Journal of Official Statistics, Vol. 17, No. 1, 2001, pp. 77–86

Comment

Photis Nanopoulos[1]

1. Introduction

The article I have been asked to discuss has the title ‘‘Can a statistician deliver?’’ and deals with the question whether it is possible to create a global theoretical framework for the production of official statistics, capable of associating to each figure a quality value in terms of a ‘‘total error.’’ In the presentation of their arguments the authors analyze the whole process of production of official statistics and have a critical view of the role of various types of ‘‘statisticians’’ regarding quality and error measure.

The article’s greatest merit is the launching of a global discussion about official statistics. This is the right moment to launch such a discussion as statistical agencies are facing important administrative and professional challenges due to several phenomena external and internal to their administrations.

The most important consideration is that the 21st century has started with two extraordinary challenges for official statistics, the Information Society and Globalization.

The ‘‘Information Society’’ constitutes simultaneously a new operational and a new conceptual framework for official statistics. It calls for changes in all the operating modes of the process of production and dissemination of official statistics. It creates a completely new administrative environment inside the agency and for its communication with the external world. It creates a completely new observational field with the e-economy and the Internet society.

Simultaneously we enter the ‘‘Globalization’’ era, due to the creation of the World Trade Organization and the liberalization of movements of goods, services and capitals, sustained by the Information Society, which is creating a completely new economic and social environment with new demands for statistics and for quality in the services offered.

National statistical agencies have to observe phenomena that are losing their national character. New operational and organizational modes of enterprise across country boundaries and important uncontrolled migration and telework blur the picture of labour force, production, investment and other socio-economic factors, making less and less clear the nature of a given country’s economic and social evolution. From the statistical point of view, it changes the conceptual framework as it affects basic concepts such as ‘‘units’’ and ‘‘population’’ and the definitions of many variables with the adjective ‘‘national’’ in their name, and also observed populations (companies, people, etc.).

In their internal sphere, statistical agencies have to move to a completely new operating

[1] EUROSTAT, 2920 Luxembourg, Luxemburg.

environment concerning all their activities. A change of culture is needed in the administrations in order to respond to increasing demands for calling for more efficiency, less response burden, fewer surveys and more use of existing administrative sources. At the same time all administrations from the central government to the local ones are becoming data collectors and data base managers, bringing them very close to becoming statistics producers, and entering into competition with the designated national official statistical agencies. Facing the customers' new requirements they have to produce more, faster, and better statistics and provide a high quality service, with fewer resources.

2. Main Issues Addressed in the Article

As said before, the central theme discussed in the article is the need for a global theoretical framework for official statistics that will provide precise error computations, offering all possible scientific guarantees to the users in support of the confidence they have in statistical data. The authors analyze the complementarily between ''design'' and ''process'' (which we prefer to call ''methodology'' and ''production'' in Eurostat) and produce a ''typology of statisticians'' handling all the operations in the ''design-process'' workflow.

The article launches an important discussion on a series of fundamental problems in official statistics, more precisely the trilogy of ''quality in official statistics,'' the ''theoretical scientific basis of official statistics'' tightening them together by the ''total quality'' wire.

They seem to identify as the main reasons why ''statisticians cannot deliver'' the following:

- the absence of a holistic, comprehensive, total survey theory,
- the absence of co-operation among the various types of statisticians, compartmentalized according to their various domains of activities, and
- the absence of a clear demand for ''total survey error.''

2.1. *The total theory versus Current Best Methods (CBM)*

I can understand the vision of the authors with regard to having a general theory for official statistics and the utility of the computation of ''a total survey error'' decomposed in an additive way to a series of components corresponding to the various operational phases. The user would be perfectly aware of the accuracy of the data and could use the decomposition in his or her decision making process and the elaboration of his or her belief. The agency manager would find it easy to define priorities and allocate resources once he or she had the total error and its components. The line managers and statisticians would be able to benchmark and compare their performance and take the necessary measures to improve it.

One of several strong positions taken by the authors in this article is their complaint about the lack of a holistic scientific base in official statistics. They consider the absence of a total theory for official statistics as the main reason for poor handling of quality matters and the regrettable fact that statisticians of all types do not coordinate their views about quality and neglect several quality aspects, mainly in the ''process'' phase.

What is theory? The textbooks on the philosophy of science say that theory is a hypothesis that has been repeatedly confirmed empirically, though in principle it remains refutable. In other words, it is a scientific generalization. I do not wish to pursue the epistemological question whether that generalization originated with an *a priori* hypothesis or was the result of implicative induction. The practical function of theory is to guide action. I think that there are two relevant questions for us:

(1) if a CBM works, what is the **value** added by a superimposed theory? Take the example on page 9 of the article. The authors state: ''*A neglected area, from a* **theory** point of view, has been the edit operation...'' The authors then proceed, ambivalently, to outline some of the practical achievements and constraints in the field of editing. Ronald Reagan had said: ''Economists are people who see something work in practice and wonder if it would work in theory.'' Is there a message there for our profession?

(2) if a scientific generalization is, at least in some cases, the outcome of an inductive process, why do we not examine the ''*collection of practices*,'' to put it in the disparaging terms of the article (p. 2), and then construct a theory? There are at least three possible explanations:

- The different instances are not sufficiently **similar** to generate patterns and warrant generalizations. For example, no two surveys are identical, let alone three or four surveys, especially when it comes to the larger number of events required as a firm enough basis for generalizations.
- We do not fully understand the underlying **mechanisms** of statistical phenomena, for example the mechanisms that cause nonresponse in particular surveys.
- The required **meta-analysis** does not yet command the confidence of statisticians as a statistical activity.

It does look as if we may have to be content with limited theories concerning specific components. The situation is analogous to that in the social sciences, where there are middle-range theories in specific domains, such as migration (the gravity model) or crime (the pathological theory or the structuralism theory), but no comprehensive social theory. Reliance on these intermediate theories results, happily, in pragmatic empiricism.

2.2. The total survey error model

The complexity of a statistical survey is a major obstacle to the modelling and computation of total error. This is particularly evident in the European Union situation with 15 different national systems congruent to the realization of a ''European survey.''

There is no indisputably best model for survey and thus I cannot follow the authors when they state on page 3, ''*we (the authors) use the term survey in a wide sense to include sample surveys, censuses, and statistics derived from administrative registers.*'' This approach is not only unnecessary but also counterproductive. Countries such as Denmark with highly integrated register systems would want to see a different taxonomy of errors from countries such as the U.K., with heavier dependence on sample surveys and censuses. I think the authors implicitly recognized this because, although they define surveys in this broader sense, it seems that the whole edifice of their thesis in the article was constructed on sample surveys.

The situation of the European Union needs a model of many levels: ''conceptual,'' ''European,'' ''national,'' and ''respondents.''

We must clearly discuss this as perhaps our main issue. To me the questions seem to be: *What is that model*?; *Who wants it and why*?; *Is it feasible*?; *If feasible, who is to do it?*

What is it? This seems to be very simply, the aggregation of the errors that occur at the different stages of the statistical production process. These errors are wider than the statistician's notion of uncertainty and probabilities. The true unknown errors are not necessarily cumulative but can be (partially) compensating.

Who wants it? If the aggregation is feasible, it can be used to attach confidence intervals to estimates of the output or parameter θ (or, more in a Bayesian sense, estimates of the probability that θ is included in a specified interval). Even if that is not a formal statistical measure, but a **credibility confidence interval**, some users might find it reassuring to be told that the random interval, say $X+\lambda$ and $X - .\lambda.$, has, for instance, a 90 percent probability of containing the parameter θ. Confidence intervals used to be fashionable in GDP estimates but they are less widespread now.

There are two weaknesses in the concept of this aggregate error as used in the article:

- Incompleteness
- Internal inconsistency and complexity

Concerning the first, the authors said that they were concerned with the total **survey** error. However, that is an incomplete measure of the total error in many analytic situations. Missing are:

- **Model specification** error, normally the biggest single source of error. If we want to know why performances vary between schools, we have to construct (a usually hierarchical) **model** that is a fair representation of reality (pupil intake, school catchment area, teachers, etc.), then estimate the parameters (from available data or from collected new data).
- **Doubts about what to measure:** users often cannot tell us exactly what they mean, for example concerning social exclusion or quality of life. What are we supposed to measure? This issue is one of methodology, and its inadequate specification in terms of concepts, universe, etc., is another major source of error.

One question then is: if we cannot measure all the errors, is it worth going only half-way?

Concerning the second weakness, the aggregate mixes up relatively clear figures about sampling error with ''guesstimates'' of systematic errors, and it is, moreover, complex and user-unfriendly. It might even be unnecessary because when it comes to taking remedial action we have to do that at the level of **individual** component, so why go to the trouble of having a global figure?

Is it feasible? There are two issues here:

- Finding an appropriate statistical framework for evaluating feasibility.
- Empirically estimating probabilities.

The ideal framework should allow the identification and separation of all the components involved, then the summation of their uncertainties, remembering that combinations in statistics are very difficult. As we do not know the exact errors (if we did we would correct them!), the Bayes framework, in which the probability calculus is used to manipulate uncertainties, seems appropriate. It is also appropriate if feasibility assessment is approached through users' utility functions within a quantitative decision framework. The statistician will then have the role of getting the client to articulate its preferences in the form of a utility function, a task that takes the statistician beyond mere inference. Such a role is very difficult, bearing in mind, as stated above, that there are instances when users do not even know what they want.

Perhaps the most difficult area is to assess probabilities. What is the probability that my programmer will make a mess of his or her input program – remembering that the programmer is an important link in the statistical production chain. What is the probability that the survey manager will allow the timetable to lapse because he or she has been insufficiently vigilant? Any probability estimates are likely to be highly subjective and crude expressions of uncertainty and error. Then one has to combine them with the standard sampling variation and amalgamate them again across 15 countries, as said above. These practical difficulties would tend to suggest to many experts that the search for a total survey error (sampling and nonsampling) is an unrealizable utopian dream.

Who is to do it? Multidisciplinary, inter-institutional teamwork is required. Here the official statistician could very well be a customer of academic enterprise.

2.3. The role of statisticians

In answering the question ''who is the statistician'' the authors give an extensive and complete account of the various job descriptions that can be found in a statistical agency, despite the fact that the definition they give, ''The statistician in our title is anyone who contributes to the ultimate delivery of statistics and data to users'' is too broad as it includes also the data providers and the providers of resources.

I have to disagree with the arbitrary choice of the authors to explicitly regard anyone connected with the production of data as a statistician. Thus the informatician writing the input data programs and the manager responsible for adherence to predetermined timetables are statisticians. While Quality is a ubiquitous concern arising at every stage of the statistical production process, from methodology right up to dissemination, I do not think that the removal of the traditional demarcation between different disciplines is very helpful. Statistics issues are primarily, though not exclusively, about uncertainty, where uncertainty is measured by probability.

What is obvious in their discussion is that all those people try to work without ''a precise order to deliver a precise product to a given customer,'' and hence what we can expect is that everyone will give his or her own definition to quality and apply his or her own standards.

Here lies the problem with quality in official statistics: there is now ''user-defined quality standard'' and producers do not collectively define a unique and coherent standard.

In order to solve this problem, one way is to deliver to the user all information needed by him or her to determine *a posteriori* the quality of the data. The article discusses this issue in Paragraph 6, suggesting what already happens in the area of **''metadata.''** For more than a decade the word metadata has been on the agenda of many statistical meetings. EUROSTAT has already financed extensive research, starting with an international conference on this topic in 1993. Currently metadata is a well-established subject among statistical agencies, coordinated by the METIS group of the UN, the IMF, the OECD, and EUROSTAT, which has been very active.

The upcoming delivery of metadata, undertaken systematically by most national and international agencies, is also a way of handling the user's inability to spell out his or her requirements concerning data quality. Consequently the design of a survey or of the whole system of data collection in a statistical agency will aim at doing the best within the existing limitations, which mainly are resource constraints, respondents' cooperation problems, and theory limitation constraints. In this context every ''statistician,'' as described by the authors in Paragraph 2, has to analyze the constraints and determine a balanced quality framework, using the available CBM, and at the same time deliver the appropriate metadata.

2.4. The role of users

An important issue considered in the article is the relation between statisticians and users of official statistics. Statisticians have to know very well the users' needs and provide a quality service. What is the user expecting from the statistician?

Quoting the authors: *''The user asks: Can I trust the data and the statistics delivered to me? To what extent do they serve my purpose? The user seeks quality assurance. Is the statistician capable of delivering not only numbers but also adequate quality assurance? What forms should the assurance take? We believe that many feel, with us, that the statistical agencies (and the statisticians who work there) do not provide sufficiently explicit answers to these questions.''*

I can only agree with the authors regarding the fact, that users are asking those natural questions and that the statisticians cannot provide fully satisfactory answers to them. The point is that before trying to find answers to those questions we should examine the adequacy of their users' formulation. Looking to the first question of the user ''*Can I trust the data delivered to me?*'' we all agree that the statistician cannot deliver a direct and clear answer. Depending on the user's statistical culture, the statistician will have to use a scientific terminology and explain that ''trust'' in statistics is expressed in terms of ''probability distributions,'' ''bias,'' ''variance,'' and ''confidence levels,'' a language that many users in fact do not understand. A ''yes or no'' answer can be given only if the user has been able to previously determine his or her own quality standards. This is almost never the case, and by extension the retort can be made, ''*Can a user order*?''

In my opinion, most of the users want to ''trust the institution producing the data,'' and this should not be confounded with ''trust the data,'' which is why I find important the discussion in Paragraph 6 of the article, where the authors question and refute the adequacy of ISO standards approaches for handling quality in official statistics.

The users ask also for ''quality assurance.'' This means that the statistical agency is not simply doing its best, but that they define a quality standard and then take all measures to keep this standard. The question again is who defines the standard, who provides the means to guarantee the standard and who controls.

From the discussion in the two previous paragraphs we can raise the question; ‘‘*Is an ISO-type approach appropriate for handling quality in official statistics*?’’

ISO 8402 advocates that product quality should be defined by the users: the existence of different users implies different approaches to quality in official statistics. Theoretically the statistician should expect from the users the definition of their quality standards. Unfortunately most of the users are not at all prepared for providing explicit standards for the quality of the data they are looking for. This leaves the statistician alone to make the decision without knowing the parameters and the consequences of the errors the data may contain. The situation in statistics is not much different from that in industry. Quality is not defined by the users, it is something provided to the user at a certain price and then the user makes the final choice, combining his or her perception of quality with the price that has to be paid.

3. The Total Quality Approach Rejected?

3.1. Total quality versus total error

The challenge, stated above, faced by official statistics can only be handled through a continuous adjustment of our culture and organization to the new demands of our operating environment. This implies firstly the change of our business culture, moving from the culture of a ‘‘statistical data producer’’ to that of a ‘‘statistical services provider;’’ secondly our maintenance of a high level of technological capacities, being pioneers in mastering IT in the whole process of production and dissemination of statistics and statistical services; and thirdly the constant maintenance at the highest possible level of the staff’s skills and professionalism. In order to meet these objectives, several statistical agencies have chosen to follow the track of ‘‘Total Quality Management’’ (TQM). This is the track that we ourselves have followed at EUROSTAT.

TQM is a large and comprehensive framework for reflection, inside which the discussed article under discussion should be classified and be considered as launching an initiative for structuring a general debate that takes place in many instances in an unsystematic way. EU NSIs organized in 1998 a DGINS conference on TQM in statistical agencies, the Conference of European Statisticians is frequently debating on it, researchers are often presenting papers on quality, but we lack at present a systematic and holistic debate and this is something that the article seems to initiate.

Although their philosophical approach is holistic, the authors seem to complain about the total quality approach as been being too much user-oriented, with negative consequences for giving priority to building total theory and total quality error.

It is true that the methods advised by the Total Quality gurus and the standards such as *EFQM* (European Framework for Quality Management) are based on subjective measurements, such as the ‘‘satisfaction of the users,’’ which *a priori* will not give absolute priority to the measurement of the ‘‘total error.’’

‘‘Quality’’ is defined in ISO 8402 – 1986 as: ‘‘the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs.’’ In fact the way statisticians have been handling ‘‘quality of the products’’ questions covers all aspects that are relevant to quality assessment in a way that is understandable to the users. The methodology developed within the European Statistical System, concerning this topic, is similar to what Statistics Sweden and Statistics Canada have been proposing incorporating the need for comparability of the outputs of our member states’ statistical systems. Thus the ‘‘quality reports’’ produced in all statistical domains address seven quality elements of statistical data: 1) Relevance, 2) Accuracy, 3) Timeliness, 4) Coherence, 5) Completeness, 6) Comparability, and 7) Accessibility and Clarity.

I think it would be disastrous to deliberately push ahead with the accuracy measurement at the expense of the other dimensions.

3.2. The current context of official statistics

Nevertheless the question that remains to be answered is ‘‘How far can we follow the principle of satisfying users’ needs, and at the same time stay independent and scientifically objective?’’ In this direction the Total Quality approach raises a series of important philosophical and deontological questions in the community of official statisticians.

Official statistics is an international system based upon principles accepted worldwide and upon methods elaborated under extensive collaborative work between national and international statistical agencies. The core conceptual and methodological elements of this system are adopted at United Nations level, because statistics is a universal language and its output has to benefit international acceptance and comparability. Following the United Nations Statistical Committee: ‘‘*the primary and permanent role of a national statistical system is the collection, the treatment and the publication at regular intervals of a complete set of data providing the necessary quantitative information for understanding the economic and social structures of the country, the forces acting within these structures and the interrelationships between these.*’’

This definition of the role of National Statistical System has been complemented with a resolution on the principles that should govern the production and dissemination of official statistics. It reads as follows:

‘‘The Economic Commission for Europe,

Bearing in mind that official statistical information is an essential basis for development in the economic, demographic, social and environmental fields and for mutual knowledge and trade among the States and peoples of the region,

Bearing in mind that the essential trust of the public in official statistical information depends to a large extent on respect for the fundamental values and principles which are the basis of any democratic society which seeks to understand itself and to respect the rights of its members,

Bearing in mind that the quality of official statistics, and thus the quality of the information available to the Government, the economy and the public depends largely on the cooperation of citizens, enterprises and other respondents in providing appropriate data needed for necessary statistical compilations;

a) Official statistics provide an indispensable element in the information system of a *democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.*

b) To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

c) To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

d) The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

e) Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents.''

Are these principles more compatible with the authors' views rather than with the Total Quality recommendations?

4. Final Remarks

Right through I had difficulty with the presentation of the arguments, in four respects:

- There is a constant appeal for ''integrated theory,'' without any explanation of what the authors mean by theory, without much illustration of how that integrated theory will help a practical statistician or a user, and without any suggestion as to how such an integrated theory could be developed. For example, there is a tendency to be dismissive of CBM because it might not have been underpinned by any explicit theory. However, a CBM is a CBM and many less developed NSIs would be very grateful for it even in the present state of the art. We should not make the best an enemy of the good.
- There is a thread of **ambivalence** running right through the article. Each time the authors castigate statisticians, they quickly follow up by recognizing the valuable progress achieved, as if there is a wish not to offend while offending!
- There are too many sweeping statements that are simply not generally true on this side of the Atlantic. Examples are:

  - ☐ Page 8: ''*When the survey statistician talks of accuracy, he or she often refers to just one single error, the sampling error...*'' Not generally true in the EU.
  - ☐ Page 10: ''*The statisticians are silent on this issue (the effect of nonresponse on the MSE).''* Not generally true in the EU.
  - ☐ Page 11: ''*for the statistician, imputation is a quick fix, always ready to be used, as if nonresponse is nothing to worry about; imputation enters the collective mind as a saving grace*.'' This is a wild exaggeration.

- ☐ Page 12: ‘‘*(coverage errors in frames) are seldom or never considered.*’’ Again, very much off the mark.

Of course, there is room for improvement, but I am bound to say that statisticians must make accurate observations based on representative samples!

Received December 2000