Standards and Guidelines for Reporting on Coverage

Martin Provost

Statistics Canada
R. H. Coats Building, 15th Floor, Tunney's Pasture, Ottawa ONT, Canada, K1A 0T6, martin.provost@statcan.ca

Key Words: Quality indicators, coverage, ideal, target, survey and frame populations

1. Introduction

The level of coverage provided by a survey or a statistical program has a major impact on the overall quality of the survey process. It is therefore important that users of the survey data be informed of the coverage quality provided by the survey and of any potential problem associated with identified defects in terms of coverage. Those measures of coverage add to the set of quality indicators that together provide a complete picture that facilitate an adequate use of the survey data and an optimization of the survey resources to allow for its constant improvement. But how should one define coverage (based on what units, etc.)? And how should survey managers report on it? This paper outlines the content and describes the procedures, issues and topics which will have to be covered in standards and guidelines for reporting on coverage at Statistics Canada to standardize the terminology, the definitions and the measures relating to these coverage issues. These standards and guidelines have been proposed for inclusion in Statistics Canada's Quality Assurance Framework as a supplement to this agency's Policy on Informing Users of Data Quality and Methodology.

2. Statistics Canada's Quality Assurance Framework

Statistics Canada (STC) has always invested considerable effort in achieving an adequate level of quality for all of its products and services. Over the years, those products and services have evolved, as well as the methods and the technology used to produce and support them. The measures required to manage the quality of STC's products have evolved accordingly. In 1997, Statistics Canada drafted a description of the measures it had in place to manage the quality of the data it produced in order to situate existing policies or practices within a common quality management framework. This was called "Statistics Canada's Quality Assurance Framework" (QAF). The QAF was revised and updated in 2002.

Statistics Canada's product is information. Measuring the quality of this information is not a straightforward task. There is no one single measurement that would give an absolute measure or a rating of the quality of its products. There are many things that have an impact on whether or not the data or analysis produced are of adequate quality and those things are not necessarily even intrinsic to the information itself, but rather are linked to the users and their intended use of the information. That is why Statistics Canada defines the quality of the information produced in terms of its "fitness for use". To determine whether or not a product or service is fit for use, the concept of quality has been broken down into six dimensions. As defined in the QAF, those dimensions are:

- Relevance: reflects the degree to which it meets the real needs of clients.
- Accuracy: is the degree to which the information correctly describes the phenomena it was designed to measure.
- Timeliness: refers to the delay between the reference point (or the end of the reference period) to which the information pertains, and the date on which the information becomes available.
- Accessibility: refers to the ease with which it can be obtained from the Agency.
- Interpretability: reflects the availability of the supplementary information and metadata necessary to interpret and utilize it appropriately.
- Coherence: reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time.

Since quality is defined as fitness for use, any quality rating would have to permit each user to set its own priorities among the dimensions. Some would put more emphasis on timeliness and others on accuracy, for instance. This highlights the fact that managing and achieving quality is really a matter of finding an adequate balance between each of the six dimensions.

As a reference document, the QAF relates to many policies that have been developed to ensure the proper management of quality measures and to extend some of the best practices of the Agency. There is one policy that is more closely related to

the topic of this paper: the Policy on Informing Users of Data Quality and Methodology (PIUDQM). Making sure that users of its data have access to all the information they need to determine if a product would satisfy their needs, and if its quality is good enough for their intended uses is the objective of the PIUDQM (policy 2.3 in Statistics Canada's Policy Manual).

3. Policy on Informing Users of Data Quality and Methodology

The Policy on Informing Users of Data Quality and Methodology (PIUDQM) states that:

- Statistics Canada will make available to users indicators of the quality of data it disseminates and descriptions of the underlying concepts and methodology.
- Statistical products will be accompanied by or make explicit reference to documentation on quality and methodology.
- Documentation on quality and methodology will conform to such standards and guidelines as shall from time to time be issued under this Policy.

It also indicates its scope: This policy applies to all statistical data and analytical results disseminated by Statistics Canada however collected, derived or assembled, and irrespective of the medium of dissemination or the source of funding. The document describes the reasoning behind the policy and explains everyone's responsibilities with regards to the observance of this policy. The major part of the document is made up of the Standards and Guidelines, describing the minimum mandatory requirements under the Policy (the Standards) and the additional documentation on methodology and data accuracy that should be provided for major surveys, censuses or programs (the Guidelines). The Standards and Guidelines are designed to summarize the set of measures and elements of information that should accompany each of our products, as well as to help each program in presenting those in a standardized way.

Although the Standards and Guidelines included in the Policy describe the topics that should be covered in the documentation and the way those should be arranged, they don't go into the details of how we should report on those topics. For instance, the standards require an estimate of coverage error, a response rate, imputation rates, etc., but do not detail how those measures should be obtained. To go even further in standardizing the measures of quality and the way to inform users about those, other standards and guidelines are developed. A document called *Standards and Guidelines for Reporting Nonresponse Rates*, for example, was designed to standardize the way each survey reports on the item of non-response.

The present article describes a framework for developing similar guidelines for reporting on coverage. These standards and guidelines would standardize the terminology and definitions, standardize or give guidance to the description of the target and survey population required by the PIUDQM; particularly in regard to limitations affecting conclusions, coherence and interpretability, analysis and comparability with other data sources. The next section will define the various populations encountered in the design and implementation of a survey, along with the related coverage issues. Then, ways to measure and evaluate the coverage of a study will be described, as well as what should be reported about coverage. The guidelines will apply equally, with some differences, to surveys, censuses and usage of administrative records.

4. Defining Coverage

4.1 Populations

In the literature, we usually find references to the concepts of target population and survey population.[1] The first one refers to the population that should be covered by the study and the second one refers to the population that the study will really try to cover, the goal being to have the smallest possible difference between those two populations. To effectively describe the survey process, we would in fact need to define four populations. Even for censuses or studies done using administrative records, those four concepts of population still hold.

First we start with a *conceptual universe* or *ideal population*. This is the theoretical definition of a sometimes changing and renewable population that we would be interested in and is often in practice something that can't be observed. Then we have the *target population*, which is a rationalization of this ideal population into a concrete universe that can be defined and delimited. For instance, the conceptual universe could be the elementary students going through the school system in Canada

[1] Some of the concepts presented in this section dates back to Kish (1965). The Standards and Guidelines, as well as the longer version of this paper that will be included in the conference proceedings, will incorporate a review of the literature concerning those definitions and also the way coverage errors are handled and reported on in various statistical agencies and organizations - see for instance Statistics Finland (2002) or OMB (2001).

and the target population all elementary students in Canada in a given year or for a given period of time. This target population is a static image of the conceptual universe at one of its "stages of evolution". Clearly, for research, regulation or decision-making purposes, one could be interested in something way more general than just this restricted population but it is not something that can be observed and measured. One has to "operationalize" the definition in order to work with it in practice. This is what the target population does. This is why as far as coverage is concerned, the starting point is really the target population.

Although the target population can be completely defined and delimited, it is still usually an optimistic target. It is the population that would be covered if there were no constraint in terms of time or budget, as well as no administrative and operational limitations. The third population encountered in the design and implementation of a survey is the *survey population*. It is the population that the aforementioned constraints force us to narrow down to, by excluding some units harder to identify, to reach or interview. This is the first coverage issue that we should be concerned with and users should be informed about it. Most of the time, the survey population is a subset of the target population, but this is not always the case.

In order to select and contact units from the population, a frame, a combination of frames or a hierarchy of frames is used. The fourth population will then be called the *frame population*. It is composed of the units that are included on the frame. There are almost always differences between the survey population and the frame population. Sometimes the chosen (or constructed) frame is known to be less than perfect but is the only available option.

Sometimes, because of flawed estimation procedures or non-response adjustment, for instance, the survey results are depicting a population that differs from the survey and frame populations (this could have constituted a fifth population, the measured population). For example, the results could show a higher proportion of older people than is actually the case. This difference is not part of the coverage errors but it is sometimes hard, when trying to evaluate some components of the coverage error, to differentiate those from the "representation" errors.

4.2 Coverage Limitations

There are two types of coverage limitations: the *deliberate restrictions to coverage* and the *coverage errors*.

4.2.1 Deliberate restrictions to coverage. The deliberate restrictions to coverage refer to the difference between the target and survey populations. As mentioned above, the survey population is usually a subset of the target population. For instance, some remote units are omitted from the survey population because of their high collection cost relative to their impact on the estimates. Such exclusions are not an uncommon issue and the reasons for these may be varied: in social surveys, Military, homeless, institutionalized people, among others, are often excluded from the survey population. Exclusions can be caused by the collection vehicle, for instance households without a conventional phone being excluded from telephone surveys.

But the survey population is not always a subset of the target population. In fact, the two could refer to totally different types of units. For instance, one could be interested in studying income of individuals under various conditions (those individuals would form the target population) but end up using an administrative file that is structured by household (those households would make up the survey population). In such an extreme case, it would be up to the researcher to determine if the administrative file will serve his purpose, but such choices frequently must be made.

4.2.2 Coverage Errors. The coverage errors refer to the discrepancies between the survey population and the frame population. These frame imperfections are of two types: the *undercoverage* and the *overcoverage*.

Undercoverage: units can be missing from the survey frame(s). This can be caused by administrative delays, the use of an outdated list, sampling units that are not correctly defined or simply by the choice of an inadequate frame. It is a serious problem, especially since those missing units can be different from those present on the frame in terms of the variables of interest for the study. Their exclusion can bias the survey estimates if no compensating corrective measure is taken.

Overcoverage: units can be erroneously included on the frame (inclusion of deceased or other types of out of scope units) or included more than once. In multi-stage sampling, duplication can be caused by definitions or procedures not detailed enough, causing second stage units to be included in more than one first stage units. For example, children in joint custody included in both parent's households or an inadequate design of geographical units in an area frame. Duplication can also be caused by an inadequate identification of frame overlap when multiple frames are used. Overcoverage problems are usually not as serious as those of undercoverage because as long as the wrongfully included units can be adequately identified, no bias will be introduced in the survey estimates. Nevertheless, this loss of sampling units will increase the sampling variance.

With a multi-stage design, the coverage should be assessed at each stage, with a focus on the impact that the coverage errors will have on the coverage of the units at the last stage of the sampling process and, if different, on the units used for the analysis. For example, say we conduct a survey of children under five in a certain city, using a multi-stage design where the city is first divided into neighbourhoods, some of which are randomly selected, then a list of all private households within the selected neighbourhoods is established and a sample of households is selected. Finally, information is collected about each child under five in the selected households. In this case, we are interested in children but the real coverage issue is rather at the household level. We should first evaluate if the neighbourhoods are non-overlapping and covering the whole city. This will indicate the quality of the coverage at the first stage of the design. Then we should look at how good the listing of households was. This is where the most important part of the coverage error will be found. Ultimately, we should evaluate if all children under five that ought to have been linked to the selected households indeed were. When multiple frames are used, it is the coverage provided by the combination of the frames that is important in terms of measuring and reporting on coverage, although information about coverage of each individual frame will prove very useful in improving the design.

In some cases, response errors might be confused with coverage errors. In order to differentiate these, one should ask what are the units of interest in terms of sampling and analysis. Two examples will make this clearer. In the Census of population, each household is contacted and a list of all persons living in those households is obtained. Because the Census is interested in counting both the households and the persons, omitting a person from the list for a particular dwelling constitutes a coverage error. The listing of the persons in each household should be perceived as one more stage in the sampling design. On the other hand, if a sample of businesses is selected and data about the number of employees are collected for each firm, a business omitting to count some of its employees wouldn't cause a coverage error. This would rather be considered as a response error. This distinction comes from the fact that in that case, we are interested in businesses and the number of employees is only an additional piece of information that we want to get from the selected units.

4.2.3 Representation Errors. There are many things, besides frame imperfections, that can cause the measured population to be different from the survey population. Even with a perfect frame, the two populations could be different. For example, even in the straightforward case where a simple random sample is selected from a complete list of all units in the population, if the level of non-response varies within subgroups of the population, and if that is not taken into account in the non-response adjustment and/or estimation process, then the measured population will not be a correct representation of the survey population. Some subgroups will be underrepresented and others will be over-represented. Those representation errors are not part of the coverage errors but since coverage is often evaluated by comparing estimates from the survey to numbers from other sources, it can sometimes be difficult to differentiate both.

5. Measuring Coverage

5.1 Assessing the Deliberate Restrictions to Coverage

Sometimes, it is possible to measure the size of the portion of the target population that was excluded from the survey population, or at least to get an idea of the importance of this excluded portion. This size could be expressed as an absolute number of units excluded or as a proportion of the whole target population. It is also sometimes possible to depict, using other sources, the characteristics of this portion of the target population and to get an idea of the impact this exclusion could have on the survey results. For repeated surveys, it is even more important to monitor the changes in the size of this population, and of sub-population characteristics**.**

For example, in surveys using Random Digit Dialing (RDD), households without a conventional phone are excluded from the survey population. Studies have shown that the national proportion of households without a telephone is less than 2% in Canada but that those households have characteristics different from those with a telephone (in terms of income, mobility, etc.). Although the level of telephone coverage remained constant over the last few years, the increasing popularity of cellular phones could change that fact. It is important to monitor the size and the characteristics of this excluded population in order to reassess from time to time the appropriateness of RDD methods.

Those evaluations are appropriate when the survey population is a subset of the target population. When it is not the case, like for the example given in section 4.2.1 where the survey population was made up of households and the target population of individuals, every effort should be made to compare the two populations, identify the differences and evaluate the impact on the survey estimates. In the example, it should be determined how well the households on the administrative file are covering the individuals the research was interested in and how the use of the household income instead of the individual's

could still provide useful results. Then in that case there could still be other sources of deliberate restrictions to coverage. Some small regions or some types of household could be excluded from the file. This has to be examined too.

5.2 Measuring Coverage Errors

As previously mentioned, undercoverage of the survey population on the sampling frame(s) can introduce a bias in the survey estimates. As for all that relates to bias, evaluating the quality of the coverage provided by the frame can be difficult and costly. Actually, the assessment of a potential bias can only be done by comparing with an external source. It is nevertheless possible to measure the coverage errors of the survey frame using different methods. Some of those methods are:

Evaluating the frame prior to use: Several techniques allow for evaluating the quality of a frame prior to using it and also evaluating it periodically for repeated surveys. Those techniques are described in section 3 of Statistics Canada's Quality Guidelines (1998). These methods - comparing the frame or a sample of it to a similar source, comparing it to an area frame, checking frame totals against an independent source, etc. - can, among other things, allow calculation of an undercoverage rate (estimated number of units missing on the frame divided by total/estimated number of units in the survey population) and an overcoverage rate (similarly defined), validate the appropriateness of using this frame and enable the implementation of corrective measures for coverage issues. One should keep in mind that in most cases, the external source chosen as the reference is not perfect either, will not be 100% comparable, and that a part of the estimated coverage error is caused by imperfections in and differences from the external source. Integration of survey frames - using the same frame for surveys targeting the same population - can help offset the cost of those types of evaluations. More and more STC surveys rely on the use of the Business Register, the Labour Force Survey frame, the Farm Register or the RDD technique. The most should be made of this convergence towards common sampling frames to permit more effective coverage assessment, as well as to permit better coverage.

Overcoverage as estimated from the sample: During the collection process, some sampled units may be classified as out of scope for the survey (deceased, nonexistent, etc.). Those are units that shouldn't have been included on the frame. Those units can be used to approximate the proportion of out of scope units on the sampling frame. Units present on the frame more than once are much harder to evaluate using the sample. The overcoverage as estimated from the sample is therefore a very limited measure for assessing the coverage provided by a frame, but is still useful.

Slippage rates: It is possible to take survey estimates of the size of certain groups (usually province*age*sex in social surveys) - adjusted for non-response but not poststratified – and compare those to totals from an independent source, usually demographic projections for the same groups. The percentage difference between the two sources is calculated for each group and is called the slippage rate. The slippage rate is a very useful measure to assess the coverage of a survey. It shows the subgroups that are underrepresented and over-represented in the survey estimates. Slippage rates are especially applicable for panel surveys or surveys for which the sample is selected in such a way that changes in the frame are not directly reflected or represented in the sample as initially selected.

Nevertheless, some drawbacks of those rates as measures of coverage are worth mentioning. Similarly to what was pointed out earlier, the slippage rates are calculated by comparing two estimates, using one as the "gold standard". This external source used for comparison is itself subject to errors and various differences that should be taken into account. Those errors will be reflected in the slippage rates, causing them to be higher or lower than in reality for certain groups. It is important to use the best possible source for comparison and to be well informed of all of its limitations. The most important aspect of slippage rates that should be considered is that they are not true measures of coverage, but rather an indication of the extent of the representation error. It includes the coverage errors but also a lot more than that.

Other indicators of coverage issues: The stability of the distribution of some classification or control variables over time (number of households per region, average household size, etc.) can indicate potential coverage problems and because of that, should be regularly monitored.

6. Reporting on Coverage

The Policy on Informing Users of Data Quality and Methodology is not just about declaring limitations and errors, but about conveying a clear understanding about what the data represent. This facilitates proper analysis, conclusions and comparison between sources and over time. The population and its coverage are clearly major aspects of that. As an issue, coverage and the large differences in the ideal or target population from the survey and frame populations are of even greater importance when the commonly or historically accepted population within the user community corresponds to either of the first two.

The Standards and Guidelines for Reporting on Coverage will distinguish between two types of coverage measures: some general standard measures that each survey would have to report on and some more detailed measures that should be produced for larger repeated surveys. The Standards and Guidelines will also include examples and enhanced definitions as per what was presented in sections 4 and 5. The minimum requirement in terms of reporting on coverage should be a coverage rate or coverage rating. There is still room for discussion as to what the most appropriate measures should be, depending on the type, frequency and importance of the data collection vehicle. Other indicators and evaluations of coverage could be provided to allow for appropriate use of survey results and survey planning and improvement:

- All surveys should define their target and survey population: what is the population that would have ideally been covered, what restrictions were made and why. Also, a general idea of the importance of the excluded units should be given if possible. When the units in the survey population are not exactly of the same type as those in the target population, a justification should be given, with an analysis of the impact. Any issues in terms of limitations or that may affect comparability with other sources should be mentioned. This also holds for the other points mentioned below. There could also be a description of the conceptual universe if deemed necessary.

- A description of the survey frame(s) used should be provided. It could be explained how the frame was created and for repeated uses, what are the mechanisms for updating the frame. If other frames were considered, it could be mentioned, while describing what motivated the choice of the actual frame. We must also give information about the elapsed time between the creation or last update of the frame and the start of the data collection.

- For coverage errors, a minimum would be to explain, as part of a discussion on non-sampling errors, that there are coverage issues and define the concepts of undercoverage and overcoverage. We should also explain what causes those errors, in general (deaths, etc.) but also in the specific case of the survey: what could be the existing problems – for instance, delays for adding new units to the frame, etc. - and their impact. That is where the coverage rate or coverage rating should be provided.

- When using multiple frames, we could give an approximation of the coverage of each frame, as well as explain how the coverage will be improved by the use of a combination of frames and describe what will be done to take into account the overlap between the frames, if any.

- We should describe every measure taken to correct coverage problems – for example weighting adjustment - and the expected effect of those measures.

- The other measures described in section 5.2 are not part of the mandatory reporting on coverage but any coverage measure other then those mentioned in the minimum requirements that is available should be reported.

7. Conclusion

Information on coverage is important in designing and improving surveys but is also a very important tool to put in perspective the results of those surveys. It is therefore essential that the users be informed of all available indicators and information relating to the quality of the coverage, as this is one of the most important "fitness for use" issues for users. This is the goal of the Standards and Guidelines for Reporting on Coverage.

References

Kish, L (1965), Survey Sampling, New York: Wiley

OMB (2001), Measuring and Reporting Sources of Error in Surveys, Statistical Policy Working Paper 31, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget

Statistics Canada (2002), Statistics Canada's Quality Assurance Framework, Statistics Canada publication no. 12-586-XIE.

Statistics Canada (1998), Quality Guidelines, Statistics Canada publication no. 12-539-XIE.

Statistics Finland (2002), Quality Guidelines for Official Statistics, Handbooks 43b