

Seminar on Registers in Statistics - methodology and quality

21 - 23 May, 2007 Helsinki

The first register-based Census in Norway in 2011: How to comply with international recommendations?

Espen Andersen and Harald Utne
Statistics Norway

espen.andersen@ssb.no and harald.utne@ssb.no

Chapter 1. Introduction

The last fully traditional population and housing census in Norway was conducted in 1960. The censuses from 1970 to 2001 were based on a mixed mode data collection, combining register data and traditional censuses. The use of register data increased over the years, and in 2001 only information on households and dwellings was collected from census forms. The 2011 Census in Norway will according to the plans be totally register-based. That means that all data will be collected from administrative registers and, at least in principle, no additional data collection will be necessary.

From the 1960 round of censuses, regional recommendations have been adopted by the Conference of European Statisticians (CES). The current recommendation (2010 Census round) is based on an out-put oriented definition of censuses, as opposed to the classical approach with focus on data collection methods. The recommendation is more specific on the conceptual framework and definitions, for instance on place of usual residence and housing arrangements.

Furthermore, the European Union is about to adopt a regulation on population and housing censuses¹. The regulation is based on the CES recommendations. However, replacing the former “gentlemen’s agreement” by a European law means that the importance of fulfilling all requirements for census data is more important than before.

The challenge for Statistics Norway is how to comply with international census requirements by using administrative data only. In this paper we will discuss some of the problems we are facing and describe some necessary improvements towards 2011.

Chapter 2. A brief history of the use of registers in Norwegian censuses

The first Norwegian population census dates back more than 200 years. Up until the 1960s, all censuses were traditional. In 1964 the Central Population Register (CPR) was established, and an official, personal identification number was introduced. The 1960 Census provided the basis for allocating the personal identification number. In 1970 the census was used to establish the Register of Education. In the 1980 Census, information on demography, education and income, as well as geographical characteristics, was collected from registers. Administrative registers were also used to rationalise data collection, and data capture by postal mail only was introduced. In the 2001 Census all data on persons (place of residence, demography, education, labour market, income etc.) were based entirely on registers. Data on families, households and housing was based on a

¹ The regulation is EEA relevant.

full coverage survey with a questionnaire for every household. Prior to the census a dwelling register had been established as an extension of the Register of Ground properties, Addresses and Buildings (GAB register). Data from census forms was used to establish links between resident persons and their dwellings (in the CPR) and to update information on dwellings in the GAB register. The census in 2011 will be totally register-based. Data on households, families and couples will be collected from the CPR and data on housing from the GAB register.

There are two main trends in the development of register-based censuses in Norway. The *first* one is that censuses have been used quite extensively in establishing registers. The *second* trend is the step-by-step development. In the first steps subject matter statistics are tested and published in different areas. Register-based variables are introduced in the census as soon as the quality is considered sufficient. The time elapsing from administrative registers have been established till the data are satisfying for census purposes have varied from one subject matter to another. When statistics has been developed for all areas relevant for censuses, a totally register-based census can be conducted. These same trends have been observed in other countries developing register-based censuses.

Chapter 3. Register-based censuses: the overall picture

Chapter 3.1 Essential features of a census

The CES recommendation gives every country freedom to decide on data collection methods as long as the data collected fulfill the international requirements for censuses. Complying with the essential features of population and housing censuses is of course the basic issue.

Individual enumeration: A register is defined as a systematic collection of unit-level data organised in such a way that updating is possible. The units in registers used in Norwegian censuses are resident persons and dwellings, so the condition of individual enumeration is met.

Simultaneity: Since all base registers are updated regularly, all units can be enumerated at the census day (in 2001 3 November). For labour market characteristics the reference period is the census week (week covering census day), for income characteristics we use the census year and for the rest of the characteristics the census day. In the 2001 Census a minor deviation for educational topics was accepted (1 October) and this will probably be the case also in the next census.

Universality: The census covers the entire country. The population census includes all persons registered as residents in Norway on the census day that is the *de jure* population. That means that for instance illegal immigrants will not be covered (see chapter 4). The housing census includes all conventional dwellings and institutions, but non-conventional dwellings are not included. The number of units not covered by using registers is very small in Norway and there would be problems to enumerate these groups also when conducting a traditional census.

Small area data: Registers-based statistics are covering the whole population and are well suited for producing small area statistics.

Defined periodicity: Most of the census variables will be published annually. Complete censuses will be conducted every ten years.

The main conclusion is that Norwegian register-based censuses will comply with the essential feature to a high degree.

Chapter 3.2 Advantages and disadvantages of register-based censuses

Advantages

The main advantages of register-based censuses are lower costs and reduced response burden. A register-based census is based on the same sources as annual subject matter statistics.

Consequently, the marginal costs of producing a census are low compared to a traditional census. The reduction of response burden is substantial since a census is covering the whole population.

Another advantage of a register-based approach is that census statistics and annual statistics are directly comparable, or to put it another way: Census type of statistics can be produced currently, for most variables annually. This has been considered a major advantage by many users.

Disadvantages

In a register-based census only variables that are included in registers or could be derived from register variables, are available. This is discussed in more detail in chapter 4. In many countries the population census is an important tool for data collection that is used to meet emerging information needs. This flexibility is lost when data are no longer collected by means of questionnaires. For example disability topics, as recommended for the 2011 Census, are not obtainable from administrative sources in Norway. Consequently, this type of information will only be available from sample surveys.

Additionally, some restrictions are imposed on the definitions of units and variables. In a register-based census a private household is defined as all persons living in the same dwelling (*dwelling household*). *Housekeeping units* (persons living in the same dwelling with joint board) cannot be included. The population is normally counted according to *legal* place of residence (according to the population register) and not *de facto* place of residence. For some users, housekeeping units based on *de facto* place of residence would be more appropriate for their analyses.

Chapter 4. Census topics: Possibilities and challenges when using registers

Chapter 4.1 Main quality issues

The CES Recommendations identifies the following dimensions of quality in statistics: Relevance, accuracy, timeliness, accessibility interpretability and coherence. In this paper we will mainly discuss the dimensions of relevance and accuracy when using administrative registers as the source for census data.

When conducting a traditional census, using census forms, the statistical office may ask questions directly corresponding to the definitions given in international recommendations. In this way the census data will relate to the concepts that it is desired to describe, in other words the *relevance* is high. When using register-based data it is always important to bear in mind that the data is originally collected for another purpose than making statistics. This may pose some restrictions on definitions (see chapter 4.2).

To obtain a high degree of *accuracy* is a challenge independent of data collection methods use, but the type of problems may be somewhat different when using register-based data. Topics that are important for the administrative agencies responsible for the registers are normally recorded with a high degree of accuracy.

In general, problems concerning relevance may be most difficult to solve when conducting a register-based census.

Chapter 4.2 Enumeration by using the Central Population Register

The population register and the dwelling register constitute the back-bone of a register-based population and housing census. The fact that we use the Central Population Register (CPR) to enumerate the resident population has a great impact on most of the census topics.

Place of usual residence

Place of usual residence defines the resident population of each country and each region within the country. According to the CES Recommendations *place of usual residence* may be either the place of living at the time of census (de facto address) or the legal residence. In a register-based census the legal residence would be the most obvious option. For most people the legal residential address are the same as the de facto address, but for some groups there may be deviations.

According to the CES Recommendations only persons who have lived in their place of usual residence for a continuous period of at least twelve months or if not, have the intension of staying there for at least a year, should be regarded as usual resident. In the instructions for updating the Norwegian CPR there are similar requirements with regard to where a person spends most of his/her daily rest. However, the corresponding rule of duration is six months. This of course has some consequences for the count of the resident population in Norway.

Furthermore, there are some administrative rules and practices that do not fully comply with the CES Recommendations. Students may, according to instructions for population registration, choose where they want to be registered. It is rather common for students to be registered at their parents' home address even if they are actually living in a different household, and in fact most students do so. According to the CES recommendations it is acceptable to consider students family home as their usual residence, but this is not the case for persons studying abroad.

There are similar types of deviations when it comes to spouses that actually do not spend most of their daily night-rest in the same dwelling. According to the CPR instructions spouses should normally be registered at the same address in the CPR. This rule will have an effect in situations where one spouse lives in an institution or works in a different country or other parts of Norway.

Using the concept of legal residence of course implies that persons staying illegal in the country will not be counted. However, this group will be difficult to count precisely in traditional censuses as well. But even some persons staying legally in the country are not registered as residents in the CPR, for instance most asylum seekers.

In Norway it will be possible to use additional data sources (registers) to give a better picture of the de facto resident population. This information will however, only be available for certain groups of the population. A project is on-going in Statistics Norway with the objective to produce data on de facto residence for some groups, mainly students and persons resident in homes for elderly. By using other administrative data sources in addition to the CPR, it will also be possible to include asylum seekers and exclude persons studying abroad from the resident population.

Since data from additional sources are substantially delayed compared to data from CPR, the annual and quarterly population statistics will have to be based on the CPR only. Using other sources in the census therefore will give population count different from the annual statistics. This means that improving the relevance in this case will have a negative impact on comparability between statistics.

Household topics and housing arrangements

Using legal resident address has consequences not only for the topic *total population*, but also for household and housing topics. In Norway households with students are most influenced by this fact. Compared to using de facto address there will be more households with adult children and less one-person households. Especially when tabulating household income, these differences are clearly demonstrated.

In the Norwegian CPR most people are registered as living in conventional dwelling. Only very few persons are registered as *homeless* (with “no fixed abode”) in the CPR. Most de facto homeless persons is in fact registered at a legal address, for instance with friends or relatives. They may have some sort of connection to this household, but they do not spend most of their daily rest there. Statistics Norway has recently started an investigation with the objective to find better ways of producing data for homeless persons.

The group *living in other housing units* (that is other than conventional dwellings) is almost non-existing when using CPR as the only source. No persons are registered as living in for instance boats, cabins, caravans and tents. It would be possible to register people in second homes (for instance summer cottage), but this will require changes in the rules for registration. Cottages are registered as buildings in the GAB register, which is of course not the case for tents, boats and so on.

When using legal address almost all persons will be classified as occupants of conventional dwellings even if they de facto are living in other types of housing arrangements.

Chapter 4.3 Quality of census topics

Table 1. Core and non-core topics by availability in 2006 and estimated availability in 2011.

	Total	Available 2006			Available with improvements		
		No	Partly	Completely	No	Partly	Completely
Core topics, total	38	2	17	19	0	7	31
Population census	24	0	9	15	0	3	21
Housing census	14	2	8	4	0	4	10
Non-core topics, total	71	43	10	18	31	16	24
Population census	54	30	6	18	21	10	23
Housing census	17	13	4	0	10	6	1

Core topics

Table 1 shows that in the current situation (2006) half of the core topics are completely covered. Only two topics are not available at all (*occupancy status of conventional dwelling* and *water supply system*). The rest of the core topics are describes as “partly available”. This means for instance that we have information on some, but not all values, that the information do not cover the whole population or that the quality in general is not good enough. Some examples of these problems are given below. The current situation for the population census is better than for the housing census. The main reason for this is that the dwelling register was established only five years ago and that the quality has not yet reached an acceptable level for the majority of variables.

By 2011 Statistics Norway should be able to produce all core topics. To achieve this goal, several activities improving data quality must be carried out, mainly by the register owners, but also by Statistics Norway. These improvements are described in some more detail in chapter 5.

The estimate is that all core topics will be available, but seven topics only partly. For some topics the problems are more serious than for others. For instance, we only expect very limited information on *water supply*. For *ever resided abroad* the problem is limited to lack of information for persons living abroad before 1964.

Non-core topics

For non-core topics, the current situation is far less bright. Only 25 per cent of the topics are fully available and 60 per cents are not available at all. Even with the improvements planned, more than 40 per cent will not be available in 2011.

Table 2. Non-core topics by relevance, availability in 2006 and estimated availability in 2011.

	Total	Available 2006			Available with improvements		
		No	Partly	Completely	No	Partly	Completely
Non-core topics, total	71	43	10	18	31	16	24
High relevance	24	10	3	11	5	5	14
Medium relevance	25	15	5	5	8	9	8
Low relevance	17	13	2	2	13	2	2
No relevance	5	5	0	0	5	0	0

According to the CES Recommendation, decisions on producing non-core topics should mainly be based on national needs. In table 2 non-core topics are classified by relevance (that is, expected importance for Norwegian census statistics²). Not surprisingly the coverage is best for the most relevant topics. In planning improvements we are mainly concentrating on high relevance topics. Topics that are labelled medium relevance will be included only if a limited amount of resources is needed to produce them. Variables with low relevance (for instance *hot water*), will only be included if no extra work is needed. The topics classified as not relevant are either agricultural topics or topics not relevant due to the definitions used (for instance *single or shared occupancy* is not relevant for countries using the dwelling household concept).

Topics difficult to cover in a register-based census, some examples

As mentioned in chapter 2 it normally takes some years from administrative registers are being established till the data are satisfying for census purposes. This is the main reason for the quality problems we are currently facing for data on dwellings and buildings. The GAB register is the source for this kind of information. The 2001 Census was used to extend the GAB-register and to collect additional information on dwellings built before 1983. The GAB-register is administrated by the Norwegian Mapping Authorities, but the municipalities are responsible for input and updating. Variables that are not important for administrative purposes, or are mandatory to up-date, have in general a poorer quality than other variables. Examples are: *toilet and bathing facilities, number of rooms and type of heating*. For these variables the coverage is currently not more than 80 to 85 per cent. The information on *water supply* is more or less non-existing.

Prior to the 2001 Census a dwelling register had been established as an extension of the GAB register. The introduction of a unique address³ to all dwellings made it possible to establish links between dwellings and resident persons in the CPR with the address as the key variable. Today 83 percent of the residents living in multi-dwelling buildings are registered with a unique address in the CPR. The main reason for this low percentage is the poor coverage in the capital Oslo. In Oslo the percentage is as low as 68. This of course also influences on the quality of the

² The classification is done by relevance as census topics. Variables that are not considered relevant in a census may of course be important in other statistics.

³ For an address to be unique in a multi-dwelling building it must comprise a dwelling number in addition to the street address.

household variables. To improve the quality, the 2001 Census is used as an additional information source (see chapter 5).

A national register of employees was established in 1978, but *occupation* was not included in the register before 2001. *Occupation* was planned as a register-based topic in the 2001 Census, but the coverage was too low and it was not possible to include information on occupation in the census. The coverage has improved, but still *occupation* could only be published for approximately 70 per cent of all employed persons.

Another familiar problem is the lack of data describing the situation before the register in question was established. In the census this is in particular associated with topics like *ever resided abroad* and *place of birth*. In Norway we do not have this kind of information for the period before 1964 when the CPR was established. As mentioned above we had much of the same situation concerning buildings and dwellings built before the GAB register was established (1983), but the 2001 Census was used to update the register. This type of imperfections is gradually getting less important as years goes by.

Some variables are not available directly from any register, but could be derived by combining different kinds of register information. In Norway there exists no “cohabiting register”. To determine whether two persons living in the same household are cohabitants or not, we use information on the resident address, whether the persons have a child in common, if the persons are relatives, the mutual age difference and the date of moving into the dwelling. This method makes it possible to identify cohabiting couples of opposite sex in a reliable way. However the method does not allow for identifying same sex cohabiting couples⁴. This example shows one of the problems of identifying relationships between members of a household. We do not have a register that cover people’s emotional relationship to each other.

A problem area is education taken by immigrants before they entered Norway. In 1990 and 1999 Statistics Norway carried out surveys among immigrants asking about their education. The register-based information in this field is still not satisfactory. To have this information updated for the next census, a new survey is necessary for immigrants that have entered the country after 1999.

It is not possible, and in some cases not even desirable for the society, to have administrative registers covering all topics relevant for statistical purposes. We have already mentioned the example of *cohabitation* and other examples are *homemakers* and *family workers without contract of employment*. For this type of variables we do however have some relevant register information, and imputation or estimation of values will, at least in principle, be possible. Other topics are related to human behavior and attitudes, examples being *mode of transport to work/school*, *ethnicity*, and *type of energy used for heating*. This kind of topics will probably never be covered by register data, and combining data from sample surveys and registers will be the only option.

Chapter 5. Necessary improvements towards 2011

As seen in chapter 4, a register-based census in Norway can at present not fulfil all international requirements. To be able to produce a complete set of census data in 2011, several improvements will have to be made. The most important task is of course to improve the quality of data as recorded in administrative registers. This is the responsibility of the register owner, but Statistics Norway also plays an important role both formally (with reference to the Statistics Act) and practically by everyday cooperation. Furthermore, the processing of data in Statistics Norway, transforming administrative data into statistical data, must be improved in several ways. New derived variables will be created by combining data from several sources. Quality

⁴ Same sex registered partners are identified in the CPR

assessments are another important part of the census project. Especially for topics that have not been register-based in previous censuses, this is a major task.

It must be emphasized that the improvements necessary for census purposes are equally important for subject matter statistics. This means that most of the necessary activities are not organized as a part of the census project in Statistics Norway, but are carried out by the units responsible for statistical base registers and the subject matter statistics. The census project has mainly a coordinating role at this stage.

Improving the quality of administrative data - the role of Statistics Norway

As described above, the situation for data on housing and households are less satisfactory than for the rest of the census data. Improving the relevant data in the CPR and the GAB register is a main condition for a successful census in 2011. Basically there are quality problems in two fields: The dwelling variables in the GAB register and the link between dwellings (in the GAB register) and resident persons (in the CPR). We will use the last problem as an example.

The links are not satisfactory for persons living in multi-dwelling buildings. What can Statistics Norway do to improve this data? The Dwelling Address Project was established in 2000 as a co-operation between the tax authorities (responsible for the CPR), the mapping authorities (responsible for the GAB-register) and Statistics Norway. The project as such is finished, but there are still several following-up activities going on and Statistics Norway still plays an important role.

The first task is the quality monitoring. Every quarter Statistics Norway produces tables that describe the quality. The most important are: The proportion of persons living at a multi-dwelling address with a unique address in the CPR and the proportion of removal notices registered with a unique address for persons moving into a multi-dwelling building. Both tables are produced at a municipality level.

Statistics Norway also plays an important role as a coordinator and an initiative-taker. The greatest needs for quality improvement are in the municipality of Oslo. Several activities are carried out by different partners. For instance, the cooperative building association in Oslo (OBOS) has become part of the project. There are also plans to send questionnaires to every household registered on a multi-dwelling address, but without a unique address in the CPR. This is a rather extensive project that provides funding from the Ministry of Finance. In addition to this we also encourage different firms and organisations such as telecom companies, cable TV operators, insurance companies, municipalities (social services), ambulance operators and the fire brigades to take the dwelling number into use. The Norwegian Post has so far not been willing to include the dwelling number in their postal addresses. The idea is that increased use of the dwelling number will make citizens more aware of its existence and this will in the long run improve the quality in the register data. Statistics Norway plays an active part in all these activities.

Developing register-based statistics for families, households and dwellings

In spite of the quality problems mentioned, Statistics Norway is using register-based data on families, households and dwellings for producing statistics. This means that we have to develop methods to handle the quality problems.

A complete register-based statistics on families and households was published for the first time for the year 2005. Two methods were taken into consideration. The first method was to use all existing dwelling addresses as a basis, even if some of them are not unique (dwelling numbers missing). The result would be too few and too large households. The second method was to update all households established in the 2001 Census (based on census forms) with register information on births, deaths, removals, weddings, divorces and so on. This method would give too many and too small households. The chosen method was actually to combine these two

approaches, and the result was reliable data for all parts of the country except for Oslo. To overcome the quality problems in Oslo, it was necessary to correct the statistics by weighting. As the quality of administrative data is improving, this kind of corrections could hopefully be avoided.

A register-based dwelling statistics was first published for the year 2006. At present only the total number of dwellings is published (not separating occupied from vacant dwellings) and the distributions are by *type of building* and *period of construction* only. For the remaining dwelling topics, the coverage is at present too low for publishing statistics. Using legal address (see chapter 4.2) and considering the quality problems mentioned above, there are also some problems to classify dwellings as occupied or vacant. The dwelling statistics will be gradually extended as the data quality is improving.

As mentioned above the households statistics is not based on a register count only. This means that Statistics Norway will have to develop a method for harmonization between household data and data on dwellings. According to the definitions, the number of (dwelling) households must be the same as the number of occupied dwellings.

Improved and more extensive use of existing data

In order to produce the desired census topics in 2011, it will be necessary for Statistics Norway to combine existing data in new ways. In planning census activities, the census core topics will be given top priority. Examples of such topics are *current activity status*, *status in employment* and *type of family nucleus* (reconstituted families). For other topics it will probably be necessary to look for additional register data, examples being *tenure status* and *occupancy status*.

For non-core topics the priority will be given according to national relevance and of course the availability of reliable data. Examples of suggested topics are *usual activity status*, *main source of livelihood* and *distance travelled to work*. A plan for ranking of priorities will be finished by the end of 2007.

Use of additional data in the 2011 Census

For the 2011 Census, as for previous censuses, Statistics Norway is planning for a quality assurance programme. This program will comprise quality monitoring in all parts of the process. One important part is a quality assessment of census data, comparing register-based data with data from other sources. As in the 2001 Census we will use data from existing sample surveys (as the labour force survey and the survey of living conditions) as well as survey data collected especially for the census. The plan is, however, to use this data only for quality assessments and not for improving census data.

For Statistics Norway to conduct a successful register-based census in Norway in 2011, some quality improvements are necessary. When we use administrative data, the success depends quite heavily on the activities carried out by the authorities responsible for the registers. What to do if data quality is not improved sufficiently by 2011? Even if such a perspective is not included in the current census plans, it could be necessary to use data from sample surveys to improve register data. This however, means that the additional surveys must be broadened in several ways. At present our judgement is that it will not be necessary to use such methods. The goal is still to conduct a totally register-based census, and Statistics Norway will do whatever possible to reach this goal.

References

Longva S., Thomsen I. and Severeide P.I. (1998): Reducing Costs of Censuses in Norway through Use of Administrative Registers. Reprint from *International Statistical Review*, Vol. 66, 223-234

Statistics Norway (2006): The role of censuses in a country with a register-based statistical system: Norwegian experiences and plans. Paper for the CES Plenary Session in Paris 13-15 June 2006.

UNECE (2006): Conference of European Statisticians Recommendation for the 2010 Censuses of Population and Housing.

UNECE (2007): Register-based statistics in the Nordic countries. Review of best practices with focus on population and social statistics (to be published in May/June 2007)