Statistical Information Communication: past insights, future innovations

Reinoud Stoel, Vivian Meertens, Nynke Krol, Martijn Tennekes, Edwin de Jonge, Marco Puts, Nicole Bongers, Duco Hoogendoorn, Darren den Boer & Ger Snijkers (Statistics Netherlands)

1. INTRODUCTION

National Statistical Institutes (NSI's) publish statistical information on a daily basis in many forms: press releases, corporate articles on website, reports, and tables. Sometimes infographics are used to illustrate text or in a standalone form. The target audience for all these types of output is broad. It can be an interested citizen, a student searching for data for a thesis, a journalist, a researcher, a government official, or a member of parliament. All of them have their own skills and wishes, and for all of them it's important that they understand and properly interpret the published information (in whatever form), and that they can use the information for the intended purposes.

Just like any NSI, Statistics Netherlands (CBS) puts a lot of effort in the quality of the statistical output to be published. Quality is important at every stage of the production process. Data collection, sampling, processing, and estimation methods are designed to ensure that our statistics are 'unbiased and precise'. Quality frameworks, like the Total Survey Error (TSE) Framework or the GSBPM (Generic Statistical Business Process Model), are used to describe the quality issues at every step in the process. At the output stage quality criteria like timeliness and relevance are used (Eurostat Statistics Code of Practice), but also clarity. CBS does yet not have explicit methods in place to empirically study the quality of our statistical information communication, as is the case for other quality criteria.

Current approaches of communicating official statistics are often based on convenience, user experience and common journalistic principles and practices. Quality aspects with respect to the understanding of the message of statistical communication are less known, and have not received much attention yet. This was the topic of our presentation[1] at the 2023 Conference on New Techniques and Technologies for Statistics (NTTS). We stressed that it is unclear how published information comes across and how this in turn effects the understanding and interpretation of the published statistics. Naïve statistical information communication may introduce the risk of perception bias in a chart or prime a reader to draw premature conclusions, and may result in an adequate assessment of the (in)accuracy of statistics. We suggested several research directions that can improve the effectiveness of (visual) communication of NSI's. Although it may seem a trivial question, surprisingly little research has been done in this direction. Research has been done in several academic disciplines, such as information visualization, statistics, computer science (information theory), and psychology, but a holistic, interdisciplinary, approach is missing. In our 2023 presentation we encouraged the official statistics community to study how users perceive, interpret, and use their statistical information, and to develop methods and guidelines to elevate the quality of communication above "best practices". We presented our plans to conduct user studies at CBS, both in a controlled environment and in a free environment (for qualitative studies).

1.1. A new theme in the CBS Research Program Methodology

CBS aims for the production of reliable, detailed, timely and phenomenon-oriented statistics, while balancing costs and burden for society. Innovations in the statistical process and business operations should make this possible. Our CBS Research Program Methodology (RPM), run by the R&D Department, paves the way at CBS and delivers knowledge (of processes, methods, and data sources) and skills to facilitate future innovations, transitions and/or improvements in the statistical process. Within our RPM the work is divided into themes, with each theme having a specific niche. Every five years our RPM is calibrated against new societal developments, new production process demands and missing methods, and subsequent consequences for the production of our statistics. Because of the importance stressed in the introduction, our new 2025-2030 RPM will contain a new theme with the title “Statistical Information Communication”.

In the current presentation we will present what we have developed so far, what our plans are for the near future with this new theme, and how the collaboration between the Methodology Department and the Communication Department of CBS was formed. An important goal of this theme is methods development: developing methods to assess the quality of statistical information communication.

2. FUTURE INNOVATIONS: STATISTICAL INFORMATION COMMUNICATION

Research into the quality of statistical information communication is not yet a routine at CBS and faces challenges from many knowledge domains. One of the first challenges is determining unambiguously the relevant quality criteria, as well as the operationalisation of these criteria, i.e. how to assess quality according to these criteria. This is certainly non-trivial. For example, just as our statistical information should be ‘unbiased and precise', we could also state that our publications should be, and can be, correctly interpreted, valued and used. The communication format affects how the information is interpreted, valued and used. Additionally, for instance, the attractiveness and accessibility of a publication can also be used as quality criteria: an infographic should also be appealing to look at (and thus attract attention); a long block of text is not very attractive to read for all users. There are multiple quality criteria, and certain criteria may be in conflict with each other.

As for correctly interpreting statistical information (one of the above-mentioned quality criteria), we see several challenges. Our underlying assumption is that people (i.e., users of our information) make an effort to take in and process this information systematically. However, psychological research shows the opposite. What plays a role here is that people are reading less well and paying less attention: they have less patience for taking in and processing information. If a piece of text takes up more than one page, many people often find it too long already. Texts are increasingly being supplemented with 'infographics'. CBS takes these factors into account in all standard communication with the outside world. Additionally, we know that 'innumeracy', or a lack of understanding of numbers, is a widespread phenomenon. In general, people are bad at working with numbers and interpreting them. This is also an aspect we need to deal with: we publish statistical data about the Netherlands, and as such, our data requires interpretation. Statistical data are not hard facts, but rather have an uncertainty margin. Here we speak of 'statistical literacy'. There has been a lot published on innumeracy and statistical literacy. The motivation to process information, and reading and numerical skills will vary per user and user group. Furthermore, cognitive factors, such as how our brain (graphically) processes information, and its limitations (such as colour blindness), also play a role. This requires a multidisciplinary approach. The challenge is to create statistical information that is directly usable by users without them first having to learn how to work with numbers. The focus for the research theme Statistical Information Communication will be on empirical research in which we put the user at the core.

As for operationalizing the quality of statistical information communication for different user groups, we see the following research questions:

- How can we study the quality of our statistical information communication?
- Which quality criteria for texts and graphics can we distil from this?
- Which knowledge domains are supportive here?
- What do these disciplines say about the quality of statistical information communication?
- What knowledge do we already possess, and what do we need to develop?

For measuring the quality of our statistical information communication, we have the following research questions:

- Which research techniques (such as for example user research or 'eye-tracking' techniques) are already available to study the quality of our statistical information communication?
- Which ones work well in practice and which ones don't?
- Are the techniques also methodologically sound?

Finally, depending on the knowledge gained from the first and second research questions, we can determine how our statistical information can be improved.

- What changes to texts, tables and visualisations can improve the readability of our statistical information?

3. PAST INSIGHTS

In recent years, several exploratory studies have been undertaken to evaluate the quality of statistical information communication at CBS. These studies aimed to identify how users interpret and interact with statistical information, laying the groundwork for future research:

- A qualitative Study on Graph Interpretation: This study explored how a broad audience interprets statistical graphs. Participants were shown a graph and asked to describe what they understood from it.
- A study on the Interpretation of the word 'most': In this ad-hoc experiment, CBS colleagues were asked to interpret the phrase, "In a jar of marbles of different colours, most of them are red." Responses varied widely, indicating that even simple terms like "most" can be understood differently.
- An exploratory Study on Colour Perception: This study focused on how users perceive and interpret colours in graphs and infographics. The study highlighted the need for more thoughtful colour selection in data visualizations, particularly in regard to accessibility (e.g., colour blindness).
- Pre-testing Methods for Questionnaire Quality: we reviewed existing pre-test methods used to assess the quality of survey questionnaires and examined their potential application for evaluating statistical information communication. The study categorized and analysed methods, providing insights into how similar approaches could be adapted to test the clarity and effectiveness of published statistical information.

These exploratory efforts underline the importance of moving toward a more structured and systematic research agenda. The insights gained thus far will inform the development of methods and criteria that can be systematically applied to improve the quality of statistical information communication, ensuring they are more easily understood and accurately interpreted by diverse user groups.

The main findings of these exploratory studies will be discussed in detail at the conference, particularly with regard to the development of methods for evaluating statistical information communication. A key takeaway is that methodological rigor is essential in designing and

implementing studies to increase the likelihood of correct interpretation of the statistical information provided.

4. REFERENCES

[1] Martijn Tennekes, Marco Puts, Edwin de Jonge, Vivian Meertens, and Reinoud Stoel *(2023). Visualization and communication of statistics*. Presented at the 2023 Conference on New Techniques and Technologies for Statistics.