**Comments on the scientific basis of the register-based census**

Anders Wallgren and Britt Wallgren

*BA Statistiksystem AB, formerly of Statistics Sweden*

*E-mail:* *ba.statistik@telia.com*

### Introduction

A recent paper by Alphonse L. MacDonald [1] compares the scientific basis of the classical or traditional census with the scientific basis of the new survey methods used in the register-based population and housing census. These methods were developed by Statistics Denmark and are described in a 1995 Eurostat publication [2]. Our own books on register-based statistics build on the Danish book, our experiences from Statistic Sweden [3] and from our contacts with statistical institutes that want to replace the traditional census with statistical registers [4].

We are quoted in MacDonald’s paper. In the abstract McDonald says: *“… some countries started to redesign their statistical organisation favouring registers and replacing the census, with a system of combined registers and sample surveys. Proponents of this approach indicate that this procedure has no theoretical basis …”*

We are the proponents MacDonald mentions and we want to clarify our view on this issue. MacDonald quotes the following text on page 3 from the second edition of our book [3]:

*“… no well-established theory in the field exists. There are no recognised terms or principles, which makes the development of register-based statistics and register-statistical methodology all the more difficult. As a consequence, ad hoc methods are used instead of methods based on a generally accepted theory.”*

### Is there a theory?

We interpret the concept “survey theory” in the same way as Robert M. Groves [5]. The costs for the survey should be reduced and total survey errors should be minimised. So, when we compare the design of the traditional census with the register-based census, we should compare costs and total survey errors and then decide which design is the best in our country under the present preconditions.

Many different types of people have recently started to work with statistical registers and most of them do not follow a unifying theory with a common terminology. Perhaps many also instinctively look to a leading NSO for inspiration such as the Census Bureau in the US when they instead should go to Statistics Denmark, which is the pioneer NSO regarding register-based censuses.

A paper by David J. Hand [6] was presented before the Royal Statistical Society in 2018. Hand notes: *“Given that so many official and economic statistics are based on administrative data…, we might have expected there to be a substantial literature in the leading methodological statistical journals describing the statistical challenges and how to overcome them. This appears not to be so, with such journals carrying relatively few papers on the statistical challenges of administrative data (being mostly focused on the consequences of sampling theory)… More generally, papers on the topic seem to be widely scattered and often appear in the proceedings of conferences and workshops, or perhaps as reports … from official statistics offices.”*

Perhaps the present state of register survey theory is similar to the situation in 1987 when Groves wrote his paper. The first section in his paper has the title “Survey research as a Methodology Without a Unifying Theory”.

Our conclusion is that this present state results in a diversity of terms and principles instead of a unified theory. Hand also points out that statistical textbooks often only mention random sampling and sampling errors, which gives the students (and their teachers, our comment) a limited view of statistics.

The foundations of a theory regarding register surveys and register systems were developed as early as during the 1960s. Svein Nordbotten [7], then at Statistics Norway, developed ideas on *statistical information systems*. He explained that administrative sources should be used for statistical purposes and introduced what he called *statistical file systems* (what we now call register systems). With the first register-based census, Statistics Denmark developed new survey methods so that administrative registers could be used to estimate the same parameters as with the traditional census. Our conclusion is that there is a theory, but the theory is not yet well-established.

### The traditional census and the register-based census - main differences

*1. Populations and frames*

The classical or traditional census uses an area frame linked to cartography. The target population consists of the persons and households present in the country at the census day.

The register-based census is based on a statistical Population Register. Since the Population Register is linked to the Dwelling Register, there are three different register populations that can be used for a register-based census:

First, the Population Register is based on the population of all *demographic events* as births, deaths and migration. All of these events have time references: the day when the event occurred and the day when the event was registered in the register.

From this population of events, a second population can be created of *persons* registered as active members of the population at the census day. In this way, all registered persons and all dwelling households are placed in both time and space according to the census requirements.

The third register is the Dwelling Register. This register could be used as an area frame where interviewers could be sent out to interview the *households* in a random sample of dwellings. This sample survey could be used for quality assurance regarding coverage errors in the register. If the rules for inclusion in the Population Register are not suitable for the census, then the errors can be measured with this sample survey.

Our conclusion is that in a national statistical system with a Population Register and a Dwelling Register, the preconditions for a register-based census that meets the needs for the short form in a traditional census are at least as good as for the short form in a traditional census.

*2. Variables*

MacDonald argues that characteristics are not observed in a register-based census. Instead, they are constructed by using already existing information in registers, which he maintains has no scientific basis. We think that the interviewers in a traditional census do not observe the characteristics of persons; they do not see age, level and kind of education or income. They ask questions, and the answers are combinations of the true values plus measurement errors. They can observe characteristics regarding housing, and they can ask question regarding ethnicity and religion that are not in registers due to privacy concerns.

Errors in the recorded variables also exist in the registers, which is perhaps because the information is old and has not been updated. Quality assurance of the variables should be done on the micro level. The register variables and the sample survey variables should be compared for the same persons. We have conducted exercises of this kind and the results can be found in our book – there are errors in both registers and in sample surveys – quality assurance of this kind must be done on a regular basis.

### Conclusions – what about the scientific basis?

The scientific basis of the classical or traditional census is well-established, and advanced methods have been developed to handle issues regarding coverage and other error sources. The register-based census also has a scientific basis. The new methods are developing and will gradually become established. The choice between the two methodologies should be based on a simultaneous judgment of costs and quality. Timeliness is the quality factor that is the main quality problem with the traditional census. Due to the Corona crisis many developing countries will have problems with the funding of a traditional census.

We came to Statistics Sweden during 1989, just before the last classical census for 1990. We noticed that several competent statisticians were negative about the change. The census was ‘real statistical methodology’ and the quality survey regarding the census was ‘real quality assessment’. And all this would be lost when registers gradually replace the census! With this background, we can understand Alphonse L. MacDonald attitudes regarding the census. But finally, money and quality will talk.

### References

1. MacDonald, Alphonse L. Of science and statistics: The scientific basis of the census. Statistical Journal of the IAOS 36 17–34.
DOI 10.3233/SJI-190596, IOS Press 2020.
2. Statistics Denmark. Statistics on Persons in Denmark – A register-based statistical system. Eurostat 1995.
3. Wallgren A, Wallgren B. Register-based statistics. Statistical methods for administrative data. Second edition, Chichester,
UK: John Wiley & Sons Ltd: 2014.
4. Wallgren A, Wallgren B. Register-based statistics. Registers and the National Statistical System. Third edition, Chichester,
UK: John Wiley & Sons Ltd: 2021.
5. Groves, R. M. Research on Survey Data Quality. Public Opinion Quarterly 51:S156-S172. University of Chicago Press 1987.
6. Hand, D. Statistical challenges of administrative and transaction data. J. R. Statist. Soc. A. 181, Part 3, 555-605. Paper read before the Royal Statistical Society 2018.
7. Nordbotten, S. Purposes, Problems and Ideas Related to Statistical File Systems. Proceedings from the 36th Session of the International Statistical Institute. Invited paper, Sydney (1967). Available for free downloading from [www.nordbotten.com](http://www.nordbotten.com)