

Discussion of “Probability vs. Nonprobability Sampling: From the Birth of Survey Sampling to the Present Day” by Graham Kalton

Ralf Münnich¹

Let me first thank Dr. Kalton for his amazing historical review of the development of survey sampling from its origin, contrasting purposive sampling, until now, where some elements of purposive sampling in terms of web or big data seem to supersede the well-elaborated theory of survey statistics. Shall the message be that we do not need any sampling courses at universities anymore, that official statistics should turn to modelling using data with unknown data generating processes, or actually even be substituted by (commercial) *data krakens*? Hardly so! Graham Kalton emphasises a modern thinking about the use of these new data sources which may also have some advantages and he urges future research on data integration methods using (very) different kinds of data while strongly taking quality aspects into account.

Within the last decade, we could observe many new uses of classical data like administrative data and new types of data stemming from internet sources or technical measurement processes such as satellite, mobile phone or scanner data. Already the availability of these new data leads to a huge increase in developing new methodologies and uses. Indeed, official statistics also forced research on new data types, such as scanner data or web-scraped data and others. In Europe, these statistics are often called experimental statistics to emphasise that these statistics cannot (yet) be evaluated using the classical quality concepts, as, e.g. proposed within the European Statistics Code of Practice (<https://ec.europa.eu/eurostat/web/quality/european-quality-standards/european-statistics-code-of-practice>). Some examples can be drawn from https://www.destatis.de/EN/Service/EXDAT/_node.html or <https://ec.europa.eu/eurostat/web/experimental-statistics>.

During the Covid crisis, and especially in light of the political discussion in Germany, however, one could observe little understanding of data quality and statistics. Timeliness – with its urge of getting data and producing statistics immediately – often lead to the use of available (infection) data, which certainly were influenced by unknown biases. The impact of statistics on these available data in terms of evidence-based policy could hardly be understood at the time, but still legal processes like

¹ Economics, Economic and Social Statistics Department, Trier University, Germany.

E-mail: muennich@uni-trier.de. ORCID: <https://orcid.org/0000-0001-8285-5667>.



lockdowns were initiated. To state this message more strongly: whenever a legislation process is involved, and especially so if a direct impact on society is the outcome, we must make sure that high quality requirements on data gathering and statistical methodology are set as well as met. High quality typically cannot be achieved with low costs. England was one of the few very good examples during the pandemic, since they were setting up a special Covid survey to better understand the pandemic and to provide adequate and reliable information.

Certainly, this example already shows some critical aspects in data gathering and data quality. Dr. Kalton was emphasising timeliness and accuracy as very important goals of data quality. For sure, these are of utmost importance! However, in practice, both quality principles suffer from budget constraints and cost controls. This directly leads to two questions: Do modern data help to provide more timely and accurate statistics at lower costs? Is there, in case of conflicts, an *ultimate* quality principle?

The first question is already answered by Dr. Kalton. Of course, modern web or big data can help to gather information quickly. Interesting approaches are of course the use of satellite or scanner data. With electronic cash systems, price changes could be tracked much faster than via the use of survey data. However, one always has to understand the advantages as well as the disadvantages of these data generation processes, and one must be able to measure the quality of the output.

Let me briefly sketch one current German debate which, in my view, perfectly fits into this discussion. In the past years, more and more internet surveys were preferred to data from traditional market and opinion research. This immediately led to a discussion on the quality of the outcomes. And certainly, timeliness, accuracy, and costs played an important role within this discussion. The two major arguments were the following: internet surveys suffer from unknown biases. Classical surveys, in the meantime, have to consider response rates considerably below 20%. Under these conditions, most likely both areas have to consider statistical models with strong assumptions to at least reduce possible biases induced by either web surveys or non-response. In my view, one important question has not been raised yet. What is the aim of the survey?

The ultimate aim that necessitates data collection in the first place is of crucial importance for evaluating the importance of the different quality principles. In case one is interested in getting information on current public opinion, probably timeliness and costs are more important than high accuracy. However, in evidence-based policy making, and especially when information for legislative action is needed, I must stress that accuracy must always be considered to be the major principle. This is even more important when large budgets or financial equalization schemes are involved. Additionally, in these cases one must also be able to measure the quality of the outcome of the statistics. This is still a major drawback of using web or big data. And to stress this point, in legislation processes, I strongly urge to involve independent official statistics with its transparent data production process.

With this discussion, I do not want to be misunderstood. Modern data and modern statistical methods are important. And the direction of research, as Dr. Kalton pointed out, will be complex modelling and data integration. Also administrative, register, and related data are important and can provide very good information. However, with all these data, we always have to understand their quality and we should be able to measure the quality of the resulting statistics. Especially in the context of big data, quality measurement may have to be enhanced (cf. Münnich and Articus, 2022, and the citations therein).

Sampling itself may also follow new directions. Classical sampling optimization may be adequately applied in more special cases that allow focusing on specific goals, e.g. the design optimization in the German Censuses 2011 and 2022 (see Münnich et al., 2012, and Burgard, Münnich, and Rupp, 2020). However, likely robustness of methods against assumptions has to be incorporated in design optimization. On the other hand, data integration, multi-source environments, geo-spatial modelling, small area estimation and other modern methods may yield new ideas and directions in sampling theory and application. One example may be sampling from big data sources to reduce complexity.

Despite the mentioned new directions, many ideas have been well-known for a long time. In data analytics, we differentiate between descriptive, predictive, and prescriptive aims. Data that were gathered to describe a state of a system cannot be used to analyse interventions on the system. Indeed, we need the right data and not just merely available data. In conclusion, the exact purpose of the statistics under consideration plays an extremely important role for the selection of data and the priority of the different quality principles.

References

- Burgard, J. P., Münnich, R., & Rupp, M., (2020). Qualitätszielfunktionen für stark variierende Gemeindegrößen im Zensus 2021. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 14(1), pp. 5–65. With discussion.
- Münnich, R., Articus, C., (2022): Big Data und Qualität – ist viel gleich gut? Pp 85–101. In: Wawrzyniak, B., Herter, M. (Ed.): *Neue Dimensionen in Data Science*. Wichmann.
- Münnich, R., Gabler, S., Ganninger, M., Burgard, J. P.; Kolb, J.-P., (2012): Stichprobenoptimierung und Schätzung im Zensus 2011. *Destatis: Wiesbaden, Statistik und Wissenschaft*, Vol. 21, https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Monografien_Archiv/2012_07_Destatis_Stochprobenoptimierung_und_Schaetzung_im_Zensus.pdf?__blob=publicationFile&v=12