



Advisory Scientific Board
Suad Elezović, U/ARK
Tiina Orusild, PCA/MIH

Meeting with the Advisory Scientific Board of Statistics Sweden April 16-17, 2015

Board members

Stefan Lundgren, Statistics Sweden
Lilli Japac, Statistics Sweden, chair
Tiina Orusild, Statistics Sweden, secretary
Suad Elezović, Statistics Sweden, secretary
Professor Jan Björnstad, Statistics Norway
Professor Sune Karlsson, Örebro University
Professor Xavier de Luna, Umeå University
Professor Thomas Laitila, Statistics Sweden and Örebro University
Professor Daniel Thorburn, Stockholm University
Professor Geert Loosveldt, Catholic University of Leuven, Belgium,
Professor Julia Lane, Wagner School, New York University, USA ,
Professor Natalie Shlomo, University of Manchester, UK
Professor Edith de Leeuwe, University of Utrecht
Professor Lars Lyberg, Stockholm University

Other attendees

Annika Fröberg, Statistics Sweden
Eva Bolin, Statistics Sweden
Mats Bergdahl, Statistics Sweden
Birgitta Mannfelt, Statistics Sweden
Frank Weideskog, Statistics Sweden
Jennie Bergman, Statistics Sweden
Sofia Nilsson, Statistics Sweden
Can Tongur, Statistics Sweden
Martin Axelson, Statistics Sweden
Gustaf Strandell, Statistics Sweden
Pär Karlsson, Statistics Sweden
Peter Lundquist, Statistics Sweden
Åke Pettersson, Statistics Sweden
Johan Eklund, Statistics Sweden
Sara Westling, Statistics Sweden
Anton Johansson, Statistics Sweden
Fredrik Jonsson, Statistics Sweden
Frida Videll, Statistics Sweden

Day 1

Current issues at Statistics Sweden

Speaker: Stefan Lundgren, Director General

Director General Stefan Lundgren informed the board about the new Deputy Director General Ms Helen Stoye, the new Head of National Accounts Department and the other current issues at SCB:

- On-going process of Peer-review
- Quality Auditing process
- Remarks concerning Bengt Westerberg's investigation:
 - Should SCB inform other agencies about the tools and methods
 - Coordinating role of Statistics Sweden
- Declining response rates

Reply to recommendations

Speaker: Lilli Japac, Head of the Department for Research and Development

Lilli mentioned that recommendations from the Board have been well received by Statistics Sweden.

Regarding SIMSTAT, the work carried out on applying a sampling design on intra-EU trade is a response to the recommendations. Statistics Sweden will continue working on different methodological challenges if SIMSTAT is to be implemented.

Concerning the topic "Disclosure" Statistics Sweden will consider suggested alternative methods and will test and compare them to the methods suggested by the Board. In addition, the shortcomings of the ABS method have been investigated and the results will be available after the summer. A follow-up discussion by presenting the results at a future Board meeting will then be feasible, if there is an interest from the Board.

The recommendations about "Editing" have also been appreciated and the group will continue to work according to recommendations.

Feedback: Mixed-mode in the party preference survey

Speaker: Tiina Orusild and Johan Eklund

Tiina and Johan presented the work that has been done since the meeting in May 2014. Main goal: Higher response rates and cheaper data collection

A question has been asked: Why does Statistics Sweden do this survey at all?

Lars answered: This is a contribution to democratic society.

Sune asked a question about estimators: What estimators are about to be used and why?

Tiina's answer: Evaluate the auxiliary variables used in the estimator and also evaluate the use of mixed-mode. To change the estimator is a big issue.

Topic 1: Indicators for data collection

Speaker: Martin Axelson

Summary of presentation

Data collection is an integral part in the production of official statistics – without data, no statistics.

Data collection is typically very costly

- Total annual turnover for Stats Sweden: ~1 000 000 000 SEK.
- Total annual turnover for the two data collection departments

The problem is accentuated by decreasing response rates

- For Stats Sweden, this is mainly related to data collection from individuals and households.
- The Swedish LFS is a telling example.
- Once again the Swedish LFS can serve as an example. Rough estimates indicate that in a given month:
 - Interviewers spend less than 20 % of their work-time conducting interviews.

The project

- Two-fold aim:
 - To develop indicators that can be used to monitor and control/steer/guide/manage the data collection process for the Labour Force Survey (LFS).
 - To suggest a new and more cost-effective data collection strategy for the LFS.
- Budget: equivalent to 1200 hours
- Work: input from DIH, PCA and U
- Delivery:
 - A set of indicators to be used daily, to follow up the work done the previous day but also to have a basis for deciding any corrections to the collection work to be performed that day. Continuous analysis of the proposed indicators will provide relevant information on which to base future decisions on a new data collection strategy.
 - Examples!
 - In case of collection strategy, the project team state that a functioning strategy
 - must be explicit and well documented
 - should include production targets

The approach proposed project should be seen as a first step towards a finer approach - there may be reasons to first introduce these proposals and make them work in practice before further changes are introduced.

- Suggestion 1: Divide sampling units should into “treatment groups” after information on prior participation in the LFS
 - Five groups:
 - Units included for the first time
 - Units that participated last time (i.e., three months ago)
 - Units that refused to participate last time
 - Units that weren’t reached (no-contacts) last time

- Others (small, but probably very different group)
- Suggestion 2: Focus on result by reference week rather than reference month
 - Each reference month corresponds to four/five reference weeks
 - The monthly sample is split by design into four/five subsamples
 - Subsamples are allocated to reference weeks
- Suggestion 3: For a number of key indicators, agree on targets against which ongoing data collection is to be monitored and evaluated

Questions to the Board

- Assuming that we impose the restriction that we must deal with reality as it is, in terms of availability to (process) data, are there any obvious gaps/flaws in the work done so far? Are the results fit-for-use?
- Considering that the aim is to monitor, evaluate and eventually manage a complex, large-scale production process, what other methods/models, statistical as well as others, are worth exploring? Are the council aware of successful implementation/use of such methods/models elsewhere within the field of large scale statistical production? (If yes, are there any obvious lessons to be learnt?)
- Any tools/methods/models implemented, will in the end be used primarily by staff with little or no formal training in statistical theory. Is this likely to be a big problem? If yes, what types of short-term measures can be taken to overcome the problem?
- On a more general note, according to the council, what types of skills/education/?? are vital for enabling large-scale data collection, which is both cost-effective and under control (both productivity-wise and quality-wise)?

Discussant: Lars Lyberg

Lars summarized the history behind the topic. His comment on the paper is that it is kind of puzzling- unsure whether the paper is about nonresponse or about the cost or about developing a new standardized process or something else. He also criticized the incompleteness of the paper.

He commented on issues concerning the data collection process:

- The current process must be described and problems identified.
- A new and hopefully improved process is defined.
- Process variables are defined.
- Paradata are collected and analyzed.
- Action if necessary.
- Statistical process control and continuous improvement.
- Stop beating around the bush
- Gaps include lack of descriptions of current process and problems and how project results should be used.
- The literature is very rich and applications of responsive design are many.
- Quite a few organizations worth benchmarking (University of Michigan, the U.S. Census Bureau, RTI, GESIS, ONS, Statistics Canada.

- If the problem is so big why are improvement resources so tiny?

Comments on questions

- Stop beating around the bush.
- Gaps include lack of descriptions of current process and problems and how project results should be used.
- The literature is very rich and applications of responsive design are many.
- Quite a few organizations worth benchmarking (University of Michigan, the U.S. Census Bureau, RTI, GESIS, ONS, Statistics Canada).
- If the problem is so big why are improvement resources so tiny?
- Yes, lack of training can be a problem. Several examples of initiatives that did not work because of that. Important that dashboard people are well-trained.
- A new standard process should be priority no. 1. Start small when it comes to paradata and indicators and then gradually expand.
- Methodologists must update themselves in several areas (statistical process control, paradata, responsive design, total survey error framework)

Discussion

Julia Lane asked a question about necessity of all questions in surveys and also a question about potential possibility to simulate instead of using all those surveys. Martin commented that the paper was actually not so much about reducing non-response but more about increasing effectiveness of that we already do.

Jan Björnstad commented that there are many ways to correct for nonresponse bias but what we do in estimation phase is important. He added that in the future it will be unavoidable to reduce the cost of data collection. Reducing the cost and doing the things differently will be inevitable in the future, such as in the case with mixed-mode example.

Geert commented that one approach could be to condition on some other characteristics and not only on respondent characteristics.

Martin commented that Statistics Sweden have applied indicators for different groupings.

Geert also asked if there is a possibility to ask a question about next participation. Martin answered that Statistics Sweden already does that but the results are not as good as expected.

Daniel commented that the paper summarizes lot of good ideas but the outcome is unclear. Perhaps some follow-up studies would be pertinent. Maybe control studies as well. He also pointed that we should not bother about nonresponse if we have good background data. Hence, the important issue could be how to use good background data in an efficient way.

Martin's general comment was that the remarks and suggestions are more related to the Labour Force Survey than to the process perspective. The issue here is about spending money to collect data we do not get, so money might be used for other purposes.

Sune commented that calibration is one possible way out in which case the problem would be about modeling. Fundamentally, we do not know why people choose not to participate no matter what we do.

Some other comments:

- Statistic Sweden's procedures should not be model based (in this case).
- Using data from other sources, e.g. big data etc.

Topic 2: Process data and contact attempts

Speaker: Gustaf Strandell

Summary of presentation

Non-contact is the most common cause of non-response in SCB:s individual and household surveys conducted by telephone. It is therefore important to have at hand well defined indicators through which the process of seeking contact can be analyzed and hopefully also improved. In this report we have tested a collection of process indicators which relates the inflow of data in a survey to the conducted contact trials.

Questions to the Board

The main purpose of this report has been to give a few examples on how process data regarding contact attempts and non-response can be analyzed and visualized. The next set of question to try to answer is related to how information like this can be used in the steering of the data collection process.

- For example, how should the followed-up curve ideally look over the contact rounds if our purpose is to produce reliable statistics for the money available?
- At what point should we give up seeking contact with the sampled persons.
- At which times during the day should the contact trials be made?

Discussant: Lars Lyberg

Lars summarized the paper and pointed out some important issues from the Statistics Sweden's point of view. He remarked that the paper was somewhat oversimplified given the complexity of the problem. He mentioned that SCB must analyze why contact is not established in so many cases. One approach is to ask interviewers about their work procedures. He pointed out that increased response rate might increase non-response bias meaning that contact data cannot take non-response into account. He also suggested experiments on different maximum numbers of allowed contact attempts.

Discussion

Julia's suggestion was to perform sensitivity analysis. Interviewers should be involved in the project as much as possible.

Daniel's comment: Why to use maximum number of interviewers at a given cost? What is the probability of getting refusal given you established a contact? Essentially, you have to increase sample size given the cost.

Gustaf commented that there are some rules about contact days etc.

Natalie suggested studying more literature about similar studies.

Thomas commented that the response probability depends on attempts that you make.

Jan's comments:

- How many contacts do you have to make?
- Look at how the estimates behave after a while if they stabilize.
- Non-contacts usually increase (drastically).
- Most important groups cause most bias.
- What is the proportion of those who cannot find telephone number?
Many non-functioning numbers are common.

Julia's suggestion: If you really want to know you must get in touch with people you usually do not get in touch with- perhaps by paying amount of money for participation.

Day 2

Feedback: SIMSTAT

Speaker: Can Tongur & Jennie Bergman

Summary of presentation

Can and Jennie summarized follow-up on SIMSTAT (topic from October 2014). Recommendations from Advisory Scientific Board used to design a Swedish proposal (a sample survey design). This was presented to Eurostat and Member States in January 2015. Work by Eurostat and Member States will continue in parallel until early 2016.

Jennie and Can presented the Swedish proposal to re-design of Intrastat. Two sampling alternatives have been introduced. A couple of questions was asked to the Board at the end of presentation: How about quality and how to assess and communicate this new concept?

Topic 1: The behavior of balance indicators under various response models

Speaker: Pär Karlsson

Summary of presentation

Responsive designs have been suggested in order to reduce the potential negative effects of nonresponse, by means of changing the collection strategy during the data collection. A prerequisite for adapting the strategy is to have an indicator of the data collection process. The individuals to be contacted in the next round are then chosen in order to generate a "better" value of the indicator when the round is over. Several such indicators have been suggested in the literature, for example the representativeness indicator by (Scouten, Cobben, & Bethlehem, 2009) and the various balance/distance indicators suggested by (Särndal, 2011). Beside these two publications, these authors have discussed the indicators in various other publications. A simulation study of a responsive design is presented in (Lundquist & Särndal, 2013).

Here the properties of the indicators have been evaluated by simulations, where subsamples are drawn from a real sample. These subsamples will be the sets of respondents in the calculations of the indicator.

The real sample comes from one of the surveys that Statistics Sweden performs. The identity of the survey is not of importance for the purpose of this paper,

rather if known it might generate discussions related to this particular survey. The sample consists of about 7000 individuals. Register variables such as age, gender, income, education etc. are available for all individuals in the sample. The variables used here are anonymised to x_1 , x_2 , and x_3 . The response rate in the original survey was about 60%.

In this report you have seen how the R-indicator varies during the data collection phase of a hypothetical survey. The same methodology could be applied to any of the other indicators.

The main question related to the ultimate goal of implementing responsive design is:

- What objective function should be optimised in a responsive design? or at least what kind of information should be incorporated into the objective function?

Other questions:

- Do you think these kinds of simulations are useful in interpreting the indicators calculated from a real survey?
- How can this approach be used in order to find the best model to base the indicators on?

Discussant: Daniel Thorburn

Daniel welcomed the efforts in the respect with the topics discussed in the paper, such as balanced sampling, surveillance of the data collection, adaptive sampling design etc.

Some comments:

- The balance indicator is one-dimensional, but the problem is multidimensional.
- A good rule may be always to oversample in groups with large non-responses. But the optimum is not where the balance indicator indicates the best value.
- Do not use an indicator based on calibration (propensity ...) if the estimates are done by calibration (propensity scores ...). Better to use the information in an alternate way.
- Non-parametric – find a distance function. Estimate the nonresponse probability by averaging the response over the 100 closest sample points. (or vice versa).

Some conclusions:

- The R-indicator is useless for small samples. About 1000 individuals must be drawn before any conclusions may be drawn, even though the response set is constructed so the greatest response bias (in absolute terms) should occur quite early. In practice one must often have a design that reacts faster. (How fast can one see the differences between the curves).
- The R-indicator is sensitive to over-fitting. The inclusion of the spurious interactions as explaining variable makes the indicator go down considerably, even when there is nothing to explain. Without knowing the exact models and relations I refrain from telling why this occurs here.
- There is no description how the one-dimensional indicator of this type should be used for the multivariate problem of deciding, which sample units that should be prioritized and which should not be pursued further.

- Even when the non-response does not depend on the auxiliaries more than by chance, improvements can be made. (The response set is unbalanced by chance. Balancing the response set may improve the estimates in the same way as post-stratification may do).
- Pär mentioned that costs should be included in the calculations. I agree.
- If good responsive design should be made one must know the costs for new call attempts – the chances to get an answer in subsequent attempts and other metadata. But also costs up till now, since oversampling is always an alternative.
- Even with equal costs optimum is not easy to find.

Answers to the questions:

- What objective function should be optimised ... ?
 - I do not see the importance of having an objective function at all. If costs are neglected always try to get a balanced sample (in this case all estimated response probabilities are the same). Even when there is no effect of the explaining variables (other than by chance) this will give an improvement.
- Are these kinds of simulations useful ... ?
 - Yes, in principle we recommend that all suggested methods are tried out in simulations before being tried in practice. No, not in this particular case. The interpretation in real surveys is more complicated than looking at one single figure.
- How can this approach be used to find the best model to base the indicators on?
 - I do not know. One reason is that it is nowhere said how this indicator should be used in practice. The simulations showed how sensitive these models were to over-fit, but I do not believe that there is any risk of over-fitting here. A response set that is more balanced than it would have been by chance is to be preferred. The answer should also depend on the objective of the study.
- More general recommendations are postponed to after the next talk (Peter's presentation).

Discussion

Natalie commented Pär's paper: When we look at R-indicators we look at CV and do not get that picture you showed. Great simulation but there is a lot of literature to find relevant issues.

Peter commented on the question about the need for indicators. We need them because we do not have good control over data collection.

Jan's comments:

- R-indicator is not a good indicator of representativity. It works terribly if you compare it over time.
- This kind of indicator does not have any value in regard with the data collection.
- Do not use R-indicator as an indicator of balance! In data collection better to use CV.

Geert: Daniel's suggestion is about to over-sample some groups.

Topic 2

Design, data collection and estimation

Speaker: Peter Lundquist

Summary of presentation

Three important areas in the production of an interviewer administered household survey are the design, the data collection, and the estimation. Survey design and estimation are by tradition assignments for the survey statistician whereas the data collection is left to other methodologists who not always work in the same organization. The survey climate today with increased nonresponse has introduced new strategies such as responsive design (or adaptive design). By these methods it has become more clear that the data collection is an important area for the survey statistician to consider. Interventions are made during data collection with the ambition to produce a well-balanced or well representative set of respondents. Irrespectively if the interventions are based on a plan in advance or based on observations during the data collection, accurate estimation remains the ultimate goal.

An objective in this presentation is to examine the question:

- What benefits can be expected at the estimation stage from having improved the balance of the response set during data collection?
- Somewhat improved balance does not by itself guarantee estimates with low bias. Is balance worth a perhaps costly and demanding effort at the data collection stage?
- Or could one have done equally well by saving the use of the auxiliary information until the estimation stage?

Complete elimination of the bias is not achieved in the estimation phase either, but the goal is best accuracy (least bias) given the circumstances. Data from a Swedish survey will be used in simulations to illustrate the questions above. The evaluation will be based on recently developed indicators for nonresponse.

Discussant: Daniel Thorburn

Daniel presented the topic emphasizing that studies of this kind must be used for finding good adaptive designs. Non-response is indeed a big problem and one has to use all available data in order to correct the estimates as well as possible. However there will always be a bias part that cannot be corrected.

Daniel's comments on the example in Peter's presentation:

- Interesting. Peter wants to study the effects of
 - The level of relation between the auxiliary and the study variables,
 - a planned stop in data collection for units when a certain response level has been reached,or
 - the estimated population total,
 - one special effect variable, assessed income,
 - two real studies.
- It would have been interesting to see the effects on comparability over time, between subgroups and on other variables like inequality measures. The effects of a smaller sample is more important for comparisons and when the relations may change over time and vary between groups. (For LCS these are the main objectives).

There are two propositions in the paper:

- First proposition: The variance of a GREG estimator consists of two parts where the second is roughly proportional to a balance indicator and disappears when the sample is balanced.
- Second proposition: The bias is proportional to which is linear in the balance indicator but independent of artificial changes in R

In practice:

- Statistics, Sweden performs these surveys regularly. They get a lot of information on response behaviour. Already before the survey starts they know which units that have high non-response probability.
- Why wait to use that data till half the study is done. If a balanced response set is the goal plan for that already before the start e.g.
 - Oversample difficult strata/units
 - Start the data-collection earlier for those units
 - Exclude difficult units from the frame – designing other studies for them.
- Reaction during the sampling phase should only be to unforeseen problems.

Some recommendations:

- The Scientific Board welcomes the use of adaptive sampling/responsive sampling in order to get more informative samples and to cope with unexpected problems.
- We also welcome simulations of the performance of different adaptive methods.
- Statistics Sweden must realise that balancing improves the variance slightly but not (the main part of) the bias. There is no shortcut in removing the bias. This must be attacked by traditional statistical means like better field work or follow ups.
- In practice, Statistics Sweden already before the data collection knows which groups that have small response propensity. Statistics Sweden should plan the sampling with this in mind already at start. Possible ways are prioritizing in the sampling the difficult groups, starting the data collection period one week earlier for difficult groups or oversampling of difficult groups. Regardless of how this is done the estimation should take the auxiliary variables into account.
- Statistics, Sweden should not worry too much about indicators like the R-indicator. They are not indicators of the size of the bias. They mainly measure how much difference there is between estimators taking known auxiliaries into account and the simple average. There are better ways to estimate the size of the possible non-response bias.
- Statistics, Sweden should try to estimate the response probabilities as functions of the auxiliaries. Elements in the non-response set with high non-response probabilities should get a priority in the sampling procedures.
- When cost aspects are taken into account the goal should seldom be a fully balanced response set, since this means a smaller total sample size. Usually the optimum is somewhere between non-adaptive sampling and fully balanced response sets.
- One problem though is that using different methods for balancing and estimation creates problems when estimation variances. Standard methods should not be used.
- In order to adapt the sampling optimally, it is important to have good cost data. We recommend Statistics Sweden to develop good meta-data

in order to estimate the probabilities of getting responses in future call attempts and also the relevant costs for different design decisions.

Discussion

General discussion about the both topics presented in the session.

Xavier commented that some people have probability zero to be contacted. He also asked whether the eq (1) assumes linearity. If yes, then it is just an approximation.

Jan: Response rate as another measure of imbalance.

Natalie: What about confidence intervals?

After the general discussion, Lilli closed the meeting by thanking everyone for participating.