Official Statistics
Research Report - Session 5

Youhee Kil (r0768512)

Fall 2021

Contents

1   Session 5. Census

1.1   Introduction

The census means complete enumeration survey and UNECE (2006) defined a population and housing census as "the process that produces interrelated information on the population and housing stock". The purpose of the census is to find out the characteristics of the population of the individual country or regions for better and accurate government policy and infrastructure based on the results of the census. Recent social changes lead more rejection of response and increase cost of traditional based survey, most of countries are trying to implement new census methodology for cost efficiency and quality improvement. In this report, two methodologies will be covered mainly, virtual census and rolling census, by comparing census methods of Netherlands and France. According to UNECE (2006), the noticeable trend is more and more countries are moving or had moved to either a combined or register-based census methodology which is use combination of data from registers and field data collection or use data from registers only. Most of North European countries selected register-based census methodologies, on the other hand, Canada, USA, and France selected to use a bit developed version of traditional censuses. Section 2, comparisons between two cases, Netherlands and France, where using two different census methodologies will be performed. Section 3, the method of data collection of each census methodology is explained with details. Also, the advantages and downsides of each methods are discussed and checked if each census is satisfying five essential features of the census.

1.2   Census Methodology in each country

The Netherlands uses *Virtual Census*, on the other hand, the France uses *Rolling Census*. The process of implementing each census method in Netherlands and France, respectively, will be briefly discussed in this section.

The Netherlands Case: Virtual Census

In 1971, the Netherlands Statistics gave up on traditional census. After several experimental tests, Virtual Census was implemented in 2001. Unlike Norther European countries where fully dependant on the register-based census, the Netherlands uses the combined-census which is the combination of the field data collection and data from registers. The most reason why Netherlands despaired the fully traditional census was the increases of the rate of non-response. Besides, they realized using the data from registers can save cost and more convenient, but, at that time, Netherlands didn't have enough conditions to switch census method to fully registered-based. Along with the system

condition, there were two reason why traditional censuses were needed. Nordholt (2005) explained that first reason is "to determine the size of the population and to get statistical information about its socio-economic characteristics". Second, the municipal population registers has to be checked and updated. The social-fiscal number was introduced in 1997 and allowed to be used in many administrative data for data linkage. Eventually, in 2001, Census Programme Statistics Netherlands innovated its data collection and data processing methods and named it as "Virtual Census".

The France Case: Rolling Census

As Durr (2005) described several reasons why the national institute of statistics (IN-SEE) initiated a project to redesign the census at the end of the 90s in France, first, keeping the budget of the operation since France hasn't population register, second, the French population is very sensitive to privacy. In the 1970s, Kish (1979) proposed bases of the rotating (or rolling) census as an alternative to the traditional model, for example, executing regular surveys as accumulating information. The method based on Kish's idea was developed by Michel Isnard (1995) with the sampling frame. The rolling census provided more current and timely census data, as well as it spreads the costs and burden of conducting a census evenly over the census cycle (Baffour, King and Valente (2013)). In 2002, the rolling census was enacted in the national census law. France epitomized the rolling census since 2004. Communes are treated differently in relation to the size of the population of the Municipality. The concept of the rolling census is that every commune below 10000 inhabitants will be surveyed once within a five year period, on the other hand, larger communes will be split off into five rotation groups of addresses, and each rotation group being surveyed in one of the five years (2004) (Durr (2005)).

1.3   Data Collection and Data processing Methods

In this section, methods of data collection and data processing for each census method will be investigated and compared with its advantages and drawbacks.

The Netherlands Case: Virtual Census

The Virtual Census was implemented from census program 2001 and used the Social Statistical Database (SSD) as its sources containing a huge amount of data on demographic and social economic issues (Nordholt (2005)). The SSD is constructed by micro-linking several administrative registers and household sample surveys with a new approach of repeated weighting. The reason why Netherlands need traditional censuses is that, first, the size of the population and statistical information about its socio-economic characteristics are still available from the traditional censuses. Second, the municipal population registers are needed to check and update consistently. The 2001 Census program was consisted of many different data sources, such as Population Register (PR),

Jobs register and sample survey for employees, FiBase-register, Labour Force Survey, etc. In this section, we will be mainly focused on Population Register which built from the municipal population registers containing demographic information on inhabitant of the Netherlands and household information, for example, PR was consisted with 16 million records in 2001 (reference period) with various census variables (e.g., Sex, Age, Country of Birth, Country of citizenship, Country of residence a year prior to census, Region of residence, Marital status, Family status, Family situation, Family nucleus type, Private household composition, Household status, Household size, Number of children, Other persons in household (outside the family nucleus). Moreover, the Dutch population registration system has an unique linkage key to each individual which commonly referred as the Social Security-Tax Number (So-Fi number), Record Identification Number (RIN-person) which was used to prevent misuse of So-fi number. Besides, personal identifiers are replaced by age at the reference date and RIN-address for privacy protection of the citizens. Therefore, the Population Register (PR) which is the backbone of the Census can be connected to all social statistical data file and allowed to have multiple *linkage relationship* such as one person can have more than one jobs or several social benefits at the same time, but there would be an issue with mismatches. In order to maximize the number of matches and minimize the number of mismatches, the Dutch Statistical Office designed the matching strategy, and they could achieve to 96-97% matching rate by matching the records of the PR and the Labor Force Survey (LFS) with a linkage key only with one of variables, sex, year of birth, month of birth or day of birth. They also developed micro-integration approach with 9 steps to correct inaccurate data by recognizing the strong and weak points of all data sources used in linkage with PR.

In order to produce overall numerical consistency between all tables, the Statistics Netherlands developed Repeated Weighting approach which is a new estimation method applied in Census in 2001 to ensure the numerically consistent table sets even with different data source. In the Repeated Weighting technique, relevant information is extracted from the register or sample survey as a first step and inserted into a rectangular data block. Then a series of weight is assigned to each data block as a second step. Lastly, each data block is used to produce numerically consistent result tables. But as Nordholt (2005) stressed in the paper, "an important prerequisite in the preparation to repeated weighting estimation is that internal consistency of all the relevant integrated micro-data must be guaranteed."

the feature of individual enumeration is satisfied in this approach as separate information is collected regarding the characteristics of each individual. As regards "universality within a defined territory", this criterion is satisfied as the enumeration is taken from a population register in which the fields for attributes are populated from subsidiary registers relating to specific topics. With regard to "simultaneity", the timing of the census extraction may require careful thought where register update cycles vary. With respect to "periodicity", this approach allows extraction at desired frequencies, includ-

ing "at least once in 10 years" noting again the need to manage the updating cycles for the registers. Finally, in most cases, the requirement in terms of producing small area statistics is largely met, as the information in the registers allows for such aggregates to be generated.

As previous section explained, the Virtual Census satisfies five essential features of the census, (a) Individual enumeration - separate information is collected regarding the characteristics of each individual, (b) Simultaneity, (c) Universality within a defined territory - the enumeration is taken from a population register where subsidiary registers are related to specific topics, (d) Small-area data which can be obtained reliable estimates from combined samples for several survey exercises, (e) Defined periodicity. Furthermore, the Dutch census statistics have both strengths and weaknesses at the same time in terms of quality of the census. Strengths are that (1) statistics by subject of the census are sufficiently integrated into the overall statistics production program of the National Statistical Office in the Netherlands, and (2) the production cycle of policy-related data is more frequent (3) The timeliness of information is improving, (4) there is little burden of response to the population, (5) the cost of conducting a census is very low, and (6) the work of the Dutch National Statistical Office is concentrated at one point every 10 years. (7) Since the specific questionnaire is longer, it is possible to include more survey items so that it is possible to have reliable observation. Weaknesses of the virtual census are that (1) an updated register with high reliability should be secured, (2) attention should be paid to consistency issues since different administrative data are used, and (3) information on sub-regions and subgroups is insufficient, (4) it is necessary to pay attention to the fundamental problems or systematic errors of the data itself by using the registration data extensively, and (5) the concepts and definitions used in the register have some degree of dependence.

Field	Types of Administrative data
population	Population Register, Job Register, Self-Employed Register, Education Register, Income Register, Social Security Register, Unemployment Register, Pension Register, Population, family, household
Housing	Housing Register, Administrative data about Real Estate, buildings, Residence
Economy (Enterprise)	General Business Register, Administrative data about Enterprises or Businesses

Table 1: Types of administrative data for conducting the Dutch Population and Housing Census

The France Case: Rolling Census

The fieldwork in the France is organized each year as the basis of the estimates produced on an annual basis. France is consisted of 36,682 municipalities (Communes) and split into two groups where the threshold of population is 100,000s in size. As Brown and Beaujouan (2013) stated the aim of the sample design approaches to these two groups differently is to cover approximately one seventh of the population in a year with a five year rolling cycle. Therefore, they have two different sampling strategy for smaller and larger communes. It is true that these different sampling approaches require special attention for data collection purposes and strategy.

Brown and Beaujouan (2013) explained each sampling strategy by using the example of the 1995 Census, all the small Communes were split into five rotation groups by using balanced sampling approach such as each group was assigned to one of representations of the populations (geography, age, sex, residential or communal type of dwelling) and a year for data collection on a rolling basis which is covered over a five year period. On the other hand, the larger Communes were split into three layers, for example, Stratum CE, Stratum LA, and Stratum RA contain communal establishment addresses, multi-household addresses and residential addresses, respectively. Any new residential addresses can be place into the fourth stratum NA which is not identify the number of dwellings associated with the address, or the characteristics of any residents (Brown and Beaujouan (2013)). Each Stratum split into five rotation groups and each rotation group was assigned to a year for data collection on a rolling basis which would be covered a five year period. More details of structure of the design within the large Communes can be checked in Figure 1.
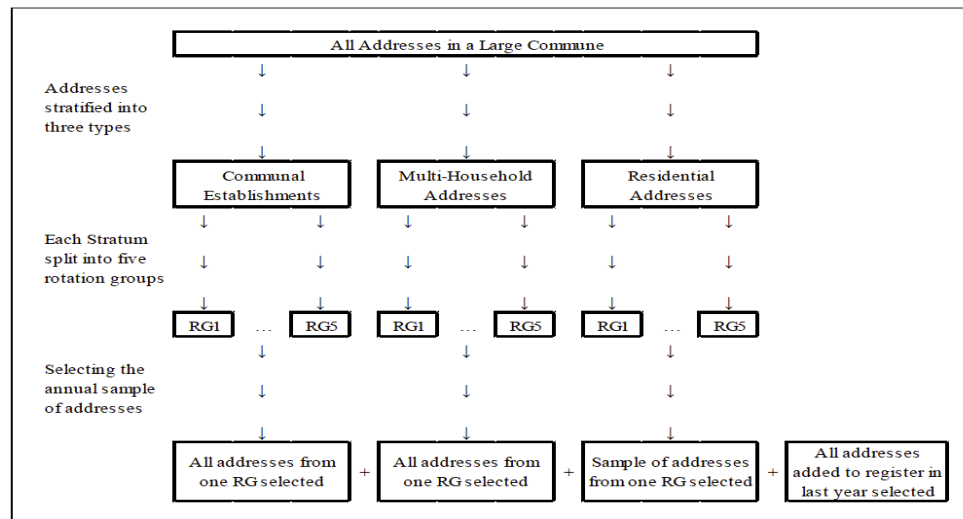

Figure 1: Structure of the design within the large Communes. Reprinted from 'Review of the Rolling Census Approach: and other survey-based options' by Brown and Beaujouan (2013)

As we stated earlier, the larger Communes are sampled every year, these annual fieldwork is organized in similar way as a traditional census which is delivering a form to

the sampled addresses in a specific year by an enumerator with a call-back to collect the form. It is manageable in France because France's localized structure of the enumeration which is managed and organized within each Commune enables to run smoother and feasible annual data collection. This structure in France prevents from an extreme spike in staffing and secures more stable staffing levels leading a possibility of long term reductions in budget. Roux (2020) enunciated that the rolling census has been providing results with very low non-response rate which is under 4% of the households. The localized publicity in Commune allows tighter fieldwork controls and more detailed chasing for the smaller data collection. Besides, when enumerators are unable to achieve the response from a residential household then they fill in a dummy form during field work, then imputation is used in the final outputs. Definitely, there are further investigation for non-response adjustment such as using Permanent Demographic sample or matching to additional administrative sources.

As previous section explained, clearly, the Rolling Census satisfies five essential features of the census, (a) Individual enumeration, (b) Simultaneity, (c) Universality within a defined territory, (d) Small-area data, (e) Defined periodicity. Furthermore, there are some advantages and drawbacks of the Rolling Census. The localized structure of the enumeration and its with the data processing system allow to develop the system over time rather than creating a whole new system every ten years. There has been a few changes of questionnaires and utilization of internet. It is possible to produce results based on the changes on national or regional level from one-year data collection, however, 5 years has to be waited to produce data at all geographical levels related to these new changes due to the feature of the methodology. Roux (2020) indicates that the annual census allows to assess the effects of changes by checking the comparability of the several data collections and the delay of 5 years is an important feature of the rolling census to take into account. On the other hand, the concept of the relevant population of the methodology could be problematic since it implies that not everyone in the country is counted at the same time (Baffour et al. (2013)).

1.4   Conclusion

The evaluation and selection of a census needs to encompass what the exercise has added to the known information about population of a country and the value this has to users. Even though we outlined five essential features of census and six different dimensions that encapsulate the construction of quality, in order to evaluate the census, we need to keep in mind that the historic information about the country's population is the essential component of census evaluation. As a project manager of a census of South Korea, I would prefer "Virtual Census". But the answer could be different if I am the project manager of different country's census. Similar to Netherlands, each Korean also has personal identifier number. In the Netherlands, social security numbers (sofi-numbers) appear in

almost all registers, and they are used to help create census statistics. Probably this identifier number in Korea could be used like Netherlands to integrate all administrative data for census statistics. As Virtual Census requires, an important prerequisite for the preparation of statistics based on administrative data, especially for the implementation of the census, is that the data contained in the administrative data must be extensive and cover many variables related to the relevant unit. In the end, if these data are not sufficient, the information can be supplemented by using the traditional method directly otherwise the statistical content can be bound to be limited. Similar to Netherlands, in Korea's administrative data, a problem will be remained that the degree of inconsistency between the registered address of resident registration and the place of residence and the actual place of residence will be a major obstacle for the Register to replace the Population and Housing Census.

References

Baffour, B., King, T. and Valente, P. (2013), 'The modern census: evolution, examples and evaluation', *International Statistical Review* **81**(3), 407–425.

Brown, J. J. and Beaujouan, E. (2013), 'Review of the rolling census approach: and other survey-based options'.

Durr, J.-M. (2005), 'The french new rolling census', *Statistical Journal of the United Nations Economic Commission for Europe* **22**(1), 3–12.

Nordholt, E. S. (2005), 'The dutch virtual census 2001: A new approach by combining different sources', *Statistical Journal of the United Nations Economic Commission for Europe* **22**(1), 25–37.

Roux, V. (2020), 'The french rolling census: A census that allows a progressive modernization', *Statistical Journal of the IAOS* **36**(1), 125–134.

UNECE (2006), 'Conference of european statisticians recommendations for the 2010 censuses of population and housing.'.