Journal of Official Statistics, Vol. 17, No. 1, 2001, pp. 113–127

Rejoinder

Richard Platek[1] *and Carl-Erik Särndal*[2]

1. The Title

We wish to thank all the discussants for their extensive and penetrating comments on our article *Can a Statistician Deliver*? It is evident that careful reading and reflection lies behind most of the discussions. They force us into another close examination of a number of issues.

Some of the discussants seem to be intrigued by the title, even agitated about it, in particular about the word ''deliver'' followed by a question mark. A title is a name, a label, not a summary of the contents. The title *Hamlet* conveys by itself nothing of the drama contained in this Shakespearean play. Scientific or professional writing is not much different in this regard: the title can, at best, suggest a theme, as does ours.

Needless to say, we have deliberately chosen the word ''deliver'' and used it in the sense of ''fulfil a promise.'' In our opinion that promise is imbedded in a statistician's role and accountability vis-à-vis the public to provide quality information. This role may be viewed differently by different types of statistician.

Our text, which is entirely nontechnical, raises issues which will generate a variety of opinions among readers, as the 14 discussions show. Some discussants have taken the opportunity to reflect on their own views and convictions, from the standpoint of their professional role. We are pleased to have provided an opportunity for this. To make a synthesis of these discussions is a challenge. To keep this rejoinder concise, a number of interesting points in the discussions will unfortunately have to receive insufficient attention in the following.

The discussants represent a spectrum of important roles that statisticians engage in: leaders and professionals in statistical agencies or in private survey institutes, university-based experts in behavioural, statistical and information sciences. Some discussants react to our text from the standpoint of an affluent country, others from the perspective of more limited resources. It is not surprising that the discussants focus on different issues, and that, on one and the same issue, there are differences in opinion.

We note, however, a common denominator: None of the discussants seems to view the future of statistics production as a wide-open boulevard, free of challenges and difficulties. There is a feeling that far reaching changes now emerging around the world will have considerable effects on statistics production, especially in National Statistical Institutes

[1] 1125 Amble side, Ottawa, Ontario, K2B 8E2, Canada.
[2] 2115 Erinbrook Crescent no. 44, Ottawa, Ontario, K1B 4J5, Canada. E-mail: carl.sarndal@home.com
Acknowledgments: We acknowledge once more the contributions of 14 discussants, whose frank comments motivated us to formulate this rejoinder. We also acknowledge two other contributions. Marker and Morganstein's contribution arrived too late and Elvers and Nordberg's contribution did not reach us due to a mistake on the part of the Chief Editor. We regret that it has not been possible for us to consider their comments.

(NSI's). All agree that no unified theoretical framework exists as a basis for this production.

Our objective in writing this article was not to vent feelings of any kind, such as optimism or pessimism. We view our text as an objective statement, although coloured, obviously, by our own orientation towards statistical survey methodology. Some discussants perceive pessimism in regard to certain issues in our text. We were not the first to express such sentiments. If a term were to be used to describe the character of our article, we would prefer ''introspection,'' ''self-evaluation'' or even ''self-criticism,'' words which are evoked by other discussants. An article must not always attempt to show the infallibility of the author's own position. No shadow is cast on anyone's individual effort if it is suggested that collectively, as a community of researchers and professionals, we could perhaps have achieved more, in particular in answering certain questions of importance, answers which would perhaps have provided a better readiness for the onslaught of ''the information society,'' ''globalization'' and ''the new economy.''

2. The Statistician

We use a broad definition of ''statistician,'' common in the media and in professional circles: The term may refer to any one of the professional types that contribute to the production of the statistical agency: the statistical methodologist, the behavioural methodologist, the informatics specialist, the economist, the sociologist, the survey manager, and so on. (''Statisticians'' also include bio-statisticians, econometricians, mathematical/statistical theoreticians, and yet others. But if their work falls outside the sphere of surveys and statistics production, they are not addressed here.)

The title can be interpreted as *Can a statistical agency deliver?* There certainly exists a statistical agency point of view on what we say, and some of the discussions reflect this. Statisticians work inside the agency, in a number of different professional capacities with different orientations and professional ideals. They attend different scientific and professional meetings. Thus we have also envisaged that the title *Can a statistician deliver*? can address the individual statistician/specialist's capacity to live up to his or her own professional standards and ideals. His or her personal obligation of ''fulfilling a promise'' is perhaps different from the agency's striving to fulfill its promise to society and to users. The statistician/specialist may consider that he is delivering, by his own standards, while at the same time he may be less convinced that the NSI is delivering to the fullest extent. Several discussants view our text from the standpoint of a particular statistician/specialist.

3. The New Era

Our timing for presenting the article is also deliberate. The question mark in our title hints an uncertainty about the future. Several discussants do address future prospects. The world is subject to vast technological and social changes. Globalization of the economic sector creates an information revolution. Electronic data gathering and dissemination will play a dominating role. Colossal amounts of information are required and large quantities of data can be produced at a moment's notice, but it is not evident that these data can be turned into reliable, high quality information. Different geographical configurations, regions and

unions of countries, create new demands for data. Fields such as industry, economics, trade, health, education, and several others rely heavily on good data. Are their needs satisfied? Can the present organizational systems and methodology for statistics production meet such expectations? Set in its traditional values, the statistical agency is, as it were, puzzled by the vast possibilities of the new technological advances.

4. The NSI: Its Present and Future Role

The rest of our rejoinder focuses on three themes emphasized in the discussions: the scientific and professional credo of the NSI, which inspired our article in the first place (Section 4), the role of theory for the NSI's activity (Section 5), and the concept of Quality in regard to the NSI's activity (Section 6).

Essentially all of the discussants agree with us on a range of issues. Some discussants disagree with us on isolated points. Also, we incur some criticism because our coverage of certain issues was not extensive enough. We have reacted to 14 of the 16 discussants in the following way. The statements in this rejoinder represent our views and opinions. A certain statement of ours may be inspired by a theme or a thought raised by one or more of the discussants; their names appear within parentheses. They may agree or disagree with us on the issue in question. The reader can consult the discussant's own text to find his or her exact position. We seldom give exact quotes from a discussant. Proceeding in this way avoids a constant repetition of phrases such as ''we (partly) agree with . . .'' and ''we (partly) disagree with . . .'' In a majority of cases there is at least partial agreement.

Several discussants debate the NSI's prospects for a continued delivery of high quality statistical products. Fellegi's view reflects a position of the NSI as the hub of a country's centralized statistical production, an organization very much active and alive. We agree with him that quality can be seen, and is seen by many observers, as the quality of the whole entity, somewhat in the manner that an observer on the shore regards a passenger liner: it is the beauty of the ship that is important. Other discussants also examine the future role and prospects of the NSI (Madaleno, Martin-Guzman, Nanopoulos, and Trewin), and frequent use is made of words depicting a harmonious relationship between the public and the NSI: the public's *trust* and *confidence* in the NSI, the NSI's striving for *credibility* and favourable *reputation* in the eyes of the public. A vehicle used to attain and preserve these desirable conditions is quality assurance, based on a multidimensional Quality concept.

The NSI and the public relate to each other in two important ways. The NSI has, like the Roman deity Janus, two faces looking in opposite directions: It looks at the public (i) as a *provider* of personal and other information for the agency's many statistical programs, (ii) as a *user* or a *customer* for the resulting statistical output. The discussants reflect on both of these faces of Janus.

The first aspect, the NSI facing the public as the provider of input data, encourages Fellegi to raise the question of ''survival beyond quality.'' In the future, can the NSI count on adequate co-operation, high rates and high quality of response? Unconditional respect for the provider of input data is a key to preserving such co-operation. Although non-response has been on the rise, one has the impression that even today relations with respondents are very good in many countries. It has been getting tougher, but it could

be much worse. Nevertheless, nonresponse in excess of 20 percent, common today even in highly respected surveys, was ''unthinkable'' 30 years ago. How flexible will future statisticians be in their tolerance of questionable quality of data input, while still preserving their professional conviction that they produce good statistical output? When would ''getting worse'' be tantamount to ''catastrophe?'' The issue lies ''beyond quality'' in that none of the components of the carefully thought out Quality concept measures the temperature of ''respondent relations.'' The prestige of an NSI is difficult to create and maintain, easy to destroy (Martin-Guzman; Trewin).

It is to be noted that some aspects of respect for the public as respondents were implicitly covered by our Section 4, because such respect lies behind a careful preparation for survey operations in the field. Inadequate training of interviewers, poorly designed questionnaires, overly sensitive questions, ambiguous definitions may not only create measurement errors and nonresponse, but also confuse and antagonize respondents to the point where the agency's credibility will suffer.

Issues ultimately linked to survival are likely to occupy more of the agency's attention in the future. Understanding and accommodating the respondents will draw more resources. The behavioural sciences and the management sciences will have a bearing on the optimization of this effort. Improved managing of the response burden (Fellegi) will be one outgrowth of it. More generally, the NSI will significantly increase its skills in regard to the procedures which improve the chances that a high quality data set will ultimately be at hand for the compilation and estimation phases. Such procedures tend to come at high cost. Using the two aspects of research on nonresponse as an illustration, attempts at reducing it before it occurs is relatively expensive, compensating for it once it has occurred (by reweighting and/or imputation) is relatively inexpensive (and effective only if powerful auxiliary information is at hand). Both aspects of research will continue to be important; a hardened climate vis-à-vis respondents may necessitate an increased concentration on the former.

The second face, the NSI looking at the public as a user and consumer of statistical products, touches on a second aspect of survival: Will the NSI be able to compete in an ''electronic future'' (Nordbotten; Nanopoulos), when the NSI and existing private survey institutes will perhaps face competition from many new statistical service outlets, that is, firms which gather and process input data and sell the resulting statistical product (Nordbotten)? How successful and competitive will a future NSI be as the ''seller'' of processed data? Will the NSI deliver superior ''value added,'' compared to a small-scale competitor, private or governmental?

In regard to the users, the NSI's face questions such as: (a) who are the future users and how does the NSI identify them? (ii) which of the diverse user types should the NSI focus on? It may become increasingly difficult to identify all customers (Nordbotten; Stanley McCarthy); vast numbers will be coming from the Internet, seeking access to essentially free data.

Users' sophistication varies. Users have very different conceptions of what Quality is (Stanley McCarthy). Some appreciate and understand well the statistical qualities relating to Accuracy (Bailar; Holt). But many less discriminating users seek accessible and timely information and pay little attention to Accuracy; they trust ''official statistics'' and take Accuracy for granted (Martin-Guzman; Madaleno). Yet when asked, almost every user

will say that Accuracy is a fundamental virtue (Martin-Guzman). As the total number of users increases, the proportion of less sophisticated users may increase, but the number of sophisticated users may also increase.

It is not evident which of the components of Quality that the NSI should spend relatively more of their future resources on, in order to maintain reputation and trust. In the past the users were relatively few and well known to the agency, but in the future, Relevance *to a majority* of a large number of users can perhaps not be guaranteed (Nordbotten).

The Eurostat perspective (Nanopoulos), presented as a view from ''this (European) side of the Atlantic,'' contrasts with well established survey practices in North America. Nanopoulos correctly hints, though, that much of the current modus operandi of statistics production is conditioned by the word ''national,'' the objective of the NSI being to serve the nation state. How the supranational perspective, fuzzy at the moment, will affect survey methodology appears to us now as an adventure of the mind; it remains to be seen how it will complement or supersede the more traditional views of the NSI's of the nation states of Europe.

The discussions relate in various ways to *trust* and *confidence* in the NSI, to *credibility* and *reputation* of the NSI. The public's feeling of trust and confidence in the NSI relates to two issues, at least: (i) the public needs to trust that privacy and confidentiality will be scrupulously and unconditionally respected; (ii) the users need to trust that the statistics produced are ''correct within reasonable limits,'' that is, highly accurate. The issues (i) and (ii) are related: without respect for the public, there will not be high quality response, there will not be accurate statistical output. It is usually apparent to the producer (the NSI) whether or not it enjoys this respect.

None of the discussants, nor we, have a clear vision of how vital the component Accuracy really is in relation to trust and confidence. Accuracy is not open to easy inspection by the public. The public at large rarely see what Accuracy is about; they get a glimpse of it in those very rare cases where a gross error is revealed in some statistic and perhaps blown up disproportionately by the media. A majority of users trust the NSI for Accuracy. Some sophisticated users will ask more detailed questions about it (Bailar) and the statisticians, time permitting, will provide more ample explanations. The average user with some interest in Accuracy may try to form his or her own opinion from more or less informative metadata.

A comparison of the discussions hints at the differences existing between countries because of the size and the resources available to the NSI. It would carry us too far here to analyze these differences, which can have important consequences for Quality.

A central yet diversified NSI benefits from a vigorous mix of expertise in different critical fields: statistical science, economics, sociology, demography, and others. This is our own preferred view of an NSI. A diverse competence will remain a shield of protection for the NSI. Its survival could be threatened, for example, by a wave of privatization of sectors of the total national statistics production. But there is room for optimism. In countries where the status and nature of the NSI has been the subject of debate, there have been strong voices both in the media and in professional circles in support of a highly competent, independent NSI, and this shows awareness of such an NSI's unique contribution, and its obligations, to a democratic society.

5. Theories for the Production of Statistics: Accomplishments, Shortcomings, and Future Prospects

5.1. Generalities

The role of some statisticians is to deliver guidance by the application of theory to assist the activity of the NSI. Several discussants dwell on this aspect. Theory is an elusive word, used occasionally in an ''undisciplined'' way about possibly trivial matters. Whatever theory is, in the current context of statistics production, it is definitely not identical to mathematics. We emphatically deny that the characterization ''survey methodology as a set of practices with no unifying theory covering the entire process'' should in our minds be so limited as to mean an absence of unifying *mathematical* theory (Bailar). Mathematics alone is incapable of breeding and embracing such a unifying theory. We purposely mentioned music and literature as examples of fields far from the hard sciences, yet endowed with nontrivial, essentially nonmathematical theory or theories. (Bailar adds the example of Chemistry.) We took a rather critical stand on myopic mathematical examination of survey methodology issues of lesser importance.

Some survey statisticians frown on the mathematical aspects of survey science. This is not surprising because the field spans several disciplines. However, some of the criticism of formal, mathematical development is unfair. Nothing has a more persuasive power than a forceful mathematical argument on a significant methodological issue. The history of survey sampling literature gives ample evidence of this.

Although not as closely, we, too, had the privilege of knowing Morris Hansen and some of his collaborators. To hear William Madow, a member of that group, talk about the progress in this era, the 1940's and 1950's, made it abundantly clear that what they experienced at the time was the enchantment and the wonder of scientific discovery. They held the Accuracy banner high; Hansen was not one to condone unbridled reliance on models of questionable validity. For example, he is known, in addition to other achievements, for his rejection of model based inference and consequently for rejection of an excessive reliance on imputation.

Although Quality is seen as multi dimensional, we still find it justified to devote considerable space, in this rejoinder as well, to one of its components, Accuracy, especially since several discussants dwell on it. Indeed, many future users, in particular the more sophisticated ones, may demand to be more explicitly informed about Accuracy (Nordbotten). If the electronic future brings undisciplined dissemination of data, some of very poor quality, there will perhaps be more users capable of identifying good quality data and of discriminating between data producers (Stokes).

However strong, theory does not eliminate professional judgment in regard to the steps of a survey (Holt). Theory is a guide. The aim of the scientific process is to reduce everything to a comparatively small number of fundamental ideas, derived by selecting and abstracting from the total real-life experience. The set of fundamental ideas is crystallized through the efforts of researchers and scientists. Once such a set of ideas has proved its merit and has become commonly accepted, we tend to call it professional judgment when it is applied by practitioners in their daily work. When they are aware of and accept the guidance of these ideas in their work, theory has fulfilled one of its roles.

5.2. Theory for total survey error: the old, unfruitful attempt

In a striking comment, Biemer notes that we seem to be implying that the statistician ''has not delivered much in the past 40 years,'' but he immediately adds ''we must admit they are correct,'' at least in regard to the slow progress in developing the Total Survey Error (TSE) modelling concept of Hansen and his collaborators. One of the concepts behind that model, the idea of TSE as a sum of error components, has both helped us (by focusing research and development on specific sources of errors) and disappointed us (in not providing a routine measurement of individual errors and of the total error).

We noted in the article that the statistician, viewed now as a representative of a modern statistical agency, has realized progress in those years because information technology has revolutionized the collection and treatment of information and the transfer of information to users. Advances have been made in other areas, such as behavioural survey methodology, based on behavioural and cognitive science theory (Bailar; Groves and Mathiowetz; Biemer; Stanley McCarthy). Advances have been realized in statistical survey methodology, based on statistical science and mathematics (Desrosières et al.; Holt; Pahkinen). In Section 5.3 we come back to these accomplishments, which target specific survey operations.

The lack of progress on the TSE model and on the estimation of its components was regretted long before this article, but it may not have found such a concerted expression as in these discussions, involving something of a ''death sentence'' on this modelling attempt (Biemer; Holt; Groves and Mathiowetz; Martin-Guzman; Nanopoulos; Pahkinen). To realize routine measurement of systematic and variable errors by means of that theory is seen as an ''unattainable and unrealistic'' ideal (Biemer), as a ''utopian project'' (Pahkinen), an ''unrealistic utopian dream'' (Nanopoulos). ''Unpredictable people are involved in so many points of this incredibly complex process'' (Stanley McCarthy), so how can we expect to measure TSE?

The theory builder should not commit the mistake of trying to fit reality into an inadequate, too narrow framework of fundamental ideas. A reproach that may be levied in hindsight against the TSE theory of Hansen et al. is that it is too narrow. (This is no criticism of Hansen et al., whose laudable objective at the time was to extend theory beyond sampling error.) The set of fundamental ideas of the TSE theory, its skeleton, is: a sequence of survey operations, an error for each operation, total error as a sum of such errors, the concept of probability, the probable error of each operation, the resulting probable total error. The crucial concepts of ''probability'' and ''probable error'' are supposed to hold the theory together, but it does not work. There is nothing wrong with these two concepts in themselves, but here they do not carry the argument forward to the desired goal. Persons trained mainly in statistical science are therefore particularly disappointed.

It clarifies this discussion if we distinguish *specific theory* and *comprehensive theory*. The former helps to understand and develop practices for one or a couple of related survey operations; ''local theory'' would be another descriptive name. Comprehensive theory takes a broader view of the whole process of statistics production, more particularly of the Accuracy of the process. A more comprehensive theory is what we regret not having, at least some of us. Specific theory can be very helpful, but only for limited purposes.

Several discussants believe that we should be content with the progress made with

specific theories on separate operations, since ''we are getting better at it every year'' (Bailar; Groves and Mathiowetz) and since it may be the pragmatic approach (Nanopoulos). Consequently, we should accept the status quo of statistics production as ''a collection of practices supported by some theory'' here and there. This attitude has, in our opinion, a smack of defeatism.

We have no qualms about saying that the theoretical basis for official statistics production is weak. Several discussants agree with us. The theory for sampling error is not weak. That of questionnaire design is not weak. We are talking about the lack of a more satisfying broader framework. The argument is made (Bailar) that other fields exist that draw on several traditional sciences of which statistics is one, meteorology being one example. It is implied that official statistics production fares no worse than these fields, so statisticians should be happy.

Statistics production does involve other aspects than the statistical one: there are legal aspects, considerations of rights of respondents, privacy, confidentiality, standards and policies for classification and for information dissemination, national accounting principles, and so on. However, central for statistics production is Accuracy (in the wide sense of the word), and *not even* this *statistical aspect* of official statistics has a solid, coherent theory. Is the statistical information accurate and reliable enough for the user, or is it not? Admittedly, a more focused phrase on our part would have been ''the theoretical basis for *Accuracy* of official statistics production is weak.'' Is it possible in the future to improve in this regard? In Section 5.4 we consider possibilities of a ''new vision.'' But first, Section 5.3 discusses specific theories.

5.3. Specific theories

We must consider two outlooks on the effort to understand Accuracy: (i) The first is tied to the interpretation of Accuracy in the standard numerical, quantitative way, as the ''deviation from the truth,'' $\hat{Y} - Y$, equal to the sum of the errors attributed to the several steps in a sequence of survey operations, and where the magnitude of each individual error is also of interest. (It is not Accuracy in a direct computational sense, because we cannot compute the deviation from the truth, only try to estimate it.) This outlook inspires statisticians to seek the best possible estimates, made from the ultimately recorded data with their various imperfections, including estimates with built-in features for nonresponse adjustment (reweighting and/or imputation), and so on. (ii) The second focuses on conditions that will facilitate or pave the way for numerical Accuracy. It strives to reduce the chances for imperfections in the set of data that will ultimately be recorded and used for estimation. This outlook motivates the search for good techniques for questionnaire design, for eliciting response, more generally for identifying factors that help understand respondents, interviewers and the interview process, so that these factors can be incorporated into the design. Efficient methods for sample selection are of interest to both types of thinking. The respect for the respondents mentioned by several discussants comes to bear in particular on (ii). (Our notation (i) and (ii) is slightly inappropriate in that it goes against the chronological order of the survey.)

The two paths tend to be pursued by two different types of researchers. The distinction between *behavioural survey methodologist* and *statistical survey methodologist* was

recently given a sharper profile than it may have had before (Dillman 2000). The former group works predominantly although not exclusively under the outlook (ii) on Accuracy; the latter group works mainly under the outlook (i), using probabilistic (mathematical) arguments and techniques.

The grouping is illustrated by the partition of researchers on nonresponse into ''reducers'' and ''adjusters,'' a distinction that cannot fail to be noticed by participants at the yearly international workshops on nonresponse issues. The former group, which counts many behavioural survey methodologists, tries to understand the causes for nonresponse and measurement error, and develop methods to prevent nonresponse from occurring. The latter group, mainly statistical survey methodologists, works with probabilistic (mathematical) tools to invent techniques to reduce bias and other undesirable effects of nonresponse in any estimates that are made with the data after nonresponse has occurred. For example, they explore different uses of auxiliary information with this goal in mind. The two groups may politely listen to each other's presentations, but the scientific intricacies on either side, flowing from different specific theories, easily reach a point where a full appreciation of the contributions of the other side is lacking. Both efforts are relevant, but they mirror two scientific solitudes in one and the same field.

Some discussants underline, for good reasons, the desirability of a stronger focus, in the future, on the reduction of nonsampling errors (Martin-Guzman). The relative lack, especially in the statistics departments, of university courses dealing with nonsampling errors is regrettable (Biemer; Nordbotten; Stokes). The discussion, in the classroom or elsewhere, of various types of bias in surveys is perhaps not mathematically elegant but is nevertheless very important (Bailar; Stokes), and it requires rare teacher talent.

There are advantages to viewing progress on Accuracy as the joint result of outlooks (i) and (ii). For example, some may believe that advances in regard to the nonresponse problem are a result of outlook (ii) only; this is not the case. There are advances also from outlook (i), set in motion by the idea of response mechanisms and of individual response probabilities attached to the different units, as discussed by Holt and Desrosières et al., and dealt with in the often cited book *Model Assisted Survey Sampling*.

The dichotomy into sampling error and nonsampling error, although an enormously popular terminology, appears archaic to us. Just a simple count - a survey has one sampling error and maybe 25 different nonsampling errors - shows the imbalance of this terminology. A more significant distinction in regard to work on Accuracy is in our opinion defined by (i) and (ii), that is, between methods *before* data collection and methods *after* data collection. Roughly, the distinction corresponds to the one between ''design'' and ''process'' in our article.

To the outlooks (i) and (ii) correspond, today, two blocks of specific theory, leading, respectively, to behavioural survey methodology and statistical survey methodology. Our article mentioned that progress has been realized in recent times in the study of several error components. We are pleased to recognize more clearly the contributions stemming from outlook (ii) and based on the behavioural and cognitive sciences (Bailar; Biemer; Groves and Mathiowetz; Stanley McCarthy). The examples given of such progress, and the references provided by Groves and Mathiowetz, are valuable additions to this discussion.

We wish to underline the contrast, noted in a few of the discussions, between the research traditions behind the two methodologies (Biemer; Groves and Mathiowetz;

Desrosières et al.). To the statistical survey methodologists, the error models are off-shoots of variance component models: the error model just postulates a number of possible variance components, and it usually has no parameterization for seeking out the *causes* behind the components. By contrast, behavioural theories try much more explicitly to identify the causes of survey errors. An illustrative example of the difference is the one of Groves and Mathiowetz involving the intraclass correlation and the interviewer workload.

The two traditions are different. In statistical science, an innovative formulation of a variance components model forms the nucleus; such a model (called a theory) is considered successful if the researcher arrives first at expressing, then at assessing the magnitude of the (perhaps numerous) terms (variances, interactions, and squared bias terms). In a practical application, some of these variance components may be insignificant, indicating no cause-effect attributable to the factor in question. But this insignificance does not worry the statistical science researcher; he or she is not concerned with any real world situation in particular. By contrast, in behavioural science, the tradition is to discover a new causal relationship among variables of interest, and to formulate an expression (a theory) for this relationship. Such a theory is deemed successful if it is capable of explaining other (but similar) situations. If the theory is not corroborated by finding the same cause-effect relation in similar real world situations, there is reason for the behavioural scientist to be dissatisfied with the theory. As Groves and Mathiowetz point out, many of the behavioral theories are not yet translated into models of statistical error; this should become part of the future program.

We would like to emphasize an indisputable presence on the scientific scene. We are referring to the fact that survey sampling, a very practical field, gave birth, several decades ago, to what is today a respected academic field of mathematical/statistical study. That this field was *born* out of practical survey sampling and official statistics is worth underlining; it is no small achievement. The movement is not present in all universities, but is lively in some. (By contrast, survey questionnaire design, another essential element of statistics production, is not a child of survey sampling; psychological measurement was developed independently.) It is not a theory for statistics production – there is none – but a mathematically framed sampling and estimation theory, stripped of much of the ''nuisance'' of real world surveys. Consequently, some call it unrealistic, impractical or useless. It is given its structure by a few fundamental ideas; it leads to interesting (at least in a mathematical sense) generalizations, which is a requirement of a good theory. The insight that the identifiable units of a finite population can be given different inclusion probabilities produced *design based survey inference*, in the period from 1934 until the end of the 1960's. From the 1970's and on, it was widely recognized that a model can be used as a catalyst for inference (leading to *model assisted design based inference*) or as a base in itself (leading to *model based inference*), or as a happy marriage of the two. Many participants in this movement are pure sampling error specialists; a reader of some of their papers may wonder if they have ever heard about survey nonresponse, for example. But the contribution is nevertheless stimulating to many, and some of it does have repercussions on practice.

Specific theory is also instrumental in recent progress in regard to some other survey operations. One example is Edit and Imputation (Nordbotten), where neural networks

theory may become important in the future. Other progress has come from Economics and Geography, for example; this has not been emphasized in the discussion around our article.

5.4. Comprehensive theory: a new vision

Some discussants appear to be satisfied with, or see little alternative to, the present situation, that is, the status quo in which the only way that survey methodology unfolds is with the aid of specific theories applied to isolated parts of the production process (Bailar; Stanley McCarthy; Nanopoulos; Trewin). However, others signal the desirability of a broader, more satisfactory theoretical framework for the statistics production process (Biemer; Holt; Nordbotten), and at least one (Biemer) is taking a step we could perhaps expect to see as a result of our article, namely proposing a new vision for research on Accuracy (in the wide sense of Section 5.3). Biemer asks how we as a profession (of survey statisticians of all categories) can *begin* to achieve a new vision, a new research tradition. He notes that the routine reporting of nonsampling error components is not a desirable future goal, even if it could be realized in practice. Even reporting of sampling error is viewed as sometimes not very meaningful (Trewin). The reporting of various meta data, including rates for different types of nonresponse, imputation rates, statement of imputation methods, and so on, is also not by itself a satisfactory future objective, nor do error profiles appear to satisfy the perceived need. These activities achieve one thing, namely a public declaration of the agency's commitment to quality (Holt), but perhaps little else.

Information scientists have produced overviews of the production process based on data flows, from the early stages of a survey to the publication of results. These flow charts tend to view the whole process from ''far above'' without much guidance for thought on the many important sub-components of the flow.

The aim of science is to extend the range of solved empirical and practical problems. A research tradition is a formulation of a set of important concepts in a field of study, and of methods suitable for making progress in the field. The formulation has to convince a majority of the participants, researchers and practitioners, as something that works now and holds future promise for fruitful extensions and generalizations.

For the case that we are debating, accuracy of statistics production, what might this set of fundamental concepts look like? It will probably contain new concepts as well as some familiar ones. It will combine concepts from several traditional sciences. Its strength and ability to convince will lie in the combination of concepts. The discussion around our article suggests that such a future survey science will borrow in important ways from the behavioural sciences, since survey participation and response touches various unpredictable aspects of human behaviour.

The new framework will consider both outlooks on Accuracy, the outlook *before* data acquisition and the one *after* data acquisition. It will respect the fact (Desrosières et al.) that for sampling error we have an approach *independent of the survey*, valid for all surveys. But except for sampling error, a general approach (one that is applicable for studying all the parameters of a survey operation and *valid for all surveys*) is out of range. For errors other than the sampling error, a general system must be replaced with particular systems, created perhaps by comparing across countries, surveys or some other reference set.

Our article invites readers to reflect on new frameworks for thought on survey errors and Accuracy. Biemer describes the objective as ''raising the bar.'' We do need something ''higher,'' something less constraining, than the skeleton of the old TSE theory. If a convincing and fruitful new framework emerges – it does not happen overnight – then researchers will flock to it.

Important elements in Biemer's proposal for a new research tradition for survey errors seem to us to be: (a) for allocation of survey resources: study-integration approaches (rather than approaches specific to a survey); and (b) for the reduction of errors of specific operations, to identify major sources of error by reference to a critical (national) survey, or to the (international) survey literature. As he concedes, it is a sketch; much work would remain to be done, but the suggestion holds promise.

6. The Quality Concept

Other prominent themes in our article were the Quality concept for statistics production, and how to achieve and measure Quality. The current Quality definition (or definitions) consists in a number of main components (the number depends on to whom we are listening: Statistics Sweden and The Australian Bureau of Statistics 5, Statistics Canada 6, Eurostat 7). Each main component has sub-components. Accuracy is always one of the main components. Clearly, to find the ''proper'' definition of Quality is a never-ending search. Also, the term Quality has quite a different signification if by it we mean Quality of the whole statistical service (the NSI), rather than Quality of specific statistics (Fellegi).

Aside from an unavoidable arbitrariness in the definition, it is clear that the concept has provided some challenge for thought in the last 20 years. Some discussants express belief in formal quality management, TQM or other such approach (Stanley McCarthy; Madaleno; Nanopoulos; Trewin). Madaleno perceives a gradual change of focus in some NSI's, in which the notion of ''error'' is gradually replaced by the notion of ''quality,'' and ''correction of errors'' is replaced by ''eliminating sources of error and preventing their occurrence.'' (Complete elimination of nonresponse and measurement error being impossible, techniques to apply ''after the fact'' are still necessary.) The amorphous Quality concept differs from a typical scientific structure, in that it does not lend itself to extensions and generalizations in the manner of a fruitful theory.

Other discussants have important things to say about Quality without referring to TQM or some such approach (Holt; Pahkinen; Stokes). The comments of Pahkinen and Stokes reflect an uneasiness that we believe is felt by many representatives of statistical science who are both university based and well informed about official statistics production. To them, a disturbing aspect is that users of data have limited control over and insight into the procedures used by the NSI in the production of data. The sophisticated users, at least, will ask: Can we trust the data for the analyses and research we want to do? High quality in the sense of ''customer satisfaction'' does not necessarily imply high quality of data for analysis and research.

The NSI enjoys, in some countries and in certain respects, a monopolistic position (Pahkinen; Nordbotten). It sets to a large extent its own production norms. This is in contrast to the pharmaceutical industry, for example, where stricter normative

international standards exist (Pahkinen), which illustrates that statistical thought for fields other than official statistics has acquired a more mature structure. It is suggested that too dominating a position of an NSI creates an almost harmful self-reliance. It can prevent the statisticians in the agency from clearly and objectively evaluating Quality, including Accuracy. Thus there is a need for an external Quality evaluation system (Pahkinen), or a user-defined Quality standard (Nanopoulos), although it is recognized that some objective criticism flows from various scientific advisory bodies, composed of experts from outside the NSI and now a tool widely used by NSI's.

The financial and other resources that are available to an NSI varies between countries. We need to consider this aspect here, because the discussants represent countries that differ considerably in regard to available resources. The NSI's of some not-so-large yet technically advanced countries have limited room only in their budget for absorbing the high cost of keeping many high level experts on regular payroll. However, users of data in these countries are no less sophisticated or demanding in regard to data quality. A fruitful avenue for some not-so-large countries has been the recent creation of continuous networks for scientific cooperation between the universities and the NSI (Pahkinen). The benefits accruing to the NSI are both objective criticism from the outside and help with methodology development.

In the discussions, we perceive a difference between a more scientific point of view and a point of view anchored in a quality management approach. In a TQM view of Quality, ''customer satisfaction is the main concern'' (Madaleno). Catering to highly visible aspects of Quality, such as Timeliness and Accessibility, can become a temptation for the NSI (Martin-Guzman). The future of the NSI, according to Nanopoulos, will entail ''more, faster and better statistics produced with fewer resources.'' Agreed, ''more,'' ''faster'' and ''with fewer resources'' are quality aspects contained in ''better product.'' But whether other important aspects of ''better product'' also follow from this expedient outlook is highly questionable.

Our own point of view lies closer to the scientific point of view. Customer satisfaction, however defined, is not unimportant, but many will agree with us that statistical quality (Accuracy, in the sense of Section 5.3) has to be a main future objective. ''Every user will say that Accuracy is important'' (Martin-Guzman).

We end by addressing three points on Quality raised in several of the discussions:

- There is a need for a stronger framework for Quality assessment (Holt). Despite all that is being said about Quality, a ''systematic debate'' about the concept is still missing (Nanopoulos). Different users attribute different importance to the 6 or so components of Quality (Nordbotten). Some experiences and problems in regard to the reporting of components such as Comparability, Consistency, Timeliness are expressed by Martin-Guzman. Instruments for monitoring the main Quality dimensions and the sub-dimensions have been developed in some agencies. Andersson, Lindström, and Polfeldt (1999) give a systematic list of measures and indicators suitable for Statistics Sweden. The vast majority of the suggested measures relate to the sub-components of Accuracy, a strikingly small number to all sub-components of Content, Timeliness and Accessibility together. Making routine use of such a measurement protocol requires resources; it adds to the workload of survey

managers, already hard pressed by production deadlines, to also regularly provide quantified measurement on a large number of Quality-related items.

- Current Best Methods (CBM) manuals are used as instruments for standardization of processes and to promote Quality more generally. Similar objectives are realized by Best Practices manuals (Bailar) or Procedures manuals (Madaleno) or Benchmarking (Trewin). A CBM is usually limited to practices for one or a couple of related survey operations, supported by specific theory (in the sense of Section 5.2) and empirical evidence. Few CBMs seem to have been written concerning the components Coherence and Accessibility; practice may be less firmly established for these than for Accuracy. The point is made (Desrosières et al.) that for a given operation, ''the best'' may be too expensive, and the combination of what is ''best'' for sampling, for reducing measurement errors, for nonresponse prevention, for nonresponse treatment may, for cost reasons, be an impossibility. One may add that the ''B'' for ''best'' in CBM usually means ''best for a particular environment:'' a recommended principle for producing a CBM on a given topic for a given workplace is to account for the assembled experiences of professionals in that environment. The ''B'' in CBM differs from the more unconditional bestness usually sought by a researcher in survey methodology. Nevertheless, a CBM manual that becomes widely used within a given local environment can lead to important gains due to standardization of procedures.
- Some discussants comment on the Quality component Relevance in interesting ways, indicating how Relevance is, in some sense, inseparable from Accuracy (Fellegi; Holt; Desrosières et al.). To qualify as information, statistics must be more than accurate, they must also be relevant. Fellegi stresses that the delivery of information that is of benefit to the country is contingent on receiving appropriate signals about the Relevance of the product line. The processes used by the NSI to get these signals (inputs from government, research communities, expert boards outside the agency, and so on) are of the greatest importance. Holt notes that statistics produced by administrative sources may not always measure the concepts we would rather like to have measured, so Relevance is, in that light, more important than Accuracy. Desrosières et al. sketch the idea that Relevance is always limited by the more or less fuzzy character of any statistical concept with regard to which the statisticians set out to deliver information. Relevance, mainly in the hands of the subject matter specialists, and Accuracy, mainly in the hands of the methodologists, are thus intertwined. There is much more to ''finding the truth'' than just finding an estimate, $\hat{Y}$, that is in some statistical sense near the unknown parameter value, Y. The idea and its consequences deserve to be further explored.

7. Conclusion

This rejoinder contains two important parts: a discussion of theory (Section 5) and a discussion of the Quality concept (Section 6). This is for good reasons. Although we did not foresee this at the time of writing the article, a partitioning emerges from a reading of the 14 contributions: there is a management point of view, and a scientific point of view. The division is not clear-cut; some discussants may fall somewhere in between. The

management objective is to ''keep the ship sailing'' as smoothly and efficiently as possible. The multidimensional Quality concept, and in some cases a formal TQM approach, helps in this regard. Statistics production must go on. The other is the scientific point of view, the desire for scientific rigour and advancement, which pervades several of the discussions. It is a shared ideal, regardless of whether the spokesperson represents behavioural, statistical or information science. Therefore, at one extreme we may have persons who say that better (unifying, more comprehensive) theory is not needed. But if such theory were forthcoming and succeeded in resolving a number of the outstanding problems, the same persons would be hailing this as a breakthrough. It is a characteristic of the human condition that that which we do not behold or cannot fancy, we do not need.

8. Additional references

Andersson, C., Lindström, H.L., and Polfeldt, T. (1999). Att mäta statistikens kvalitet (Measuring the Quality of Statistics). Statistics Sweden, R&D Report 1999:3. (In Swedish.)

Dillman, D.A. (2000). The Role of Behavioral Survey Methodologists in National Statistical Agencies. (With discussion.) International Statistical Review, 68, 200–220.

Received February 2001