

Comment

*Pilar Martín-Guzmán*¹

This excellent and very comprehensive article by Platek and Särndal deals with one of the main challenges that a Statistical Office has to face: how the quality of the statistical production can be evaluated, and how this quality is appreciated by the users. In a time of growing competition with the private sector, the image of quality is a crucial asset for these Offices.

But here comes a first question: to what extent is the image of quality correlated with what we, statisticians, would consider quality?

Starting from this point of view, I am going to comment on some of the many stimulating questions presented in this article. I will concentrate on three basic issues:

- A. The different perceptions of quality.
- B. Quality as perceived by NSO.
- C. The development of a total quality model.

A. Quality and image of quality are not exactly the same thing. The first is an objective concept, that can be subject to measurement when appropriate instruments are available. The second is the general perception of society, that results mainly from the subjective evaluation made by the users. There should be a certain positive correlation between these two concepts, but it does not necessarily have to be strong. On the other hand, the level of this correlation is subject to a considerable variation in time and space, for it is greatly affected by the cultural characteristics and the level of statistical literacy of the country.

The image of quality is essential for an NSO, because it is deeply related to credibility, and credibility is the basis of the virtuous circle that helps to improve the quality of statistics. Credibility is the most important asset for an NSO, but it is also a very frail item: it takes years to consolidate but it can be very easily damaged. That explains the fact that NSOs are increasingly interested in monitoring their external images, and that a growing number of them are carrying out surveys of user satisfaction on a regular basis.

How do users perceive quality? The example of the Rolls-Royce is a good one, because every driver knows what kind of services to expect from a good car. But for many users it is not so clear what can be expected from a good statistic.

The five dimensions stated in the Encyclopedia of Official Statistics referred to in the article are important to the users. But they can effectively evaluate only four of them. When confronted with the question, every user will state that accuracy is a fundamental virtue of a statistic. But at the same time, it is something that usually has to be taken for granted, because the ordinary user does not have the chance to replicate the process in order to check this aspect of quality. On this point the users are, as the authors say, “at the mercy of the statistical office.”

¹ Universidad Autónoma de Madrid, Madrid, Spain

On the other hand users can be, and usually are, very critical about the other four dimensions of quality. This attitude can sometimes be attributed to real deficiencies of the product. But in many cases a low degree of statistical literacy on the part of the users can account for most of the criticism. The consequence is that, although faced with the same problems, different NSOs will tend to follow different policies to solve them, according to the statistical culture of the country. And when this culture is comparatively limited, these policies will probably include some effective plans to improve it.

Ordinary users will certainly appreciate the quality of contents – and by that they understand the availability of information that will help them to make the right decisions and to monitor the results obtained with them. They will also appreciate timeliness, comparability through time and consistency of the different statistical sources. In addition to that, the delivery of information in a good, friendly format and accompanied by transparent methodological information is now increasingly valued.

In countries that do not enjoy a high degree of statistical literacy, a first problem to deal with is the fact that a number of users tend to identify statistics with the theoretical models of physical science. As a result, they have difficulties in understanding the limitations of the statistical production in relation to issues like the consistency of information coming from different sources. There is an important gap between their expectations and what can be actually offered to them. This is usually one of the main difficulties that a Statistical Office has to face when trying to improve credibility.

In this sense, any comparative studies performed and published in order not only to improve consistency but also to show what are the sources of the apparent inconsistencies and how some of them can easily be explained by differences in definitions or methodology are a very good investment. Unfortunately, for countries whose regulations do not accept the assignment of a unique number to every citizen, the need to preserve confidentiality may add some extra difficulties to this kind of analysis.

Then there is the problem of comparability over time. Users complain endlessly about this issue. It is not easy to get them to understand that if statistics are intended to picture real life, and this life is undergoing continuous change, this change has to be adequately reflected. Researchers, in particular, sometimes prefer statistics that are perfectly consistent over time, even at the cost of getting an increasingly obsolete view of facts.

Timeliness is a highly desirable virtue, but it very often affects the stability of the results published. Here the statistical office faces a problematic choice: whether to accelerate the publication at the risk of being forced to proceed to substantial revisions afterwards, or to delay it in order to provide, sometime later, more accurate information without revisions. There is not a unique solution to this dilemma. In countries where the media enjoy a high level of statistical literacy the statistical office can undertake reasonable revisions without great risk, because these revisions will be quite naturally accepted. But in other countries it can happen that comparatively small revisions jeopardize the credibility of the office.

On the other hand, users tend to have unlimited demands, and here comes the problem of estimations for small areas. Users that do not have a good statistical education may possibly insist on having an estimate for the area, irrespective of its accuracy, and can blame the statistical office for the refusal to provide it based on what they consider “flimsy” reasons.

And then I would like to mention another fact that can affect the image of quality of Statistical Offices: the growing use of statistics, especially at the microdata level, by

nonstatisticians. Although increasing the number and widening the scope of users is something positive, there is a risk of misuse of the data, including incorrect application of methods that can lead to crazy results. When these results are disseminated by the media, they will probably include a reference to the organization that produced the data, so that this organization will somehow share, in the public eye, the responsibility for these results. There is not much that can be done here, except perhaps to provide some sort of “instructions for use” manual together with the microdata file.

B. For an NSO, though, accuracy is a crucial issue. In fact, for most of them it is usually the main concern.

This means that estimating the total error of a survey is a basic objective. Sampling errors are easy to control, as reasonably complete theoretical models have been developed for them. But the situation is far more complicated for nonsampling errors. There is not a general theory that helps, and most of the available methods are expensive and very often not altogether satisfactory.

The first important thing for an NSO is to be able to detect and to have some idea of the magnitude of this type of errors. For a good part of the statistical production, the National Accounts are a good instrument, because they act as an internal checking method. It is also important that analysis of the data is carried out as much as possible within the office. Unfortunately, not all NSOs do this on a regular basis, either for scarcity of means or for lack of a suitable culture. Here international and supranational organizations can exert a good influence, by increasing their quality demands. For example, the IMF is already beginning to develop some checking devices for countries that are part of the Special Data Disseminating Standards group.

Then comes the question of how to deal with these errors. Although the problems involved are basically the same in every country, the solutions adopted usually vary, as a consequence of the differences in cultural characteristics. Still, the interchange, at the international level, of know-how and of experiences, both the successful and the unsuccessful ones, can be of great help to everybody. It has to be stressed that most methods available to deal with nonsampling errors are very expensive to apply.

Good communication within the Office is also necessary. Irrespective of the internal organization patterns, a reasonable degree of flexibility usually has a positive effect. The “project team” model, including the possibility of involving the users at some stages of the process, which has been adopted by many NSOs, seems to be working very well.

Although accuracy is a fundamental component of quality, it can be useful to NSOs to remember that it is not the only one. This is particularly true when it comes to editing. Manual editing is becoming obsolete, because of the cost in terms of time and money that it implies, and also because of the risk of introducing biases due to different interpretations of editing rules by the editors. There is usually here a choice between a complete, perfect editing at the cost of a delay in the publication, or a quicker release of slightly more imperfect data. In some countries two or three inconsistencies in the microdata, detected and published by the researchers, can substantially damage the image of the office. Once again, it is a question of improving the statistical literacy of researchers, by getting them to understand that a good balance between accuracy, timeliness and cost has to be achieved in order to provide a good service to the different types of users.

Imputation is another point subject to some controversy. Should the NSO impute missing data, or should this task be left to the users? In general, users tend to prefer to be provided with complete tables. When working with microdata, they also generally show a preference for data that have been imputed by the office. Still, it is important to make clear which of the data provided have been produced through an imputation process, and to describe as completely as possible the method that has been used. This will allow the users to carry out their own imputation if they want to use a different method, and will also give them useful information on the real degrees of freedom of the data provided.

It is not only in connection with the imputation that methodological information should be given to the users. Most of them, and particularly researchers and the media, tend to be highly appreciative of any explanation of definitions or methodologies accompanying the figures, and one should be given to them as often as possible. It prevents erroneous interpretations of the data by helping the users to understand the precise meaning of them, and adds to the image of transparency of the office.

Dealing with nonsampling errors is usually expensive. And the fact that it is the image of the office which mostly counts for the users is an additional problem because, as has already been mentioned, aspects like consistency or timeliness have a higher visibility than accuracy. That may result in the reluctance of governments to provide extra funds for the improvement of accuracy, as many users do not show a high sensitivity towards this particular issue.

C. We have to accept that there is no such thing as a complete model for evaluating total errors in surveys, and that there is not much hope for one, at least not in the short term. As the authors say, we are far from realizing the integrated goal of measuring simultaneously all important survey errors within one model.

This should not come as a surprise. As we move from the theoretical to the real world, modelling becomes more and more difficult. Measuring the error of an ideal sample is a comparatively easy task, because ideal samples fit well into a theoretical model. But when we come to real samples, those that are affected by events such as incompleteness of the frame, or nonresponse, we face the problem of modelling something as difficult to control as everyday life.

So, for the time being, and presumably for a long time to come, we will have to content ourselves with “a collection of practices, backed by some theory and empirical evaluation.” But this is not only the case of statisticians. As the authors say, this happens also in other professional fields, such as law and accountancy — and, I would add, medicine. Are we, statisticians, perhaps committing the same error as our users — that is, expecting too much from statistics?

Still, even if a complete modelling of survey errors does not seem to be within reach, there is much to be done for the improvement of the statistical production. The authors mention, among other things, the increasing importance of an accuracy assessment for statistics coming from administrative registers. On this point, a handbook of good practices regarding how to organize, develop and operate a register for statistical purposes could be a great advance.

Received December 2000