**A modest attempt at measuring and communicating about quality**

Laurie Reedman, Statistics Canada, Laurie.Reedman@Canada.ca

**Abstract**

*While we were mostly disseminating aggregate statistics derived from a sample survey, we used the sampling error as the backbone of quality reporting. Now we are moving towards disseminating aggregate statistics derived from non-survey methods, and disseminating micro-data products. What should we be communicating about the quality of these products? What do we as data producers need to know about the quality of data is it traverses our processing steps, and what do the users of our data products want or need to know about its quality? This paper explores these questions in the Statistics Canada context. We are looking for common vocabulary, and a standardised way to represent different aspects / attributes / dimensions of quality. In particular we are exploring ways to measure and report accuracy, and how to reflect the impact on accuracy of processing steps such as data integration, imputation and disclosure avoidance. Finally we look at how to summarize and communicate the accuracy of a data product in a way that informs use.*

**Keywords:** quality, accuracy, quality reporting, quality dimensions

**1. Introduction**

At a National Statistical Organization (NSO) the methods and language for measuring and communicating about data quality mostly originate from sampling theory. Now as we are advancing towards a survey-supported business model rather than survey-centric, the time is right to incorporate and develop terminology and methods for measuring and reporting data quality when the data comes from sources other than sample surveys.

Section 2 of this paper proposes a quality indicator framework and explores how to communicate about quality in metadata. Section 3 looks briefly at how to measure accuracy in the absence of sampling error. Section 4 offers ideas for further study.

**2. Quality indicator framework**

*2.1 Quality indicators, indicator frameworks*

A good indicator is objective and measurable. According to a description on the [Public Health Agency of Canada’s website](https://infobase.phac-aspc.gc.ca/datalab/indicator-framework-blog-en.html), “an indicator framework is an organized way to view data from different sources. It is a simple and concise way to present gathered data and help show the relevance and connection between different indicators. In a framework, data can be grouped or categorized and are often shown alongside detailed descriptions of associated measures and methods of collection.” [Documentation on the Indicators and a monitoring framework for the SDGs](http://indicators.report/overview/) states that an indicator framework can be used as a report card, at a point in time, to inform decisions about using a particular dataset, or as a management tool, to monitor performance through time. A generic framework would include all possible indicators but not all indicators would be relevant for every dataset or for all audiences. It will be very important to engage data users in designing indicator frameworks so that their needs for quality reporting are met.

So the question is, which quality dimensions should be represented in a quality indicator framework? The commonly used dimensions of quality (relevance, accuracy and reliability, coherence and comparability, interpretability, timeliness and punctuality, and accessibility) retain their relevance and importance in any context. However in the context of data produced outside of the NSO there are other facets of the data that should also be noted. When we are deciding whether or not to use a particular dataset, we consider the perception of authority and credibility of the data producer. Are we willing to trust data on this topic from this source? We would also like to know what quality assurance practices, if any, were followed in the production of the data. We consider the cost in terms of skills, time, tools and resources to process, archive and curate the data, weighed against the value to be gained from using it. We look for unique identifiers that can be used for direct matching with other datasets, or characteristic variables that can be reliably used in probabilistic matching. We also consider the reference period of the dataset, the periodicity on which we expect to receive the data, and the coherence of definitions, concepts and formats with recognized standards.

Once we move from the input to the throughput stage, the quality focus shifts to the accuracy of the data. We clean and edit to improve internal consistency across records. We can use imputation to compensate for item non-response. We can use imputation or calibration techniques to compensate for missing units or bigger gaps in coverage. We link to other data sources to improve coverage and the analytical capacity of the data. We can measure the success and impact of these activities by calculating edit failure rates, imputation rates, and diagnostics of record linkage activities. Validation against other sources is always an important quality assurance practice, and it becomes even more important when we do not know very much about the underlying design or how the data was collected or processed. Summary statistics and other diagnostics from the validation activities contribute to our understanding of the variability and bias of the data.

Ultimately a dataset must be relevant to be useful. It is our responsibility to provide data users with enough information that they can make informed decisions about the relevance of our datasets for their purposes. This includes but is not limited to the reference period, target population, unit of observation and definitions and concepts of the characteristic variables.

Data users are perhaps most unforgiving about timeliness, therefore we must constantly strive to make our processes as efficient as possible. To facilitate timely dissemination and efficient processing many NSOs have standardized metadata templates and dissemination practices, which also serve to assure both interpretability and accessibility of the data and metadata. A possible compromise between relevance, accuracy and timeliness would be to release a preliminary dataset quickly, and then release a revised version after more data processing and validation has been done.

A possible quality indicator framework including all the above mentioned quality dimensions is presented in Table 1. The leftmost column confirms that all of the quality dimensions should be measured and monitored by the NSO. The centre column proposes which ones could be communicated to data users.

Table 1: Quality Indicator Framework

|  |  |  |
| --- | --- | --- |
| Monitor Internally | Report Externally | Quality Dimension |
| Yes | Yes | Credibility |
| Yes | yes | Quality Assurance practices followed |
| Yes | Maybe not | Processability |
| Yes | Yes | Linkability |
| Yes | Yes | Coherence with standards |
| Yes | Yes | Accuracy |
| Yes | Maybe | Reliability |
| Yes | Yes | Coverage |
| Yes | Yes | Bias |
| Yes | Sort of | Relevance |
| Yes | Maybe not | Timeliness |
| Yes | Metadata | Interpretability |
| Yes | Maybe not | Accessibility |
| Yes | Yes | Coherence with other sources |

*2.2 Stakeholder engagement*

Stakeholder engagement is critical to effective communication about quality. It would be interesting to know how data users have made use of accuracy indicators such as coefficients of variation or confidence intervals. Do they do anything different on their side if the confidence interval of one variable is wide while that of another is narrow? This discussion would inform how we communicate about accuracy for data that does not have sampling error. Perhaps we can seize this opportunity to do some capacity building among data users, at least those who are making analytical use of the data and would be motivated to learn. Other users might prefer a narrative approach, to better understand the concepts and to “get a feel for” the quality of the dataset. Still others might resonate to a simple indicator such as an infographic.

We hear frequently that data users want faster access to more detailed data. Do they understand that for the NSO to provide faster access, it means that some of the processing or validation would have to be left not done? Also, while we strive for reliable numbers at aggregate levels, accuracy and reliability erode at finer and finer levels of detail. Stakeholder consultation is needed to explore their expectations for speed, detail and accuracy.

An idea worth exploring is certification to some recognized standard. Do data users respect the authority of those doing an independent assessment, and does the certification imply credibility? The certification criteria would certainly include quality dimensions related to the data products, and would ideally also reflect the quality of the statistical processes which produced the data.

*2.3 Metadata about quality*

At a minimum, the metadata about quality should include the items in the quality indicator framework. These could be presented in a list, similar to a nutritional label on food items, with all the indicators in the same order and reported using standardized units of measure. Alternatively, a composite quality index would facilitate comparisons between datasets, however at this time we are far from having standardized quantitative inputs to create an index. An info-graphic such as a spider-web graph or symbols convey the least amount of information in the quickest amount of time, and might be appreciated by casual users. Given that data users cover a wide range in terms of their data literacy and the ways in which they use data, perhaps the best approach would be to provide different layers of metadata that target different audiences with the appropriate level of information.

The metadata about quality should also include the assumptions and compromises that were made in the cleaning, processing and analyzing of the data, as well as an explanation of why the assumptions and compromises were made and what is their impact on the usability of the data. For example, we might choose to ease the response burden of a survey by reducing the sample size and removing questions from the questionnaire, but the consequence could be weakened correlations between variables to use in modeling. Limitations in the data whether as a result of our decisions or otherwise should be described in the metadata.

Changes that cause problems for data users and should be described in metadata include a break in series, definitional changes, changes in collection methodology, and changes in management systems. Additional information that should be described in metadata includes accounting methods in administrative or commercial records, differences within or across data sources, conceptual or operational definitions, and unit definitions.

**3. Measuring and reporting accuracy**

Accuracy poses an interesting challenge when we do not have sampling error. When we use non-survey data there are potential sources of error that could be introduced at the NSO through our own handling of the data (processing error, misalignment of concepts, reference periods and classifications, data integration errors) as well as errors that could be in the data before they even get to us (bias, measurement error, coverage error). Our decisions about to what extent we explore, identify and fix errors in the incoming data are based on various things, such as our capacity to do so and a cost/benefit analysis. When we report on data quality should we include reference to errors that we found and corrected in data sourced elsewhere, or should we only report on the outgoing quality of disseminated products? Is there a need for data users to know about internal data profiling and quality improvement activities? We monitor the effort versus the value of these activities as part of our own resource management, but to what extent should this be reported to data users?

Non-sampling errors can be systematic, which results in bias, or random, which results in increased variability in the characteristics of interest, or noise. Statisticians have always found it challenging to measure bias. In the context of non-survey data, perhaps confrontation against other data sources can reveal the direction of bias and possibly even its bounds. Beaumont and Charest (2012) suggest a generalized bootstrap method for estimating model parameters that has some potential for application to non-survey data if we can assume that the data are from some super-population. Van Delden et al (2014) explore several different methods for estimating accuracy of non-survey data. They describe a sensitivity analysis on the effect of classification errors on quarterly turnover figures derived from tax data. They make some simplifying assumptions and apply a parametric bootstrap method. They suggest a resampling method that could correct for biased bootstrap confidence intervals. The bootstrap method appears to have some potential for estimating noise under certain circumstances.

**4. Next steps**

This paper poses more questions than it answers. The author welcomes collaboration with others who are facing the same challenge, to find methods, standards and terminology to facilitate measuring and communicating about quality for all types of data.

**5. References**

Barbalace, F., Brancato, G., Signore, M., Simeoni, G. (2016). Introducing a Framework for Process Quality in National Statistical Institutes. European Conference on Quality in Official Statistics (Q2016) Madrid, 31 May-3 June 2016.

Beaumont, J. F., Charest, A.-S. (2012). Bootstrap variance estimation with survey data when estimating model parameters. Computational Statistics and Data Analysis, 56 pp. 4450-4461.

Gudivada, Venkat N., Apon, Amy, Ding, Junhua. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. International Journal on Advances in Software 10.2, pp. 1-20.

Hamer, Daisy. Guidance for Measuring and Reporting on Quality when Administrative Data is used to Supplement or Replace Survey Data. Office for National Statistics, UK.

Kenen, Ron S., Shmueli, Galit. (2016). From Quality to Information Quality in Official Statistics. Journal of Official Statistics, Vol. 32, No. 4, 2016, pp. 867-885.

Pobjoy, Jill. (2016). Quality within ONS – Providing a framework for statistical producers and assurance for our users. European Conference on Quality in Official Statistics (Q2016) Madrid, 31 May-3 June 2016.

Reid, Giles, Zabala, Felipa, Holmberg, Anders. (2017). Extending TSE to Administrative Data: A quality Framework and Case Studies from Stats NZ. Journal of Official Statistics, Vol. 33, No. 2, 2017, pp. 477-511.

Van Delden, A., Scholtus, S., de Wolf, P-P., and Pannekoek, J. (2014). Methods to assess the quality of mixed-source estimates. Statistics Netherlands report.

Zhang, L-C (2009). A unit-error theory for register-based household statistics. Discussion Papers No. 598, December 2009, Statistics Norway, Statistical Methods and Standards.

Zhang, L-C (2012). Topics of statistical theory for register-based statistics and data integration. Statistica Neerlandica (2012) Vol. 66, nr. 1, pp. 41-63.