

Uploaded

Comments on “Comments on the scientific basis of the register-based census by Anders Wallgren and Britt Wallgren”.

Alphonse L. MacDonald a;b

a Honorary Senior Adviser Methodology to the Director, Algemeen Bureau voor de Statistiek, Paramaribo, Suriname

b No. 2 Tal Ingliza, Triq l-Ewropa, Victoria, Gozo, Malta

Tel.: +356 2156 1009; E-mail: fonzhan@gmail.com

Introduction

The objective of my paper was to review the development of the census as a general scientific method and its underlying scientific principles over time, including the recent emergence of alternative procedures to the census, and to make a call for a science-wide review of the scientific principles underlying the different “types of censuses”. It is generally accepted that the census in population statistics is based on well-established scientific principles, and that it has an array of proven methods to deal with issues of quality, including completeness and errors. In the second half of the 20th century a number of European countries offered an alternative to the census, using data derived from existing administrative records instead of direct field enumeration, the so-called register-based census. Although Statistics Denmark was the first country to produce a complete set of population and housing “census-like” tables using registers of data derived from administrative records in 1981 [1, p. 44] , there is at present no standard methodology for their production. Moreover, some of the main proponents of the use of registers to replace direct data collection methods have indicated that there are no theoretical bases for these procedures. In my paper I only presented, briefly, some of the ideas of B. F. M. Bakker (Statistics Netherlands) [2, pp. 3 -5: 3, p. 17] and Anders Wallgren and Britt Wallgren (Statistics Sweden) [4, p. 3]. I will not repeat what I have written on Bart Bakker’s ideas, but for clarity’s sake I repeat my original Wallgren and Wallgren quote: “Although register-based statistics are a common form of statistics used for official statistics and business reports, 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.*” [4, p. 3, emphasis in the original]. Wallgren and Wallgren have reacted to my article, and it seems that they wish to distance themselves from the statement that “no well-established theory in the field exists”.

[5]. In my comments on their comments I will remain within the universe of discourse established by their comments, with one exception. In the final assessment of the theoretical basis of the so-called register-based census I will refer to a 2011 publication of Bart Bakker which was not referred to in my article nor in the comments of the commentators.

Comments

The commentators' contribution consists of four parts: an "Introduction"; a section called "Is there a theory?" a section called "The traditional census and the register-based census - main differences"; and finally the "Conclusions – what about the scientific basis?"

In the introduction the commentators partly reproduce my quote without comment.

In the section "Is there a theory?" the commentators refer to a 1987 article by Robert M. Groves [6] and indicate that they "interpret the concept "survey theory" in the same way" as he does, without quoting his definition¹. They state that "Perhaps the present state of register survey theory is similar to the situation [of survey research, added by author] 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."

In less than 20 years Robert M. Groves' 1987 article was superseded by the 2004 (text) book on survey methodology of which he is the main author. [7] In the Preface Groves and his co-authors state "...Over the past two decades a set of theories and principles has evolved that offer a unified perspective on the design, conduct, and evaluation of surveys. This perspective is most commonly labeled the "total survey error" paradigm. [7, p. xv] The 1987 concept of "survey theory", to which the commentators subscribe, is now replaced by the concept of "survey methodology". "Survey methodology seeks to identify principles about the design, collection, processing, and analysis of surveys that are linked to the cost and quality of survey estimates. This means that the field focuses on improving quality within cost constraints, or, alternatively, reducing costs for some fixed level of quality. "Quality" is defined within a

¹ A word on the terminology, in English language literature the generic common language term "survey" describing any type of inquiry or inventory is used, simultaneous with the more technical term as in "sample survey". A census becomes a "100 % survey". The methodology of the census was internationally standardised in the second half of the 19th century. Sample survey methodology was formalised in the second half of the 20th century, when exactly is a matter of opinion.

framework labeled the total survey error paradigm (...). Survey methodology is both a scientific field and a profession.” [7, p.32] According to Groves et al. to achieve high-quality survey results requires the application of the principles from several traditional academic disciplines, among which mathematics (statistics), social sciences, cognitive psychology and computer science. Except for the principles of computer science, the principles of the other sciences were established before 1960.

According to the commentators “The foundations of a theory regarding register surveys and register systems were developed as early as during the 1960s. “ They refer to the seminal 1967 paper of Svein Nordbotten [8] in which he developed the basis of electronic statistical data file systems. He defined the data file, its components, structure and hierarchy, specifying the conditions for the acquisition, distinguishing between continuous and non-continuous collection, storage, maintenance and processing of the data, including the combining of data files to create a statistical data system. He also described the relation between continuous data collection (registers) and the census. “The continuous collection scheme contributes to a more balanced distribution of collection activities over time and to the reduction of the number and extent of large censuses. There are, however, characteristics, particularly those for which there are no contemporary administrative needs, which may be most efficiently collected in a census. The censuses may also provide convenient status sets for checking the current registration.” [8, p.744]

The commentators state” 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.”

Section 3 “The traditional census and the register-based census - main differences” has no direct bearing on the theoretical principles of the census, and will not be discussed in detail. Just one observation the commentators miss the most salient difference between the classical or traditional census and the so-called register-based census. The first covers the total population of a specific territory at a specific moment, while the latter only deals with the registered population.

The commentators in the conclusions state:” 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.” So what is the scientific basis of the register-based census? The commentators don’t say, but want to compare the two methodologies on costs and quality. However one has a well-established and known theoretical basis, the other has either no scientific basis, as other proponents claim [e.g. 2, pp. 3 -5: 3, p. 17], or, as the commentators maintain, it has an unknown theoretical basis which is under development. This is a methodological impossibility. Comparing methodologies or procedures just on cost and quality is a very reductive view. This issue is what methodology is at the basis of the so-called register-based census! In my article I have referred to Bakker’s description of different methodological options. I would like to refer here to one of the options, namely micro integration, which is fundamental to register-based statistics, According to Bakker micro integration “involves matching data from statistical units at an individual level, with the goal of compiling better information than is possible when using the separate sources.” [9, p. 4]. But what is the status of micro integration? According to Bakker “There has been a long debate over the question of whether micro integration is actually a method or not. To a significant extent, micro integration cannot be generalised, in the sense that it requires knowledge of the sources and the actual situation in society. As a result, specific micro integration rules are continually drawn up and applied.” [9, p.5]

When comparing procedures it is fundamental to incorporate the pre-conditions and operational and technical requirements for each procedure in the assessment. So we have to compare different procedures not only methodologies on costs and quality. According to Wallgren and Wallgren “There is no commonly used way of describing the quality of register-based statistics.” [4, p. 4]. This means that only the cost aspect remains as an element of comparison and assessment of the procedures. One procedure, the classical or traditional census, has an established scientific basis and advanced methods to deal with coverage and other error sources, whereas the other, the so-called register-based census, either is without an established scientific basis, or has an unknown scientific basis that is under development and has no procedures to establish the coverage and for which there is no standard way of describing quality.

Although it is alleged that the so-called register-based census is cheaper than a classical or traditional census, as far as the author knows, a proper cost analysis for the society, not just the statistical office or system, has never been carried out.

Hence, my appeal remains, as science is self-regulating, where are the guardians of the science of statistics to bring order in chaos?

References:

1. Lange A. The population and housing census in a register based statistical system. *Statistical Journal of the IAOS*. 2014; 30: 41-45. Doi 10.3233/SJI-140798.
2. United Nations Economic Commission for Europe. Conference of European Statisticians. Statistical Office. Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses (The Hague, The Netherlands, 10–11 May 2010. Statistics Netherlands. Micro-integration: State of the Art. Working paper 10. 12 May 2010.
3. Bakker, BFM. Estimating the validity of administrative variables. *Statistica Neerlandica*. 2012; 66(1): 8-17.
4. Wallgren A, Wallgren B. Register-based statistics. Statistical methods for administrative data. Second edition, Chichester, UK: John Wiley & Sons Ltd: 2014
5. Wallgren A, Wallgren B. Comments on the scientific basis of the register-based census. Comments uploaded on Population censuses; are statistical dinosaurs able to adapt? <https://officialstatistics.com/news-blog/population-censuses-are-statistical-dinosaurs-able-adapt>
6. Groves, R. M. Research on Survey Data Quality. *Public Opinion Quarterly*. 1987: 51: 4-Part 2: S156-S172.
7. Groves, R.M. et al. *Survey Methodology*. Hoboken, New Jersey. John Wiley & Sons, Inc. 2004.
8. Nordbotten, S. Purposes, Problems and Ideas Related to Statistical File Systems. *Bulletin of the International Statistical Institute*. 42. 1967: 733 – 750.
9. Bakker, B. Micro integration. *Statistical Methods* (20118). The Hague. Statistics Netherlands. 2011.