*STATISTICS IN TRANSITION new series, June 2023*
*Vol. 24, No. 3, pp. 43–45,* https://doi.org/10.59170/stattrans-2023-034
Received – 29.05.2023; accepted – 31.05.2023

# Rejoinder

**Graham Kalton**[1]

I should like to thank the discussants for their kind remarks, for their valuable comments on the present state and future directions of the field, and for the many references they cite. Since I have no disagreements with them, I will confine my rejoinder to a few issues that their contributions have surfaced for me.

I will start by rectifying an oversight in my treatment of the early history of survey research and survey sampling: Carl-Erik Särndal has reminded me of the major developments that occurred in Russia during the early years. The impetus for these developments was the need for local self-government units known as zemstva to collect data about their populations for administrative purposes. Initially such data were collected with 100% enumerations, but around 1875 sample surveys were introduced for cost savings. The survey procedures were coordinated across zemstva and a number of sampling methods were evaluated with input from theoretical statisticians. These statisticians made a number of important contributions, including an impressive early text (1924) entitled The Foundations of the Theory of the Sampling Method by A. G. Kowalsky. Although Russian statisticians were at the frontiers of developments in survey sampling until the late 1920's, their contributions were not fully recognized outside Russia. For example, Tschuprow (1923) and Kowalsky in his 1924 text both derived the optimum allocation formula for stratified sampling a decade before Neyman did so in his famous 1934 paper (after learning of Tschuprow's paper, Neyman (1952) recognized Tschuprow's priority for the results). Mespoulet (2002), Zarkovic (1956), Zarkovic (1962), and Seneta (1985) provide further details about early survey research and research on survey sampling in Russia.

Danny Pfeffermann has pointed out that probability samples are almost never representative because of nonresponse—and I would add noncoverage—that is not missing completely at random (NMAR or MCAR). Moreover, I do not think the nonresponse should be viewed as missing at random (MAR), that is MCAR after conditioning on known covariates. Using standard weighting adjustments based on known covariates will not make the sample representative. My favorite quotation from George Box is "Essentially, all models are wrong, but some are useful." Nonresponse adjustments should be viewed from this perspective as useful but not perfect. Another George Box quotation: "Statisticians, like artists, have the bad habit of falling in love with

[1] Joint Program in Survey Methodology, University of Maryland, College Park, MD, USA.
E-mail: gkalton@gmail.com. ORCID: https://orcid.org/0000-0002-9685-2616.

their models." But there is a difference: artists have artistic license to paint over a model's blemishes whereas statisticians should attempt to identify and repair the blemishes.

Risto Lehtonen points out the considerable attractions of population registers, as have existed for some time in several Scandinavian countries and are in development elsewhere. Such registers can be viewed as surveys with 100% samples, and the quality of their data should be assessed accordingly: What is their actual coverage? How up-to date are they? How accurate are the data they contain?

Risto's discussion of population registers also reminded me of a point that I should have addressed more fully: there is a wide variation in the data infrastructure for social research across countries. For example, most developing countries are not in a position to use administrative records or the internet. They rely on probability sample surveys to satisfy their data needs. Fortunately, they have not yet experienced the severe declines in response rates that are so harmful to surveys in most high-income countries.

Julie Gershunskaya and Partha Lahiri address two important current areas of research. One is the research on how to employ a probability sample to reduce the bias in estimates from a nonprobability sample, making use of auxiliary variables collected in both samples. The auxiliary variables aim to capture the key variables that are predictors of membership in the nonprobability sample. Challenges to be addressed with this approach include identifying the key variables; dealing with the fact that some response categories that occur frequently in the probability sample are very sparsely represented in the nonprobability sample; and concerns about the equivalence of the responses to the key variables obtained in the two samples that use different modes of collection. The results from this approach should be viewed with caution. However, recalling George Box's quotation above, imperfect models can be useful. Julie and Partha rightly say that the aim of these models is to reduce, not eliminate, bias. The question to be asked is how to assess whether the models have reduced bias to an acceptable level.

The second area that Julie and Partha address is small area estimation. I should have written more about this methodology whose use has now become so widespread. My first practical exposure to small area estimation occurred in the late 1990's, when I chaired a National Academy of Sciences' panel that was asked to advise about the quality of the small area estimates of the numbers of poor school-aged children that were being developed in the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program. The central issue was whether the estimates, which were produced for 3,000 counties and 14,000 school districts, were appropriate and sufficiently reliable to be used in allocating very large sums of money directly to school districts. At that time, this was a novel application of small area estimates, and subject to considerable questioning. After extensive evaluation of the area level models by both the Panel and the Census Bureau (Citro and Kalton, 2000), the Panel concluded that the small area estimates were "fit for use" for the purpose of this fund allocation, despite a recognition of substantial errors in the individual estimates. The Panel was influenced by the fact that the legislation stipulated that the funds should be distributed directly to the school districts and that, even though the small area estimates were not ideal, they were the best available. I was persuaded by my experience on the Panel that, with strong predictors and careful model development and testing, small area estimation methods have an important role to play in responding to policy makers' increasing demands for local area estimates.

Ralf Münnich emphasizes the importance of assessing the overall quality of statistical estimates in the light of the uses of the estimates. As he notes, timeliness is often in conflict with accuracy. In some situations, timeliness may be paramount, and accuracy may suffer. However, one must guard against the risk that accuracy is so low that the resulting estimates are misleading. Estimates based on big data sources or even large surveys conducted with an overriding emphasis on speed may, because of their sample sizes, appear to be well-grounded but that may well be illusory.

It is often argued that although individual estimates may be subject to serious biases, these biases will cancel out for differences between estimates, either between subgroups of the sample or across time. While the underlying model for that argument often appears reasonable, the assumptions underpinning it need to be carefully assessed in each case.

Ralf also points out the importance of cost constraints. When the cost constraints severely limit a study to a very small sample size, it may be preferable to forego the extra costs involved in selecting and fielding a probability sample, in favor of a quasi-probability sample or a nonprobability sample design. As Kish (1965, p. 29) notes: "Probability sampling is not a dogma, but a strategy, especially for large numbers."

Finally, Ralf and other discussants have pointed out the attractions of data integration. I also see these attractions, but I think that the challenges of mode effects arising from different data sources should not be underestimated.

***In conclusion, I congratulate Statistics in Transition on celebrating its 30th anniversary. It plays a distinct and important role among statistics journals. With the major changes in statistical methodology taking place in official statistics and in social research, it has a bright future for the contributions it can make.***

## References


Citro, C. F., Kalton, G. Eds., (2000). *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology*. National Academy Press, Washington D.C.

Kish, L., (1965). *Survey Sampling*. Wiley, New York.

Mespoulet, M., (2002). From typical areas to random sampling: sampling methods in Russia from 1875 to 1930. *Science in Context*, 15(3), pp. 411–425.

Neyman, J., (1952). Recognition of priority. *Journal of the Royal Statistical* Society, A, 115(4), 602.

Seneta, E., (1985). A sketch of the history of survey sampling in Russia. *Journal of the Royal Statistical* Society, A, 148(2), pp. 118–125.

Tschuprow, A. A., (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, 2(4), pp. 646–683.

Zarkovic, S. S., (1956). Note on the history of sampling methods in Russia. *Journal of the Royal Statistical Society, A,* 119(3), pp. 336–338.

Zarkovic, S. S., (1962). A supplement to "Note on the history of sampling methods in Russia". *Journal of the Royal Statistical Society, A*, 125(4), pp. 580–582.