (e.g., there may be distinct CUs at the same address); and, (4) the same CU may be linked to multiple sample units.
Sampling error	The error that results from drawing one sample instead of examining the entire target population. It also refers to the difference between the estimate and the parameter as a result of only taking one sample instead of the entire population (i.e. conducting a complete census).
Nonresponse error	There are two types of nonresponse error: (1) unit nonresponse, where a sampling unit does not respond to any part of the survey, and (2) item nonresponse, where a sample unit only partially completes the survey. Unit nonresponse may be a consequence of not being able to contact the unit, refusal by the unit to provide some or all of the requested information, or some other reason (e.g., a cognitive or language barrier). There are two viewpoints on nonresponse. First, a deterministic viewpoint that there are two types of sample units – those who never respond and those who always respond. Nonresponse bias in the sample mean is then a function of the proportion of nonrespondents and the difference between the average value of the survey estimate for the respondents and the nonrespondents. Second, a stochastic viewpoint where people have a propensity or likelihood to response. Nonresponse bias of the sample mean is then a function of the correlation between the response propensity and the substantive variable of interest being measured.
Construct validity	The extent to which the survey questions reflect the underlying construct (e.g., theme or topic) they are designed to measure. Another related and similar term to construct validity is specification error. This occurs when the concept implied by the survey question and the concept that should be measured in the survey differ.
Measurement error	The difference in the response value from the true value of the measurement. The major sources of this error are the respondent (e.g., the respondent either deliberately or unintentionally provides incorrect information), the interviewer (e.g., the interviewer incorrectly records of response), mode of data collection, and the data collection instrument (e.g., poor question wording).
Processing error	The deviation between the value used in estimation and the response value provided. It refers to errors that arise in the data processing stage, including errors in the editing of data, data entry (not done by the interviewer), and coding.
Postsurvey adjustment error	The extent to which survey estimates are affected by errors in adjustment procedures, such as weighting and imputation, that are initially designed to reduce coverage, sampling and nonresponse errors.

All the six TQM dimensions should be borne in mind in any modifications to the existing survey, or large-scale survey redesign. An attempt to improve one dimension could have unintended adverse impacts on other dimensions. Similarly, an attempt to address one source of survey error may inadvertently increase another type of survey error. Practical considerations may force trade-offs to be made between data quality dimensions and error sources, but these trade-offs should be the result of conscious and deliberate decisions, and not unintended consequences. We hope the proposed data quality definition will help to keep this “big picture” in mind whenever survey modifications are considered.

3. PRIORITIZING DATA QUALITY CONCERNS BY DATA USERS

The ability to prioritize data quality concerns is an asset in a survey environment where resources are scarce. The proposed data quality definition can help in the endeavor to prioritize data quality concerns that are related to primary uses of the data, by providing a framework for systematically gathering information about the fitness of the data in meeting the objectives of any specific data user. In the rest of this section, we outline the procedure for using the proposed data quality definition to identify and prioritize data quality concerns of specific data users, as well as provide an illustration of using the procedure.

3.1 METHOD

The procedure for using the proposed data quality definition is outlined in Table 3.

Table 3. Using the Proposed CE Data Quality Definition to Document Data Quality Concerns of Specific Data Users

1. CE Management identifies the primary data users of the CE for whom "fitness for use" is a priority.

2. For each identified user of the CE data from Step 1, use the proposed data quality definition to guide the collection of information regarding the "fitness" of the CE data for that user:

    a. Identify the relative importance of the six TQM dimensions from the user's perspective, document concerns about each dimension, and potential issues with proposed solutions.

    b. In addition, for the accuracy dimension

        i. identify the error sources of concern;
        ii. identify the underlying possible mechanism of the error source;
        iii. propose solutions to address the mechanism, where possible; and,
        iv. identify the potential impact of the proposed solution on the respondent, other primary data user(s), interviewers, processing, and program resources.

    c. Document user needs that are not currently being met.

3. Compare the information gathered across the data users, and where possible, identify a hierarchy of data quality concerns.

Implementing the proposed data quality definition in this prescribed manner facilitates the identification of ways in which the data are not meeting user needs, sources of data error and their possible underlying mechanisms that are of concern to specific data users, *and* to record this information in a consistent and systematic way.

Having information recorded in this way provides a comprehensive profile of the needs and concerns of each primary data user. In addition, it facilitates the comparison of concerns and error mechanisms across these primary data uses. The relative importance of data quality concerns and error mechanisms become transparent if they are common to more than one primary user. This information can be used to guide the establishment of priorities in the allocation of resources and research activities in support of ongoing efforts to improve the CE to meet the needs of primary data users.

In the following subsection, we provide an example of how we used the proposed data quality definition to identify a possible prioritization of data quality concerns related to specific data users.

3.2 ILLUSTRATION

Although we recognized there are many different uses of the CE data, we selected three data users for this illustration. Note however, the *process* of using the proposed data quality definition to identify a possible prioritization of data quality concerns is the same, regardless of which data user or how many data users are selected.

The three data uses we adopted for this illustration are: official CE estimates that appear in the tables published on the CE public website and biennial publication (henceforth referred to as *Official CE Estimates)*, the Consumer Price Index (CPI), and the Experimental Poverty Thresholds. The CPI uses the CE to update the cost weights in the CPI market basket of goods and services biennially. The Experimental Poverty Thresholds is an ongoing joint Census-BLS research project to develop new poverty thresholds based on the work of the National Academy of Sciences (NAS) to define basic needs and family resources. In 1995, the NAS Panel on Poverty and Family Assistance published a report, *Measuring Poverty, A New Approach* (Citro and Michael, 1995), which showed why the official current poverty measure was outdated and should be revised. The Experimental Poverty Thresholds uses the CE Interview Survey for data on out-of-pocket spending by families.

Our contacts for these data uses were William Passero for Official CE Estimates, Robert Cage for CPI, and Thesia Garner for Experimental Poverty Thresholds.

3.2.1 Documenting Information on Data Quality Concerns of Specific Data Users

We were guided in our information gathering discussions with the data use contacts by the data quality definition, using Step 2 of Table 3. We documented our discussions with the data use contacts in the following tables. The relative importance of the six TQM quality dimensions is shown in Table 4. The TSE error sources that are of concern for each specific data use are documented separately in Tables 5a – 5c. In addition to recording the concerns the data use contacts mentioned, we also attempted to identify the error mechanism (i.e., the fundamental processes involved in or causing the error), possible solutions for addressing the error, how the proposed solution might be evaluated, and the possible impact of the proposed solution on the respondent / data user / interviewer / processing systems.

We emphasize that the entries in Tables 4 and 5a - 5c are meant to illustrate how the proposed data quality framework could be implemented to prioritize data quality concerns related to specific data uses. These entries should not be interpreted as comprehensive or final (as this would require more in-depth discussions with the data use contacts than time permitted for this proposal).

3.2.2 Prioritization through Comparisons

With data quality concerns of individual data users documented in Table 4 and Tables 5a - 5c, the next step was to compare the information in these tables across the data uses for the purposes of identifying common and distinct issues of data quality concerns.

Overall ranking of the TQM dimensions. The relative importance of the TQM dimensions for the three data uses in Table 4 is summarized in Table 6. Accuracy was unanimously ranked the most important for all three data uses. The ranking of other TQM dimensions is less similar across the three data uses.

The ranking of the TQM dimensions across all three data uses was computed as *the average of the rankings across the three data uses, weighting each data use as equally important.* For example, the overall rank for accuracy was (1+1+1)/3) = 1. The overall ranking (1=most important, 5=least important; ties were permitted) among the three data users were (see Table 7):

1. Accuracy
2. Timeliness
3. Relevance
4. Interpretability
5. Access & Coherence

We recognize that the overall ranking of the six dimensions of data quality across multiple data uses can misrepresent the relative importance of these dimensions to a specific data use. For example, the dimension of interpretability is ranked number 2 for the Published Tables data use (see Table 4) but this dimension appears as number 4 in the overall ranking shown above. While how to weight each primary data use may be debated, this issue does not undermine the usefulness of the framework in documenting the data quality concerns each specific data user's and coming up with an overall ranking across the data users.

Common concerns regarding sources of error. We compared the sources of error information for each data use in Tables 5a – 5c, and summarized our comparisons in Table 7. This summary highlighted the following concerns common among data uses (not presented in any ranked order):

- *Measurement error*.
    - Difficulties in recall, as well as proxy reporting for other CU members (in the current survey design), are concerns for all three data uses.
    - Under-reporting in different expenditure categories in the two sources of data (the Diary and Interview surveys) is problematic when a data use needs complete expenditure information on a CU. This affects both the CPI and the Experimental Poverty Thresholds.
- *Coverage error*. Both the CPI and the Experimental Poverty Threshold are interested in special subgroups (e.g., the elderly, low income).
- *Construct validity*. The current questions asked in the Interview Survey do not capture sufficient detail for both the CPI and the Experimental Poverty Thresholds. For example, the Experimental Poverty Thresholds require more detailed information about CU participation in various types of government assistance programs; the CPI would like information on where purchases are made because place of purchase is important in the construction of the CPI for cohorts of interest (such as the elderly).
- *Postsurvey adjustment*. Lack of longitudinal weights. The use of expenditure distributions in the computation of the Experimental Poverty Thresholds requires (at least) one year's expenditure data on the reference CU. As changes within CU composition during the CU's tenure in the survey panel are not rare events, longitudinal weights are needed so that the CU's weight properly reflects the numbers of CUs they represent in the population.

Table 4. Illustration: Data Users' Ranking of Total Quality Dimensions

Data Quality Dimensions	Official CE Estimates	CPI	Experimental Poverty Thresholds
Relevance	**5**	**4**	**1**
	Diverse users of table data; not practical to meet all needs.	▪ Construction of cohort indexes (e.g., elderly, social security recipients, low vs high income) require: - coverage of these cohort populations - their point of purchase information ▪ Prefer total expenditure coverage on each household: - Permits more flexibility in index construction. A household-specific index requires full coverage of expenditures for a household. Advantages: can consider distribution, instead of just the mean; construction of a "democratic index" where all households are weighted equally. Currently, each household is weighted by total spending, so high spending households get more weight ("plutocratic index"). ▪ Concern that the Diary weights are too low.	▪ Data needed to compute calendar year thresholds based on expenditure distributions for Food, Clothing, Shelter, and Utilities (FCSU) - Require sufficiently large sample - at least one year's expenditures on each reference CU. - The thresholds are based on percentiles along a distribution, and thus they are more sensitive to seasonality and periodicity of the data than means.

Table 4. Illustration: Data Users' Ranking of Total Quality Dimensions

Data Quality Dimensions	Official CE Estimates	CPI	Experimental Poverty Thresholds
Accuracy	**1** Users of published tables range widely in statistical skills, many of whom "take the number as given".	**1** ▪ User believes sampling errors are more of concern that nonsampling errors in the CEQ. ▪ Sample size of infrequently purchased items too small at the Item-Area level. ▪ Wants unbiased, low variance estimate on each Item-Area-Month-Target populaion	**1** ▪ Minimize imputation of expenditures. If a split questionnaire design is being considered, consider making FCSU core questions.
Timeliness	**3**	**2** ▪ 2 year lag in CE data in superlative CPI - would like monthly/quarterly delivery of CE data.	**1** ▪ Need earlier delivery of thresholds to Census than currently possible (by July, so that the thresholds can be posted on Census website by early September).
Coherence	**6**	**2**	**5**
Access	**4**	**4** ▪ CPI does Universal Classification Code to Elementary Level Item mapping.	**5** ▪ The McDermott House Bill calls for the release of microdata and online tools necessary for the computation of the published modern poverty thresholds and poverty rates.

Table 4. Illustration: Data Users' Ranking of Total Quality Dimensions

Data Quality Dimensions	Official CE Estimates	CPI	Experimental Poverty Thresholds
Interpretability	**2**	**6**	**4**
	▪ Given the consumers of data in this form come from a broad range of statistical background, documentation and other aids for correctly interpreting the data are very important.		

* Ranking for Official CE Estimates by Bill Passero; CPI by Rob Cage, Mary Lynn Schmidt, and Joshua Klick; Experimental poverty thresholds by Thesia Garner.

Table 5a. Ilustration: Documentation of User Concerns over Sources of Error for Accuracy Dimension - Official CE Estimates

Sources of Error (TSE)	Possible Error Mechanism	Proposed Solution(s)	Measure/Indicator/Method	Possible impact on respondent/user/interviewer/processing
Sampling error			Standard error – already available for published tables	
Construct validity	*No concern specified*			
Measurement error		Necessary to understand the magnitude of this problem and the expenditure categories most susceptible to this problem.	Validation study to assess extent and areas of mismeasurement.	
	Respondent has difficulty recalling expenditure due to nature of the item.	Change / vary reference periods. Offer different methods to aid recall.	Cognitive test and/or field test of varying reference periods. Assess use and effectiveness of existing and new recall aids offered.	
	Respondent fatigue due to length of survey.	Redesign questionnaire: e.g., use Split Questionnaire design where respondents are not asked the entire questionnaire at every interview.	A field test for a Split Questionnaire design for the CEQ is planned for 2010.	Potentially more data processing with higher level of imputation with Split Questionnaire (SQ) design. Users face more statistical challenges with imputed data.
	Respondent unaware of expenditures made by another CU member.	Design that allows for multiple respondents (e.g., member diary).		

Table 5a. Illustration: Documentation of User Concerns over Sources of Error for Accuracy Dimension - Official CE Estimates

Sources of Error (TSE)	Possible Error Mechanism	Proposed Solution(s)	Measure/Indicator/Method	Possible impact on respondent/user/interviewer/processing
Processing error	Input file error (e.g., parse files, UCC mapping) – not updated correctly	Better integration of metadata systems; process checks.		Slows down access to data Reprocessing diverts staff from other activities
	Instrument error (wrong skip pattern, mis-synchronization of instrument versions, unanticipated effects on downstream variables from a change)	Better instrument documentation to aid in identifying impacts of a change in one variable on other variables.		
	Incorrect pre-edit / pre-process files from Census			
Coverage error	*No concern specified*			
Nonresponse error	*No concern specified*			
Postsurvey adjustment error	*No concern specified*			

Table 5b. Ilustration: Documentation of User Concerns over Sources of Error for Accuracy Dimension - the CPI

Sources of Error (TSE)	Possible Error Mechanism	Proposed Solution(s)	Measure/Indicator/Method	Possible impact on respondent/user/interviewer/processing
Sampling error	Sample size is too small at the area level for the Wage & Earner, and Elderly population cohorts.	Increase sample size or oversample	Standard error of estimate.	More funding.
Construct validity	Expenditure details for some CPI Item Strata are not currently asked in the CE. [e.g., packets vs. bag of sugar]	Add questions on details for CPI's Item Strata. Need to define new constructs.	Map question(s) to information required to identify what is and what is not adequately asked? This is also useful metadata for identifying question ownership & orphan questions.	Increase respondent burden.
	Point of purchase for item not asked.	Add question on point of purchase (CPI is willing to provide funding).	Item response rate.	Possible increase in respondent burden. Time added to survey. CPI drops its Telephone Point of Purchase Survey.
	Collection of all necessary detailed expenditures from one CU.	Use a one sample instead of the current two samples, allowing for multiple instruments or methods for data collection from the same CU.	Field test. (Performance measures: response rates, standard errors, data quality indicators; collection costs).	Increases respondent burden, but has the following advantages: + all collected expenditures from a CU will be used to produce survey estimates, and none will be "discarded". + provides option to create household-specific index.
Measurement error		Necessary to understand the magnitude of this problem and the expenditure categories most susceptible to this problem.	Validation study to assess extent and areas of mismeasurement.	
	Respondent has difficulty recalling expenditure due to nature of the item.	Change reference period. Offer different methods to aid recall.	Cognitive test and/or field test Assess use and effectiveness of recall aids implemented.	

Table 5b. Ilustration: Documentation of User Concerns over Sources of Error for Accuracy Dimension - the CPI

Sources of Error (TSE)	Possible Error Mechanism	Proposed Solution(s)	Measure/Indicator/Method	Possible impact on respondent/user/interviewer/processing
	Respondent fatigue due to length of survey.	Redesign questionnaire: e.g., use Split Questionnaire design where respondents are not asked the entire questionnaire at every interview.	A field test for a Split Questionnaire design for the CEQ is planned for 2010.	Potentially more data processing with higher level of imputation with SQ design. More imputed data with SQ design. Statistical challenges with using imputed data.
	Respondent unaware of expenditures made by another CU member.	Design that allows for multiple respondents (e.g., member diary).		
Processing error	Incorrect UCC mapping Inaccurate Diary item code mapping	Better integration of metadata systems; process checks.		
Coverage error	Inadequate sample of subgroups to construct special cohort indices (e.g., elderly, low-high income). Young, single under-represented	Over sample members of under-represented subgroups.		More funding is required.
Nonresponse error	*No concern specified*			
Postsurvey adjustment error	*No concern specified*			

Table 5c. Illustration: Documentation of User Concerns over Sources of Error for Accuracy Dimension - the Experimental Poverty Thresholds

Sources of Error (TSE)	Possible Error Mechanism	Proposed Solution(s)	Measure/Indicator/Method	Possible impact on respondent/user/interviewer/processing
Sampling error	*No concern specified*			
Construct validity	Computation of thresholds also requires consumption-based constructs. CE does not collect some of this type of information. *Examples:* The length of tenure in current dwelling; this information is critical for subsidized housing and rent control. Need more detailed information about participation in government assistance programs.	Identify both spending-based and consumption-based constructs needed to compute the thresholds. Improve documentation (e.g., identify variables that are in-scope for consumption in the data dictionary) Re-visit CE conceptual framework report.	Map questions - existing, or to be created - to these constructs.	Additional variables to capture data needs for thresholds may increase respondent burden.
Measurement error	Comprehension problems – vague concepts or unfamiliar terms.	The standard fixes for comprehension problems may work– pretesting, cognitive interviewing, "conversational-like" interview.		Any changes in the way concepts are currently asked may disrupt (and to an unknown extent) other components of the analytic objective.
	Recall problems for some items.	Better prompting to aid in recall. Identify appropriate reference periods for different expenditure items.	Cognitive test and/or field test Assess use and effectiveness of recall aids implemented.	

Table 5c. Ilustration: Documentation of User Concerns over Sources of Error for Accuracy Dimension - the Experimental Poverty Thresholds

Sources of Error (TSE)	Possible Error Mechanism	Proposed Solution(s)	Measure/Indicator/Method	Possible impact on respondent/user/interviewer/processing
	Perhaps not all members of the CU are being accounted for across the five waves of CE Interview Survey. Thesia's research indicated about 10% of the reference CUs changed characteristics across the waves.			
	Currently all expenditure data for the threshold computations are based on the CE Interview Survey, but the CE Diary Survey collects better food expenditure data.	Have only one data source for expenditure estimates. (Use one sample to collect data).		Potentially increase (perceived) respondent burden.
Processing error	*No concern specified*			
Coverage error	Shifts in population not captured. Examples: ▪ Families in semi-permanent shelters ▪ Elderly (different levels of assisted living) ▪ Minorities ▪ Low income	Supplement current sampling frame with additional frames/lists that target (or better) cover these groups. Possible to pretest elderly living arrangement questions and better train interviewers to handle these types of living arrangements.		More complicated construction of weighting adjustments because sampling units maybe contained on multiple lists. If the standard sampling frame is supplemented with additional lists/frames, then users could erroneously calculate weights - documentation is key.

Table 5c. Ilustration: Documentation of User Concerns over Sources of Error for Accuracy Dimension - the Experimental Poverty Thresholds

Sources of Error (TSE)	Possible Error Mechanism	Proposed Solution(s)	Measure/Indicator/Method	Possible impact on respondent/user/interviewer/processing
	"Extreme" subgroups in the population are very important for thresholds.	Oversample members of under-represented groups.		
Nonresponse error	*No concern specified*			
Postsurvey adjustment error	Threshold computations will make use of CE's panel feature. Quarterly weights that treat CUs as independent across quarters do not address changes in composition of CUs and unit nonresponse across panel.	Provide longitudinal weights.		More resources are needed to develop longitudinal weights.

Table 6. Ranking of Total Quality Management Dimensions by Data Users*

Data User	Ranking of Total Quality Management Dimensions					
	1 (most important)	2	3	4	5	6 (least important)
CPI	**ACCURACY**	**TIMELINESS** **COHERENCE**		**ACCESS** **RELEVANCE**		**INTERPRETABILITY**
Experimental Poverty Thresholds	**ACCURACY** **RELEVANCE** **TIMELINESS**			**INTERPRETABILITY**	**COHERENCE** **ACCESS**	
Published Tables	**ACCURACY**	**INTERPRETABILITY**	**TIMELINESS**	**ACCESS**	**RELEVANCE**	**COHERENCE**

Information presented here is based on Table 4. Ranking of TQM dimensions was done by the primary contact for data use; ties in rankings were permitted.

Table 7. Accuracy Dimension: Summary of Error Concerns Among Data Users

Coverage Error	Sampling Error	Nonresponse Error	Construct Validity	Measurement Error	Processing Error	Postsurvey Adjustment
CPI & Experimental Poverty Thresholds: - Under-representation of subgroups of interest (e.g., the elderly, low-income)	**CPI:** - Sample size too small for Wage and Earner, and Elderly cohorts; also for infrequently purchased items at the Item-Area level		**CPI:** - Inadequate expenditure details on some Item-Strata **Experimental Poverty Thresholds**: - Inadequate consumption-based constructs (e.g., length of residence in current dwelling unit).	**All 3 uses:** - Recall difficulty - Respondent unaware of purchases made by another CU member **Experimental Poverty Thresholds**: - Vague questions (e.g., GOVTCOST) - 2 independent CE samples for CEQ and CED: one survey collects better data on some items, but thresholds need to be based on one reference CU, so forced to use data from CEQ. **CPI:** - A single data source would be preferable - more flexibility in index construction - Concern that CED weights are too low (source for food data)	**All 3 uses:** - Incorrect pre-edit/pre-process files from Census - Instrument error - Incorrect UCC mapping	**Experimental Poverty Thresholds:** - Unavailability of longitudinal weights to adjust for changes in reference CU composition over tenure in the panel

4. ASSESSING AND MONITORING DATA QUALITY

In this section, we propose a set of data quality measures and indicators for consideration. A data quality *measure* directly assesses a specific aspect of data quality, and a data quality *indicator* does this assessment indirectly to provide insight to that aspect of quality (Office of National Statistics 2007). While it is often costly or difficult to directly measure data quality, it is possible to obtain indicators. For example, a direct measure of nonresponse bias of expenditure on alcohol is the exact amount of nonresponse bias. However, this exact quantity cannot be computed since the average expenditure on alcohol by nonrespondents is unknown. Instead, we use indicators, such as the average expenditure on alcohol by "proxy nonrespondents", to estimate the nonresponse bias of alcohol expenditures.[5]

These measures and indicators are recommended for use to routinely monitor data quality, and to evaluate effects of survey changes on data quality. It is important to agree on *standard definitions* for these proposed measures and indicators, so that whenever some or all of these measures and indicators are invoked in a study or in a monitoring process, these measures and indicators are uniformly implemented. This standardization allows the measures and indicators to be comparable across studies and time.

4.1 PROPOSED MEASURES AND INDICATORS

Most indicators and measures we propose for assessing data quality of the CE are quantitative, but some will take the form of qualitative assessments. Among the six TQM dimensions, quantitative indicators are most prevalent for the Accuracy dimension. It is possible to assess the dimensions of Timeliness, Accessibility, and Interpretability independently of each other. However, the assessment of each of the dimensions of Relevance, Accuracy, and Coherence assumes the other two dimensions are perfectly satisfied. For example, the assessment of Accuracy presupposes the standard of comparison is both Relevant and Coherent.

We offer suggestions for assessing the six TQM dimensions (Table 8), with a focus on indicators and measures of Accuracy (especially measurement error, Table 9).

[5] One way of estimating nonresponse bias for a survey statistic in the deterministic framework is to treat a group of respondents who meet some criteria as nonrespondents (hence the term "proxy nonrespondents"), and compare their response values to those of other respondents. An example of a criteria used to identify proxy nonrespondents are respondents who are difficult to contact, as measured by the proportion of contact attempts that are non-contacts.

Table 8. Proposed Indicators to Assess and Monitor Total Quality Management Dimensions

TQM Dimension	Assessment	
Relevance	Indicator	▪ User satisfaction index, and topic-specific sub-indices on different dimensions of data.
	Method	▪ User satisfaction survey of key data users to identify gaps between user needs and current data. Data are relevant when user needs are met in terms of coverage and detail. o Survey followup: describe how feedback from users will be translated into concrete actions.
Accuracy	Indicators	▪ Proposed indicators are listed in Table 9.
		▪ A recurring interviewer survey to obtain feedback about field experience on the survey (e.g., after every 2-year form change) to identify problem areas in the survey, as well as to learn what is working well . o Follow up: report interviewer survey results overall and by Regional Office back to interviewers as well – highlight common problems identified, and address how interviewer concerns are being handled (or not).
Timeliness	Indicator	▪ Average time lag between end of the reference period and data release date.
	Method	▪ Describe key user's need for recency of collected data, and how these needs are addressed.
Coherence	Method	▪ Routine comparison with other data sources and documenting why and how these alternative sources differ from CE.
Accessibility	Indicator	Track over time: ▪ Number of hits on CE public website ▪ Number of requests for microdata. ▪ Document user questions about the data. ▪ Average time to respond to different types of data inquiries.
	Method	▪ User survey to assess ease in locating CE information (web, microdata).
Interpretability	Method	▪ Assess quality and sufficiency of metadata provided to assist data users.

4.2 INDICATORS AND MEASURES OF ACCURACY

Most of the quantitative indicators for the accuracy dimension are based on information currently collected in the survey, and the Contact History Instrument (CHI). The CHI is a supplemental component to the Computer Assisted Personal Interview (CAPI) instrument that interviewers use to record information on each attempt made to contact the respondent to conduct the interview. The CHI allows the interviewer to record the date, time, mode and outcome of each contact attempt. The interviewer can also record the contact attempt strategy (or strategies) employed, and the interviewer's perception of whether the respondent has concerns about participating in the survey upon contact, and if so, what type of concern(s).

In the Table 9, we provide a brief description of the indicator, the desired direction of the indicator for accuracy, the error source(s) which the indicator indirectly measures, whether it is currently collected in the Interview and Diary surveys, and if it is, we identify the source variable. The only measures of accuracy currently proposed are the data editing rates. As further research identifies other indicators and/or measures, the list of proposed indicators and measures in Table 9 can be expected to change over time.

Table 9. Proposed Indicators and Measures for Assessment and Monitoring of Accuracy

Description of Proposed Indicator / Measure [Desired Direction]	Source of Error	Interview Survey		Diary Survey	
		Currently Collected	Source Variable	Currently Collected	Source Variable
EXPENDITURES					
Expenditures reported [Generally higher is better, but requires context, e.g., relative to CU size]	ME	Y	EXPNSUM	Y	Z_TOTAL
Number of expenditure reports [Generally higher is better, but requires context, e.g., relative to CU size]	ME	Y	NUMEXPN	Y	
EFFORT TO COLLECT DATA					
Number of contact attempts to final disposition [Lower is better]	ME, NR	Y			
Number of attempts prior to and including 1st contact with sample unit member [Lower is better]	ME, NR	Y	CTSTATUS		CTSTATUS
Number of attempts to begin interview [Lower is better]	ME, NR	Y	CTTYPE		CTTYPE
RESPONDENT COOPERATION					
Interviewer's perception of respondent concerns [Lower is better]	ME, NR	Y	RSPDNTnn		RSPDNTnn
Soft refusals [Lower is better]	ME, NR	Y	CTSTATUS, CTTYPE, NONINTR		CTSTATUS, CTTYPE, NONINTR
Mode (at the collection level, collected in the CAPI) [Personal visit is preferred]	ME, NR	Y	TELPV	Y	TELPV

Table 9. Proposed Indicators and Measures for Assessment and Monitoring of Accuracy

Description of Proposed Indicator / Measure [Desired Direction]	Source of Error	Interview Survey		Diary Survey	
		Currently Collected	Source Variable	Currently Collected	Source Variable
Mode (at the contact attempt level, collected in the CHI) [Personal visit is preferred]	ME, NR	Y	PERORTEL	Y	PERORTEL
Predominant mode used to collect data in sections for a case [Person visit is preferred]	ME	Y	HOW_INTV	Not applicable	
Sections collected entirely by phone [To be determined]	ME	Y	TELSCT01-TELSCT10	Not applicable	
Primary reason for collecting data by phone [To be determined]	ME	Y	TEL_RESN	Not applicable	
REPORTING BEHAVIOR					
Recall aids usage: Information book [Use is better]	ME	Y	INFOBOOK	Y	INFOBOOK
Recall aids usage: Receipts [Use is better]	ME	Y	RECORDS	N	
Recall aids usage: Used bill for utilities [Use is better]	ME	Y	HAVEBILL	N	
Converted refuser [Not converted refuser is preferred]	ME	Y	CONVREF	Y	CONVREF
Number of explicit "don't know" responses to expenditure items [Fewer is preferred]	ME	Y	NUMDK	N	
Number of explicit "refused" responses to expenditure items [Fewer is preferred]	ME	Y	NUMRF	N	
Length of interview [Longer is preferred]	ME	Y	TOT_TIME	Y	TOT_TIME

Table 9. Proposed Indicators and Measures for Assessment and Monitoring of Accuracy

Description of Proposed Indicator / Measure [Desired Direction]	Source of Error	Interview Survey		Diary Survey	
		Currently Collected	Source Variable	Currently Collected	Source Variable
Identify sections in the Interview Survey that are problematic for respondents (method of implementation to be determined)	ME	N		Not applicable	
DATA EDIT RATES (computed at CU level)					
Expenditure reports by allocation [Lower is preferred]	ME, CV *	Y		Y	
Expenditure reports by imputation [Fewer is preferred]	ME, NR	Y		Y	
No edited reports [Higher is preferred]	ME			Y	
Income imputed [Fewer is preferred]	ME, NR	Y	FINCBTXI	Y	FINCFTM
RESPONSE RATES					
Unit response rate (weighted and unweighted) [Lower is preferred]	NR	Y	OUTCOME	Y	OUTCOME
Item response rate [Lower is preferred]	NR	Y	<expenditure variable>	Not applicable	

Notes:

* Expenditure reports by allocation: since the Interview survey permits reports of combined expenditures, allocation of combined expenditures an indicator of measurement error or another mechanism, such as construct validity.

Column labels

- [Desired Direction] describes the direction we would like to see the indicator for accuracy.
- Source of Error: CV = Construct validity; ME = Measurement error; NR = Nonresponse error
- Currently collected: if the indicator is already collected in the survey instrument (Y), or not (N)
- Source variable(s) in the CEQ (CED): name of variable in the CE Interview Survey (Diary Survey) data file.

The classification of indicators in this table is not meant to be rigid. It should be viewed as an attempt to relate each indicator to behavior or characteristics that could reasonably be expected to cause various types of survey errors.

4.3 CONSIDERATIONS FOR USE

In addition to identifying the appropriate indicators and measures to assess and monitor data quality, further thought must also be given to how they should be used and reported. Some of these considerations include:

1. Determining the relative importance of the proposed measures and indicators.

   What are the criteria for establishing the relative importance of the proposed measures and indicators in describing data quality?

2. Determining the appropriate periodicity for reporting the various proposed measures and indicators.

   Since data from the field are sent to BLS monthly, reporting at monthly, quarterly, or annual frequencies, it may be useful to report some indicators and measures at one or more of the frequencies.

3. Utilizing the longitudinal aspect of the data.

   Although the Interview Survey is a five-wave panel survey, the production of official survey estimates does not utilize the longitudinal feature of the Interview Survey data, but treats each interview of a sample unit as an independent observation. Panel surveys pose different concerns from cross-sectional surveys – such as panel attrition and correlated errors between observations. Thus, the assessment and monitoring of the Interview Survey's data quality should incorporate examining some of the proposed measures and indicators longitudinally as well.

4. Further research

   Continuing work is needed to identify additional and / or better indicators and measures of data quality, understand how to use them, and their relationship with survey features such as interview length, proxy reporting, and interviewing structure. These research objectives have been raised for the larger Gemini Project.

5. DATA QUALITY INFORMATION SYSTEM

The preceding sections describe a conceptual framework that facilitates the measurement and assessment of data quality dimensions relevant to the CE's goal of meeting the needs of primary users of its data. While it is important to systematically collect appropriate information for assessing and monitoring data quality, an important complement to the data collection effort is the development of a data quality information system that supports the systematic assessment, monitoring, and reporting of data quality. It is important to have a data quality information system because:

- a known, single repository that houses all the quantitative and qualitative data on data quality (analogous to survey data on CERSCH) lends credence to the importance of data quality as an ongoing part of CE operations. The commitment of resources to developing and maintaining a data quality information system "signals" the importance of data quality to CE;
- it gives integrity to the data quality information since it will be treated as an integral part of data production;
- it facilitates consistent and comparable research on data quality across time, as well as from changes in the survey.

It is beyond the scope of this team to provide a concrete description of a data quality information system in this proposal. However, the importance of having this information system warrants an attempt to provide an outline of desirable elements of this information system for the CE, which will hopefully engender further reflection and discussion on the development of a data quality information system.

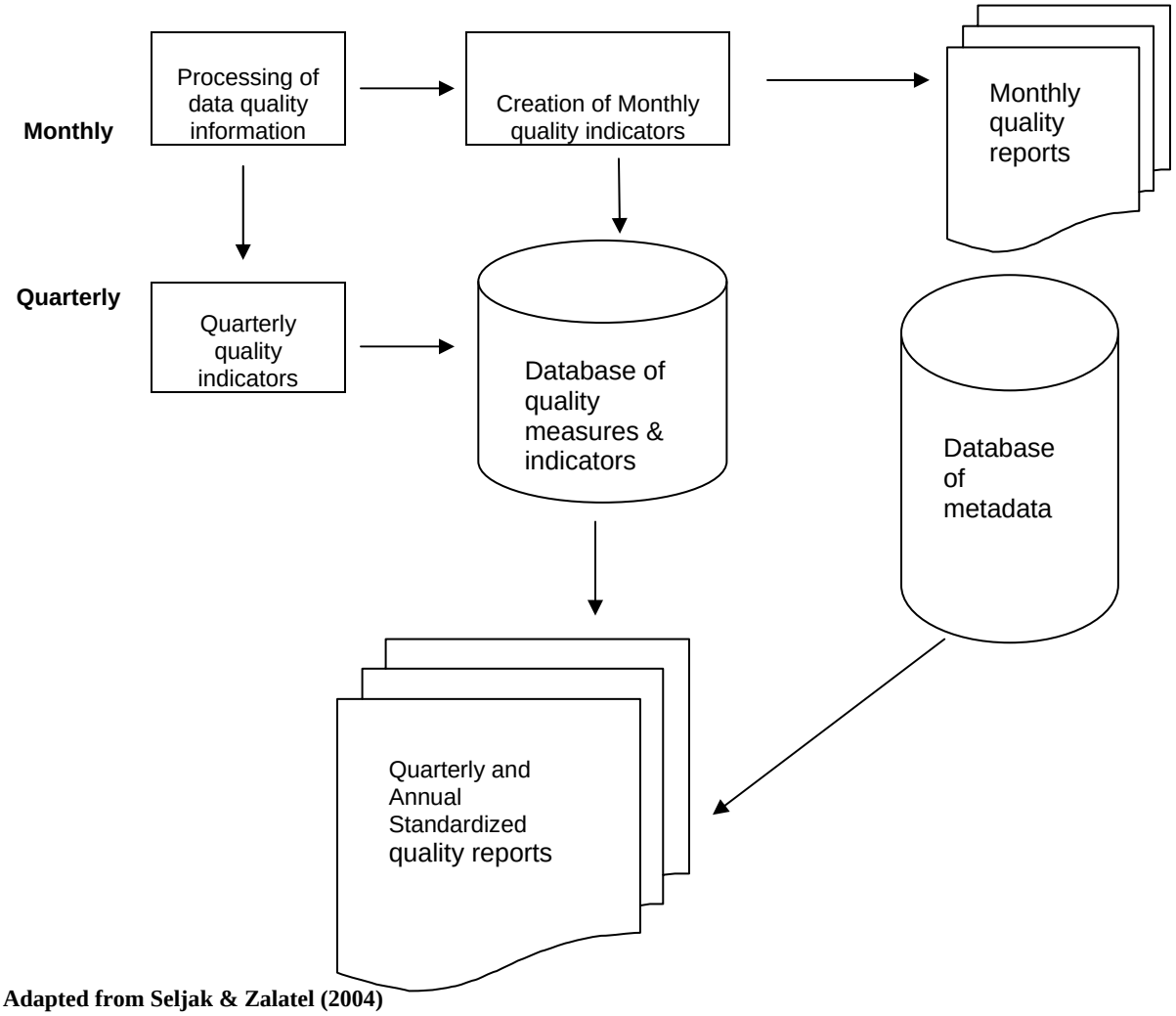
A simple high-level representation of the key components of a data quality information system for the CE is sketched out in Figure 2 (adapted from Seljak and Zaletel 2004). The key components are:

- collection and processing of data quality inputs
- storage of both quantitative and qualitative information, including survey metadata
- production of reports at different periodicities and levels of detail, for reporting internally and externally

Currently, CE already routinely produces annual and quarterly data collection quality monitor reports that are made available on the CE intranet. For the Diary Survey, the quarterly report shows rates on mode of collection, double placement, recall aids usage, final disposition of cases, response rate, length of interview, and the number of cases that failed the minimal expenditure edit. For the Interview Survey, the quarterly report shows rates on mode of collection, recall aids usage, final disposition of cases, response rate, length of interview, and data editing rates. For each indicator, the report shows the rate by Regional Office in a table for the reference quarter, and a time series graph of the rate at the national level through the reference quarter. We propose that this reporting system be expanded to incorporate the aspects of data quality addressed in this report.

In addition, consideration should also be given to reviewing existing data quality/metadata information systems in CE's Branch of Production and Control, and Branch of Information and Analysis to see if/should/how these systems can be integrated for more efficient storage, maintenance, and access to these information.

Figure 2. Illustration of Key components of a Data Quality Information System *



*** Adapted from Seljak & Zalatel (2004)**

6. SUMMARY

This team was tasked to propose a data quality definition for the CE that would be attentive to the "fitness-for-use" concerns of individual program stakeholders, and to propose how the definition could be operationalized to routinely assess the overall quality of the CE data.

We recommend that a data quality definition that is to be responsive to the "fitness-for-use" concerns of primary data users be multidimensional. Our proposed data quality definition is comprised of the six dimensions from the TQM paradigm – Relevance, Accuracy, Coherence, Timeliness, Accessibility, and Interpretability. In addition, we adopt the TSE paradigm to enumerate the sources of error that affect accuracy. The combination of the TQM and TSE paradigms in the data quality definition help to bring to the forefront the varied factors that affect data quality which any proposed modifications to the survey should consider. An attempt to improve one dimension or error source could have unintended adverse impacts on other dimensions or error sources. While practical considerations may force trade-offs to be made between data quality dimensions and error sources, these trade-offs (to the extent foreseeable) should be the result of conscious decisions, and not unintended consequences.

Having proposed a definition for data quality, we outlined a procedure for using the definition to guide the collection and documentation of data quality concerns of each data user in a consistent and systematic manner. We provided an illustration of this procedure with three uses of the CE data (Published Tables, CPI, and the Experimental Poverty Thresholds), and showed how we used this information to identify common and distinct data quality concerns among these data users. Common data quality concerns shared among multiple data users become natural candidates for prioritization in ongoing efforts to improve the CE.

For the assessment and monitoring of data quality, we suggested various indicators and measures for each of the six TQM dimensions. In addition, indicators and measures for measurement error were also proposed, and issues for consideration when determining how to implement these indicators and measures were raised.

Finally, we also proposed establishing a data quality information system that integrates databases of measures and indicators with metadata of the survey. While it is important to systematically collect appropriate information to assess data quality, an important complement to the collection effort is the development of a data quality information system that supports the routine and systematic assessment, monitoring, and reporting of data quality across production phases as well as across time. The perspective that metadata management detracts from the primary mission of collecting and reporting data reflects a misunderstanding of how a robust, integrated metadata system is integral to quality control that affects the entire survey process - from data collection through generating survey estimates, and user documentation. We strongly advocate that a major redesign of the CE is an opportune time to build an integrated data quality information system. While we realize it will require a significant investment in resources (i.e., there are large initiation costs) to build this information system, it should be the 'big picture' that guides the incremental development of such a system. As individual databases are built, they should be done so with integration in mind.

REFERENCES

Australian Bureau of Statistics. 2009. ABS Data Quality Framework. http://www.abs.gov.au/AUSSTATS/abs@.nsf///mf/1520.0?opendocument?utm_id=LN (accessed May 20, 2009).

Biemer, Paul P., and Lars E. Lyberg. 2003. *Introduction to Survey Quality*, New York: Wiley.

Brackstone, Gordan. 1999. "Managing Data Quality in a Statistical Agency," *Survey Methodology*, 25(2): 139-149.

Citro, Constance F., and Robert T. Michael (eds.), 1995. *Measuring Poverty: A New Approach*, Washington, D.C.: National Academy Press.

Chicago Manual of Style Online. http://www.chicagomanualofstyle.org/tools_citationguide.html (accessed July 17, 2009).

Dobbs, Joy, Catherine Gibbins, Jean Martin, Pam Davies, and Tricia Dodd. 1998. Reporting on Data Quality and Process Quality. Survey Research Methods Proceedings from the annual American Statistical Association Joint Statistical Meetings, pp. 32-40.

European Statistics Code of Practice. 2005. http://www.dzs.hr/Eng/international/code_of_practice_en.pdf (accessed May 20, 2009).

Groves, Robert M., Floyd J. Fowler J, Mick P. Couper, Jame M. Lepkowski, Eleanor Singer, and Roger Rourangeau. 2004. *Survey Methodology*, New York: Wiley.

Groves, Robert M. 1989. *Survey Errors and Survey Costs*, New York: Wiley.

Lee, Geoff and Bill Allen. 2001. Educated Use of Information about Data Quality, Australian Bureau of Statistics, Paper presented at the International Statistical Institute (ISI), 53rd Session, August 22-29, Seoul, South Korea. http://isi.cbs.nl/iamamember/CD2/pdf/1076.PDF (accessed May 20, 2009).

Office of National Statistics. 2007. Guidelines for measuring statistical quality. http://www.statistics.gov.uk/downloads/theme_other/Guidelines_Subject.pdf (accessed May 20, 2009).

Seljak, Rudi and Metka Zaletel. 2004. Measurement of Data Quality in the Case of Short Term Statistics. Paper presented at the European Conference on Quality and Methodology in Official Statistics, May 24-26, Mainz, Germany. http://www.oecd.org/dataoecd/6/8/35238995.pdf (accessed July 15, 2009).

Statistics Canada. 2003. Statistics Canada Quality Guidelines http://www.statcan.gc.ca/pub/12-539-x/12-539-x2003001-eng.pdf (accessed May 20, 2009).

Notes on Data Quality Issues Discussion with Data Use Contacts – Internal documents

Official CE Estimates for Tables Published on CE Public Website and Biennial Table Publication (referred to as *Published Tables*), William Passero
\\psbres2\dces-public\Teams\Gemini\3 Data Quality\DataUsers\UserMeetingNotes\20090501_ meeting with BPassero.doc

CPI, Robert Cage
\\psbres2\dces-public\Teams\Gemini\3 Data Quality\DataUsers\UserMeetingNotes\20090508_ meeting with RCage.doc

Experimental Poverty Thresholds, Thesia Garner
\\psbres2\dces-public\Teams\Gemini\3 Data Quality\DataUsers\UserMeetingNotes\20090514_ meeting with TGarner.doc