Efficiency in Population Censuses - the situation of the European register-based 2011 Censuses[1]

Eric Schulte Nordholt*
Statistics Netherlands, The Hague, The Netherlands e.schultenordholt@cbs.nl

Abstract

The last Census in the Netherlands based on a complete enumeration was held in 1971. Since the Dutch Census of 1971, the willingness of the population in the Netherlands to participate has decreased tremendously. Statistics Netherlands found an alternative in a Virtual Census, by using available registers and sample surveys as alternative data sources. Advantages of this alternative are that it is much cheaper and more socially acceptable. The combined use of registers and sample surveys for composing the Census however also leads to several methodological challenges. One of them is deciding on the method used to compensate for missing information. Another is the decision on the methods used to combine register and survey data. In this contribution we explore the options chosen by ten (partly) register-based European countries. Next to the Netherlands six fully register-based countries (Norway, Denmark, Sweden, Finland, Austria, and Slovenia) and three partly register-based countries (Switzerland, Germany, and Poland) were asked a number of questions. We are thankful that all countries approached were willing to help us in this research project. This research has been beneficial to Statistics Netherlands for making final decisions in the approach of the Census of 2011. Hopefully, it is also useful for other countries that are working or planning to work with registers in their censuses.

Keywords: Europe, registers, sample surveys, Virtual Census

1. Introduction

All European Union (EU) countries conduct a Census of 2011. How this Census is conducted is up to the countries. In the Netherlands virtual censuses are held ever since the last traditional Census in 1971. This means that census forms no longer exist and that the relevant information is provided by data in already existing registers and sample surveys (Schulte Nordholt, 2004). This approach was used for the Virtual Censuses of 1981, 1991, and 2001. The Censuses of 1981 and 1991 were of a limited character. The data compiled on 1981 and 1991 were much less detailed than the set of tables of the 2001 Census. For the 2001 Census Statistics Netherlands published information on the municipal level. For the 2011 Census more registers and fewer surveys have been combined. The Population Register forms the backbone for the integration activities that will result in coherent and detailed demographic and socio-economic statistical information on persons and households.

A generic problem in using administrative registers for statistical purposes is that the data in these sources are collected and maintained by other organizations for non-statistical purposes. The process is beyond the control of Statistics Netherlands. This not only makes Statistics Netherlands highly dependent, it may also affect the quality of the output of Statistics Netherlands (Schulte Nordholt, Ossen, and Daas, 2012).

European countries that use registers to provide data for the Population and Housing Census of 2011 usually cannot obtain all required census information from the available registers. Registers are usually not set up for statistical purposes, and

[1] The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

therefore register data may differ in content and definitions required for the Population and Housing Census of 2011. Several approaches may be used to obtain the missing information (e.g. to conduct new fieldwork or to recycle already existing sample surveys) or it may even be decided to not provide these data to the European statistical office Eurostat.
Statistics Netherlands has conducted a research project in order to compare the ways fully register-based and partly register-based countries deal with this missing data problem for the Population and Housing Census of 2011. Statistics Netherlands uses available sample surveys to obtain missing information, and therefore it was investigated whether other countries also make use of sample survey information for the Population and Housing Census of 2011. One of the goals of this research project was to compare the different estimation methods used to combine register and survey data. Several estimation methods can be used to raise the survey data to the population totals (e.g. imputation or weighting), but not all estimation methods will automatically produce a set of numerically consistent census tables (i.e. when tables with common margins are estimated from different sources the estimated margins differ). Since production of a set of numerically consistent census tables is required in the EU regulation (European Commission, 2008), Statistics Netherlands will use the estimation method of repeated weighting (Houbiers et al., 2003) for the Population and Housing Census of 2011. This method guarantees production of a set of numerically consistent tables. More information about the Dutch methodology used for estimating Census tables based on incomplete information can be found in Schulte Nordholt (2012).
Six fully register-based countries (Norway, Denmark, Sweden, Finland, Austria, and Slovenia) and three partly register-based countries (Switzerland, Germany, and Poland) were asked for participation in the research project. Together with the Netherlands, ten countries participated in this research project. All countries were asked about their specific situation on missing census information and the approach that will be used to obtain the missing information. In addition, countries were asked whether sample surveys will be used for the Population and Housing Census of 2011 and, if applicable, which method will be used to combine the register and survey data. In order to compare the approaches used by all participating countries to obtain missing information, the variable “current activity status” is used as an example, since this variable is usually difficult to derive from registers.
It should be mentioned that our research yielded information with a variety in content and length per country, since answers provided by the countries differed in their specificity and comprehensiveness. Where some countries provided an extensive description of the amount of missing information, other countries only provided one or more examples of variables with missing information. Moreover, some countries were very clear and honest about the fact that some of the data will not be provided completely in accordance with the definitions as specified in the EU regulations, while other countries did not provide any information about the way data will be provided. The information in this paper is based on the information provided by the countries, sometimes complemented with information from literature. This should be taken into account while comparing the approaches used by the participating countries. A general comparison of the methods used to compensate for missing information and for combining register and survey data can be found in sections 2 and 3. Some conclusions are drawn in section 4. This contribution could be considered as an updated and shortened version of Maris, Schulte Nordholt, and Van Zeijl (2012). A more detailed description of the approaches of the individual countries that participated has been published in Maris (2012).

2. Comparison of Census 2011 methods used to compensate for missing information

The methods used by six fully register-based countries (Norway, Denmark, Sweden,

Finland, Austria, and Slovenia) and three partly register-based countries (Switzerland, Germany, and Poland) to compensate for information that is not available from registers are compared. Moreover, these methods used are compared with the method of Statistics Netherlands. All fully register-based countries can make use of a lot of registers, such that most of the census variables can be obtained by using register data only. However, all fully register-based countries experience problems with obtaining data that is not available from registers. Several approaches will be used for the Population and Housing Census of 2011 to deal with this problem. For example, Statistics Austria and the Statistical Office of the Republic of Slovenia will impute all missing information, whereas Statistics Norway and Statistics Sweden will impute some of the missing information. Moreover, Statistics Norway and Statistics Sweden will not be able to specify some categories of the required variables and will therefore assign some persons to a different category. Furthermore, Statistics Finland will derive all missing variables by using a so-called 'register estimation method'. In addition, Statistics Denmark, the three partly register-based countries involved (Switzerland, Germany, and Poland) and Statistics Netherlands will use additional sample information (e.g. from the Labour Force Survey) to obtain all required census information.

Since most of the fully register-based countries use register data only, wrongly classified persons and categories that cannot be specified are unavoidable. This implies that some of the data will be sent to Eurostat not completely in accordance with the definitions as specified in the EU regulations. For example, most of the fully register-based countries will not be able to provide information on unemployed persons in the way it is required according to the EU regulations, since information on persons available for work and persons who are looking for work is usually not available in registers. Moreover, administrative registrations from unemployment benefit agencies and social assistance benefit agencies only cover registered unemployed persons, but there may also be unemployed persons who are not officially registered as being unemployed and who also do not receive any unemployment or social assistance benefits. Furthermore, most of the fully register-based countries have no information on persons who have completed education abroad, such that the variables "current activity status" and "educational attainment" are difficult to obtain completely in accordance with the definitions required.

Thus, although several methods will be used to deal with the missing data problem, most of the fully register-based countries will not be able to send all required census variables to Eurostat completely in accordance with their definitions as specified in the EU regulations. Where the fully register-based countries can obtain most of the census variables from registers and will provide some of the data not completely in accordance with the definitions, the three partly register-based countries included in this research project will conduct surveys for their Censuses of 2011 to obtain additional information since they can only obtain some of the variables from registers. For Statistics Netherlands, most of the variables can be obtained from registers, but there is no register information on occupation available. Moreover, there is not much register information on educational attainment for some of the residents (e.g. persons older than 45 years). Therefore, information on occupation and educational attainment must be obtained from sample surveys. In this way, the method used by Statistics Netherlands is comparable to the approaches used by the three partly register-based countries involved. However, Statistics Netherlands uses existing sample survey data on occupation and educational attainment, whereas the three partly register-based countries conduct (ad hoc) surveys for their Censuses of 2011 to obtain this information. On the other hand, Statistics Netherlands has a lot of register information and it is possible for Statistics Netherlands to derive the variable "current activity status" from register data only. In this way, the method used by Statistics Netherlands is comparable to the approaches used by fully register-based countries. Moreover, the costs of the Dutch Population and Housing Census of 2011 are comparable to the costs

of a fully register-based census (United Nations, 2010). Thus, the method used by Statistics Netherlands corresponds to a combination of a fully register-based approach and a partly register-based approach.

3. Comparison of Census 2011 methods used to combine register and survey data

In this section it described whether the participating countries make use of survey data to obtain information that is not available from registers and which methods are used to combine the register and survey data. All three partly register-based countries in this research project (Switzerland, Germany, and Poland), Statistics Denmark and Statistics Netherlands make use of survey data to obtain information missing in registers. Although Statistics Denmark is a fully register-based country, existing sample survey information will be used to obtain information on educational attainment for immigrants. Moreover, Statistics Netherlands will use the Labour Force Survey to obtain information on occupation and educational attainment. Where Statistics Denmark and Statistics Netherlands make use of available sample survey data, the three partly register-based countries involved have conducted (ad hoc) surveys for the Population and Housing Census of 2011. For example, the Federal Statistical Office of Germany has conducted several surveys to obtain information on, for instance, education, employment, and housing, whereas the Federal Statistical Office of Switzerland has conducted several surveys to obtain information on, for instance, households, families, housing, employment, and education. Furthermore, the fully register-based countries Norway, Sweden, and Austria will conduct sample surveys in order to assess the quality of the variables derived, but these sample survey data will not be used to complement the data obtained for their Population and Housing Censuses of 2011.

All three partly register-based countries in this research project will use traditional weighting methods to raise the survey data to the population totals. For both the Swiss and the Polish Population and Housing Census of 2011, the survey data will be weighted and calibrated with the use of register data. Survey data used for the German Population and Housing Census of 2011 will be calibrated with the use of the GREG-estimator, after which the number of German residents estimated from the survey data will be used to correct the number of German residents obtained from registers. Moreover, Statistics Netherlands will use the method of repeated weighting to raise the survey data to the population totals. Furthermore, Statistics Denmark will impute the non-response in the survey on the basis of register data, after which the survey data will be integrated with the register data.

The results of this research project can be used to compare different census methods for the Population and Housing Census of 2011 and to learn from experiences of other countries. There is a variety of missing information among the countries and several approaches will be used to obtain the missing information. Since all countries deal with different situations of missing information (e.g. only a small amount of missing information, problems with consistency, quality problems) several methods will be used for these different situations. This research project can be helpful to compare the methods used by other countries in a specific situation and to investigate the possibility of using a similar method for this situation. For example, when countries deal with missing data for some of the variables, it may be considered to impute all missing information. This is most frequently done when dealing with a small amount of missing information. Moreover, many countries experience difficulties with obtaining information on educational attainment for immigrants. These countries may consider using the method of Statistics Denmark to use existing sample survey information for the variable “educational attainment”. In addition, most of the countries will not be able to derive the variable “current activity status” completely in accordance with the definition as specified in the EU regulation (European Commission, 2009) and therefore other countries may also consider to derive this

variable not completely in accordance with the required EU definition. This was especially important for Statistics Netherlands, as Statistics Netherlands has since this finding decided to derive the variable "current activity status" based on registers only for the Dutch Population and Housing Census of 2011. Furthermore, it may also be investigated whether the variable "current activity status" can be derived by using an approach similar to the 'register estimation method' used by Statistics Finland or whether this method can be used for deriving variables that cannot be obtained directly from any register.

In addition, countries that have problems with consistency may investigate the possibility to use the method of repeated weighting. For example, the partly register-based countries may use the method of repeated weighting in order to obtain a set of numerically consistent census tables. However, it should be taken into account that by using repeated weighting several estimation problems may occur. For example, sampling zeros cause estimation problems such that it is sometimes not possible to estimate the required census tables completely. Statistics Netherlands therefore plans to use the technique of macro-integration in addition to repeated weighting to estimate the more detailed Census 2011 tables (Mushkudiani, Daalmans, and Pannekoek, 2012).

Although countries that do not make use of sample survey data for the Population and Housing Census of 2011 will probably have fewer problems with consistency, the method of repeated weighting may still be useful when dealing with quality problems. For example, it may be considered to use existing sample survey data for variables that are not of good quality and to apply the method of repeated weighting to combine the register and survey data. In addition, fully register-based countries that conduct quality assessment surveys may consider using the survey information obtained from this quality survey and to use the method of repeated weighting.

4. Conclusions

The Virtual Census has proved to be a successful concept in the Netherlands. It has many advantages compared to traditional censuses. The costs are now considerably lower. Still, census data on the Netherlands can be compared to results of earlier Dutch censuses and to the results of other countries that take part in the same Census Round. So far the Netherlands has conducted Virtual Censuses of 1981, 1991, and 2001. Also for the Virtual Census of 2011 it is important that the final results are comparable both over time and with other countries. Therefore, the quality of the Dutch registers used is of vital importance for the 2011 Census.

It is possible to conduct a register-based census in more and more countries. Although, in most countries not all census variables can be derived from register information. For those variables additional surveys remain a necessity. To be able to use registers for statistical purposes, it should be possible to determine the quality of these registers. In the year after this research project it has been decided how the different variables for the Dutch Census of 2011 had to be derived.

An advantage of the method used for the construction of the Virtual Census (Schulte Nordholt, 2004) is the use of micro-integration. In this way data are checked and incorrect data are adapted. The number of measurement errors thus decreases. By the introduction of the method of repeated weighting the remaining inconsistencies are solved. Given the detailed information requests of the 2011 Census, the large number of available sources for the Dutch Census, and this research project comparing approaches of different (partly) register-based countries, it is sure that we will have a lot of interesting experiences with our register-based 2011 Census in the coming years that will draw the attention of many other countries.

Thus, although all countries deal with different situations of missing information and different amounts of available register information, this research project can be helpful to learn from approaches used by other countries in a specific situation, not only for the Population and Housing Census of 2011 but also for future censuses.

References

European Commission, 2008. Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses. *Official Journal of the European Union*, L218, pp. 14-20.

European Commission, 2009. Commission Regulation (EC) No 1201/2009 of 30 November 2009 implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdown. *Official Journal of the European Union*, L329, pp. 29-68.

Houbiers, M., P. Knottnerus, A.H. Kroese, R.H. Renssen, and V. Snijders, 2003. Estimating consistent table sets: position paper on repeated weighting. Discussion paper 03005, Statistics Netherlands, Voorburg / Heerlen. http://www.cbs.nl/NR/rdonlyres/6C31D31C-831F-41E5-8A94-7F321297ADB8/0/discussionpaper03005.pdf

Maris, A.M., 2012. Construction of numerically consistent tables from population and housing census data. Research report, Statistics Netherlands, April 2012, The Hague / Heerlen.

Maris, A.M., E. Schulte Nordholt, and J. van Zeijl, 2012. Comparing approaches of different (partly) register-based countries. Paper presented at the UNECE-Eurostat Meetings on Population and Housing Censuses, 22-25 May 2012. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_3_Netherlands.pdf

Mushkudiani, N., J. Daalmans, and J. Pannekoek, 2012. Macro-integration techniques with applications to census tables and labour market statistics. Discussion paper 201201, Statistics Netherlands, The Hague / Heerlen. http://www.cbs.nl/NR/rdonlyres/AD653253-647D-4FFD-AFC4-67E2BDE602EE/0/201201x10pub.pdf

Schulte Nordholt, E., 2004. Introduction to the Dutch Virtual Census of 2001. *The Dutch Virtual Census of 2001, analysis and methodology*, eds. E. Schulte Nordholt, M. Hartgers, and R. Gircour, Statistics Netherlands, Voorburg, pp. 9-22.

Schulte Nordholt, E., 2012. Methodology used for estimating Census tables based on incomplete information. Paper presented at the UNECE-Eurostat Meetings on Population and Housing Censuses, 22-25 May 2012. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_14_Netherlands.pdf

Schulte Nordholt, E., S. Ossen, and P. Daas, 2012. Quality of registers used for the Dutch census. Paper presented at the UNECE-Eurostat Meetings on Population and Housing Censuses, 22-25 May 2012. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2012/use_of_register/WP_8_Netherlands.pdf

United Nations, 2010. Main Results of the UNECE/UNSD Survey on the 2010/2011 Round of Censuses in the UNECE Region. Paper presented by Paolo Valente at the Working Group on Demography and Census on 19-20 April 2010 in Luxembourg.