Comment

Ivan P. Fellegi[1]

1. Introduction

I have the highest regard for the authors (Platek and Särndal) and for their attempt to synthesize our current understanding of the role of statistical theory in the practice of official statistics. Furthermore, I do not disagree with their conclusions as far as they go. However, I do find that they take a narrow view: they seem to equate ''delivering'' (presumably with respect to the basic functions of official statisticians) with quality; furthermore, though they start from a broad view of quality, most of their discussion focuses just on accuracy. Hence this discussion deals not with what the authors say, but with what they have left out: the other dimensions which must be considered before answering the question about whether we are ''delivering.''

2. Delivering on What?

The article raises a question in its title: ''Can a statistician deliver?'' Without intending to quibble, I think it is necessary to try to clarify how one might interpret this broad title. As far as the word ''statistician'' is concerned, the authors leave no doubt about their own intention: by statistician they mean ''official statisticians,'' and indeed they employ a broad interpretation of the term to include the variety of professionals who work in national statistical offices. However, the term ''deliver'' is left undefined. If it is intended to mean ''deliver on some implied or explicit promise,'' then the next question is: promise regarding what? From the point of view of society as a whole, the most relevant interpretation would be to refer to an implicit social contract between an official statistical agency and its clients to deliver a statistical information service that provides the maximum benefit within the available budgetary resources. Of course, ''maximum benefit'' also needs to be defined, and I will come back to that later.

By contrast, and without discussing what they mean by ''deliver,'' the authors seem to equate ''delivery'' with quality. They do, admittedly, propose a broad concept of quality which encompasses: (1) Contents, (2) Accuracy, (3) Timeliness, (4) Coherence (especially comparability), and (5) Availability and Clarity. These are not precisely the headings that Statistics Canada identified in its Quality Assurance Framework, and I actually prefer our headings,[2] but this is a point of secondary importance.

[1] Statistics Canada, Ottawa, Ontario, Canada
[2] The Statistics Canada Quality Assurance Framework defines the elements of quality as (1) Relevance, (2) Accuracy, (3) Timeliness, (4) Accessibility, (5) Interpretability, and (6) Coherence. The two descriptions of the attributes of quality are close, but I prefer ''Relevance'' instead of ''Contents,'' ''Accessibility'' in place of ''Availability,'' and the extra condition of ''Interpretability.'' Interpretability, among other attributes, includes the adequacy of the definitions of concepts, target populations, variables and terminology underlying the data, as well as the availability of accessible information on any relevant limitations of the data. These attributes are commonly referred to as good meta-data.

Even such a broadly defined concept of ''quality'' does not quite include all the attributes on which it is essential that government statisticians ''deliver.'' I will only mention two such attributes, though both of them are quite broad. I consider these as survival issues: they are so important that without excellent performance in these two dimensions the agency will not survive – and hence issues of quality become irrelevant. They are:

- respect for respondents, and
- credibility of information (of which accuracy is an important part, but only a part).

So, my first criticism of the article is that, by zeroing in on ''quality'' as the sole attribute which distinguishes statisticians who ''deliver'' (as opposed to those who do not), it does not highlight some ''survival'' issues without which official statisticians cannot deliver anything useful - certainly not for long.

My second criticism is that, in spite of its own broad definition of quality, the article implicitly skips over all but one aspect of quality – the exception being accuracy, to which close to two-thirds of the article is devoted. In what follows I will elaborate on both of these criticisms.

3.   Survival Issues Beyond Quality

3.1.   Respect for respondents

Respect for respondents manifests itself in many ways. I want to underline three particularly important dimensions of it: respect for privacy, safeguarding confidentiality, and managing reporting burden.

3.1.1.   Privacy

I am personally convinced that the single greatest strategic risk facing statistical agencies arises out of privacy issues. The nature and magnitude of this threat might vary from country to country, depending on national history and culture. However, official statistics deeply depends on access to personal information, either through surveys or through access to administrative records. We must have the confidence of the public that we will treat such information *well* within the margin of what good privacy practice demands. Unlike almost any other collector of personal information, we have very few *direct* incentives to offer, and practically no effective penalties. Indeed, banks, insurance companies, or many government institutions provide direct benefits to individuals who seek such services in return for providing whatever personal information might be requested. Yet others have very effective sanctions to ensure the flow of information: this is clearly the case for the tax authorities, but in a more subtitle manner the same applies to information requested by physicians. By contrast, statistical offices have to rely ultimately on the willing cooperation of the public. While such cooperation is undoubtedly based on their acquired legitimacy as providers of relevant and trusted information, the fact remains that the individual who is asked to provide personal information about himself or herself receives no *direct* benefit and is typically subject to no meaningful threat of sanction.

Key privacy issues on which official statisticians must ''deliver'' include:

- We must pay careful attention to limits of what might be regarded, at any point in time, as appropriate topics for official statistical inquiry. There is a clear but difficult

balance that must be attempted: too conservative an approach will limit relevance, too much daring might lead to public outrage. Such limits are not fixed over time. For example in Canada an inquiry in a national census about common law living arrangements, particularly of same sex couples, would have been unthinkable ten years ago - and might still be so in many countries; yet it is now part of the forthcoming Canadian census. Conversely, ethnic origin which did not use to be sensitive has become so during the last 10-15 years.

- We have to be extraordinarily transparent with respect to our policies and practices regarding record linkage. The essence of information privacy is having ultimate control over the use of information about oneself. Yet most of the time when large-scale record linkage is carried out, it is impossible or at least unfeasible to ask for the informed consent of the people involved. So most large-scale record linkage studies inevitably involve a degree of real privacy violation. However, it is also true that record linkage might provide information whose public good value is extraordinary – e.g., on the potentially carcinogenic effects of previous exposure to certain chemicals. Here, as in so many other areas, the issue boils down to balance: a judgement that must be made as to whether the potential public good resulting from record linkage outweighs (in some sense) the violation of privacy that is implied by record linkage. If one wants to avoid the potential of public outrage then, whatever the judgement, it is essential that there should be a rigorous and transparent review process, that the objective should clearly be statistical/research in nature, that there should be confidentiality safeguards in place, that as a group the subjects of record linkage should not likely be disadvantaged as a result of an authorized use of their information, that the linked files should be destroyed within specified time limits, and that the resulting information should be placed into the public domain.

3.1.2. Confidentiality, including security
The safeguarding of the confidentiality of individually identifiable reports has long been recognized as a survival issue for statistical offices so I will not elaborate. It certainly includes both appropriate physical security of completed statistical questionnaires (stored in whatever medium), as well as a careful screening of publicly available statistical information.

3.1.3. Managing reporting burden
I have intentionally used the term *manage* instead of *minimize* because I believe that the issue is more complex than the latter term might suggest. This is not the place to elaborate at length, but the following are definitely involved:

- Use of every reasonable tool to reduce the aggregate reporting burden, particularly in relation to small businesses (for whom the reporting effort might represent a non-negligible competitive burden). Possible tools include: use of administrative data to replace direct surveys, sampling in place of complete enumeration, abbreviated questionnaires, etc.
- Careful management of the distribution of reporting burden, including sample rotation (distribution of reporting burden over time).

- Regular measurement and public reporting on the aggregate reporting burden that is imposed by the statistical system.
- Negotiation with larger businesses (which tend to be in several surveys) of the most effective manner of receiving reports from them (e.g., consolidation of several surveys, electronic reporting).
- Clear communication for every survey of its purposes.

3.2. Credibility of information

Credibility is the second ''survival issue'' that I would like to highlight. Indeed, statistical information is a very special commodity: it can be rendered useless entirely on the basis of perceptions, even in the absence of other substantive reasons. Since users of statistical information are rarely in a position to directly assess or otherwise verify the information, they have little alternative but to rely on the reputation of the producer. If anything, this reliance has become even more unavoidable in the age of the Internet and the unfiltered access which it provides to an enormous range of information emanating form a huge variety of sources.

Credibility, therefore is a life and death question for statistical agencies since information that is not believed will not be used - in which case we are no longer viable. Thus credibility is an element of ''delivering'' which has an existential character.

How can such a reputation be acquired and maintained? A full exploration of the issue would take me beyond the scope of this discussion, but I will mention at least four characteristics of credibility: accuracy, transparency, nonpolitical objectivity, and (less directly) relevance.

Clearly, a reputation for credibility must be based first and foremost on a sustained record of producing information of substantial **accuracy**. The full range of issues and considerations raised in the article by Platek and Särndal is, therefore, entirely relevant. They correctly point to a genuine vulnerability due to the fact that only a relatively small part of our practice is directly guided by theory. The quality of our output, hence our credibility, would certainly be based on a stronger foundation if more of our practice could be justified directly on theoretical grounds. Hence it is very much in the self-interest of national statistical offices to collaborate with all relevant partners to bring this elusive objective somewhat closer to reality. But in the meantime we can still ''deliver,'' even if doing so in this respect takes particularly great care, constant awareness of the evolving best practice, maintaining an open mind in the face of experimentation, and so on. There are many fields, from medicine to astronomy, where theoretical guidance only provides some of the needed guidance, and yet where progress in applications is rapid. Even where theory provides no *direct* guidance, best practice often utilises a way of thinking about errors (and allocating resources to reduce them) that is grounded on the (incomplete) theory of total error.

Transparency in the reporting on data quality issues and on the methods and definitions used also contributes to credibility. It is essential that users of statistical information should know that they need not speculate about possible quality issues: they can have confidence in the statistical office to make a full disclosure of all significant issues that might help to interpret the reported information. The knowledge that such information has freely

and proactively been made available can become enormously useful when quality issues become the subject of public controversy.

Nonpolitical objectivity is another determinant of credibility. In effect, the data might be as accurate as need be, but if the user believes that the statistical office is likely to be politically biased (or even subject to possibly effective political pressure), the credibility of the information that is provided will be severely affected. After all, the statistical office might have been selective in its presentation of the data or commentary.

I would finally mention **relevance** as an (indirect) contributor of credibility: obviously, you have to be *noticed* in order to be able to create a reputation for anything, including credibility. In Statistics Canada, for example, we have created two classes of release: what we (officially) call a ''major release'' and ''other releases.'' Major releases are those to which we particularly want to call attention because of their intrinsic importance. ''Other releases'' refer to more specialized information which commands a smaller general interest. By so doing we signal to the media what we think they may want to feature - and hence reduce the risk to our reputation posed by an image of being the purveyors of a lot of information which most people might regard as irrelevant.

4. Quality of Statistical Service Versus Quality of Specific Statistics

My second set of comments about the article by Platek and Särndal concern the rather important distinction between the overall quality of a statistical service, as opposed to a specific statistical series. In effect, from the point of view of society, the main question is not whether or how to improve a particular statistic: that is always possible to do with more resources. But even the question of improving the quality of a given statistic within a given budget is still not the most pertinent question socially - interesting and important as it may be from a methodological point of view. After all, there are very few statistical offices which are absolutely unable to allocate some additional resources to the improvement of *one* particular statistic. The socially crucial question, therefore, concerns whether the budget of the statistical office is, in some sense, optimally used to deliver a statistical service for the benefit of the country concerned.

I am not, of course, suggesting that a rigorous solution exists for this broader question. But then no such solution exists typically even in the context of a single statistic. Yet in that case we have evolved a productive approach: we try to learn as much as can reasonably be done about each source of error (or, in the broader context, about each aspect of performance), then gradually make adjustments and keep reassessing. A similar approach can work very well in the broader context, substituting for sources of error the various different aspects of the performance of a statistical service. The question boils down to setting out what are the relevant dimensions of quality in a statistical service, how to monitor each, and how to effect adjustments.

4.1. Dimensions of quality and how to monitor them[3]

4.1.1. Relevance

I already mentioned relevance as a subsidiary aspect of credibility. It is, however, a

---

[3] This part of the present article draws on Fellegi, I.P. and Brackstone, G. (1999). Monitoring the Performance of National Statistical Offices (NSI). Statistical Journal of the United Nations ECE, 16, 251-265.

fundamental determinant of the quality of the information service provided by the statistical system. Like most other aspects of performance monitoring, the extent to which the statistical system is relevant cannot be summarized in a single *quantitative measure*. Rather, performance monitoring here involves putting in place comprehensive *processes* designed to provide explicit signals about the relevance of our product line. While the issues arising from such processes will typically raise questions of priority (an ultimately subjective matter), the assessment process itself can and should be both objective and comprehensive, and decisions about them should be made in a transparent manner.

The point to emphasize is the importance of the processes designed to monitor relevance. Acquiring knowledge about evolving client needs does not occur by osmosis. The following are examples of such mechanisms:[4]

- bilateral committees with major government departments which meet regularly;
- committees of senior officials dealing with subjects which are under provincial jurisdiction (where applicable);
- expert advisory committees on specific subjects (demography, labour, price measurement, globalization issues, etc.);
- a blue ribbon committee to provide guidance on broad priorities and other national issues;
- commissioned expert studies on how client satisfaction might be improved in specific information domains;
- market feedback (sales, number of Internet downloads by subject, etc.);
- feedback from internal or externally conducted analytic studies.

These processes yield a variety of signals which need to be integrated and displayed, and about which explicit judgements (unavoidably subjective) have to be made. These judgements, to the extent they involve the expenditure of additional resources, *necessarily involve a trade-off against alternative possible investments, e.g., to improve the accuracy of some existing statistical series*. How well these trade-offs are carried out depends on the existence and character of a good planning system within the statistical office.[5] But there can be little doubt that the quality of service provided by the statistical office (i.e., whether or not it is ''delivering'') is crucially affected both by the processes it uses to monitor the continued relevance of its product line and by the regular trade-offs it makes about marginal adjustments in the utilization of its resources.

4.1.2. Accuracy

I will be much briefer here since much of the article by Platek and Särndal deals with the issue very thoroughly. To reiterate some previous comments:

- clearly the quality of service of the statistical office crucially depends on the adequate accuracy (relevant to the main purposes) of its outputs;
- no doubt, we could all do a better job if theory provided us with more specific guidance.

---

[4] For additional detail, see Fellegi, I.P., Characteristics of an Effective Statistical System.
[5] Reference to Barnabé paper.

4.1.3. Timeliness

Timeliness as an element of quality requires little explanation: while information may not entirely lose its relevance even after a long period of time (in fact, like old cars, beyond a point it may start increasing its value), there is little doubt that ''current'' information becomes less relevant with the passage of time. But, as in the case of the trade-off between investing in product line relevance versus increasing the accuracy of some existing series, there are important trade-offs here as well. The primary trade-off here is with accuracy, but there is interaction with other performance criteria as well: investing in timeliness competes for resources which can equally well be invested to improve other aspects of performance.

No general criteria exist here. Indeed, there are significant differences among countries (and even between clients in the same country) in the relative weight attached to timeliness versus accuracy.

4.1.4. Accessibility

Given the continued and rapid evolution affecting dissemination media, accessibility has never been as important an element of performance as it is now. But, precisely as a result of rapid changes in technology, significant resources must be devoted to the maintenance of excellence in dissemination – one more competing need. The cost relates not only to the creation and maintenance of large Internet accessible databases. Indeed, significant resources are also needed to maintain traditional (paper) products, in spite of their diminishing sales, since a reasonably large proportion of the population still cannot access data electronically.

4.1.5. Interpretability

It is essential for the effective use of information that users of the statistics be clear about the target population measured, and the concepts, operational definitions and classifications that were used. However, it is increasingly clear just how difficult a task this is. Many countries are struggling to establish what have come to be known as meta-databases. Just as difficult is the maintenance of such a base, once established. Two points can be unambiguously asserted: good and accessible meta-data are significant aspects of the performance of statistical offices – i.e., of whether they ''deliver''; and the resource needs here have to compete with all of the other demands arising out of the overriding need to improve our service.

4.1.6. Coherence

Coherence is intrinsically an aspect of quality that refers to several, as opposed to a single, statistical series: it reflects the extent to which series which ought to adhere to a certain expected relationship deviate from such a behaviour. Unquestionably, the provision of coherent statistics is an important aspect of quality – of fitness to use. To illustrate, if users want to compare the output of a certain industry with its labour input, they expect that the industrial coverage of the two series should be identical. Otherwise the measured labour productivity becomes an artifact of inconsistent measurement.

Coherence is improved through the use of common classifications, common sampling frames (both of these would be needed in the case of the example above), common

conceptual frameworks (e.g., the systematic use of the concepts behind the System of National Accounts), and common definitions where appropriate (e.g., defining employment or unemployment consistently across surveys).

Maintaining and improving coherence is not cheap. Not only does it require quite expensive infrastructure (e.g., a register of businesses which is used system-wide, a capacity in classification, a strong analytic capacity), but it also entails periodic allocation of additional resources when conceptual frameworks or classification systems have to be changed in order to maintain *their* relevance in a rapidly evolving world. Once again, the needs here have to compete with other uses of the same resources.

5.   Conclusion

Platek and Särndal have raised a crucial question and provided us with a wise exploration of issues related to accuracy. However, this represents only one dimension of their broad question. They explicitly avoided answering their own question, even within the constraints of their exploration, but the tone of their conclusions is nevertheless somewhat pessimistic.

I have tried to indicate all dimensions of the issue of what we need to do if we want to ''deliver'' on our implicit contract with society. My insistence on the broader context is not just a debating point. I believe that there are other equally important determinants of performance, over and above considerations related to accuracy – crucial though the latter are. Furthermore, while I share the authors' regrets about theory not providing us with more direct guidance with respect to quality issues, I am much more optimistic in my answer to the broad question posed by the authors: we can, indeed, ''deliver'' if we pay full attention to the various determinants of the performance of a statistical office. However, this will not happen by itself: we must design explicit processes to monitor the different aspects of performance. We must also design a system that brings together the variety of explicit (but often descriptive) indicators of performance and which therefore facilitates priority setting.

We are setting ourselves up for failure if we equate performance with accuracy: within the resources of a statistical office, we can always allocate some additional resources in order to improve the accuracy of a single series. This will even be true if one considers the accuracy of all key series. While it is important that we should be assessed (and we might as well start with self-assessment), it is essential that the parameters of such an assessment are as broad as the range of issues on which we must ''deliver.''