

STS50 - Methods and quality of administrative data used in a census**Building and maintaining quality in register populations**

Hendriks, Mr. Coen

Statistics Norway, Division for statistical populations

Oterveien 23

N 2225 Kongsvinger, Norway

E-mail: hen@ssb.no

Åmberg, Mr. Johan

Statistics Norway, Division for statistical populations

Oterveien 23

N 2225 Kongsvinger, Norway

E-mail: jag@ssb.no

Introduction

Statistics Norway (SN) has been using registers for statistical purposes for decades to produce censuses (Longva et al. 1998) and for other statistical purposes. The use of registers in censuses has gradually increased since 1970. The census for 2011 will be the first Norwegian census to be based entirely on registers. This is a result of SN's policy to use administrative register for statistical purposes and relies heavily on a good cooperation with various governmental register owners. The policy was developed during the 1990's and supported politically by means of the Statistics Act of 1989. During the late 1990's SN agreed to establish common practices for developing, updating, and using statistical base registers. Gradually the concept of statistical population management was introduced as SN's common approach for the way the statistical base registers were being developed into statistical populations and made available for the production of statistics.

The paper describes the concept of statistical population management, the duties of the population managers and the consequences for the production of statistics. To exemplify the general principles of population management, the paper discusses experiences and challenges from establishing statistical populations for households and dwellings, primarily for the Census of 2011, but also for running statistics on households and dwellings. Finally the paper presents an example on the use of information on place of usual residence from other sources than the (Central Population Register) CPR.

Part 1 – General principles**1 The concept of population management**

As many other national statistical institutes, SN is experiencing an increasing demand for statistics. At the same time there is a constant demand for the reduction in response burden and for a more efficient use of resources. This has resulted in extensive use of administrative registers in a close and good cooperation with register owners in the public and private sector. One of the means to enhance the use of registers for statistical purposes was to engage strongly in the development of electronic coordination and reporting, based on the principle of re-use of electronic data.

This policy was supported politically by the Act Concerning Official Statistics and Statistics Norway of 1989 and developed in SN during the 1990's. According to the Statistics Act, SN has the right to use registers and must be notified in case changes are planned in the source. These rights are being acknowledged by the register owners. When SN wants to develop a new register based source

for statistical purposes, SN refers to the Statistics Act § 2-2 on the obligation to provide information and § 3-2 on administrative data-processing systems¹.

By the end of the 1990's SN agreed on common practices for developing, maintaining, updating, and using statistical base registers. There is a common understanding on the concept of a statistical base register, also referred to as a statistical population register. A statistical population register is a register which covers all the units in the statistical population, is quality assured, updated and its contents is well documented. The register is adjusted to the needs of the users. The main issue is to keep track of the units in the statistical population by keeping the basic information on the units updated. Basic information is made up from variables which identify and describe the units, typically variables for stratification of a population and core variables for common statistical use. By doing this gradually the concept of population management was introduced.

During a reorganisation in 2008, SN established a new Department for Data Capture. Under this department a Division for Statistical Populations was established. The new division replaced the Division for Business Register, which already had been in charge of the statistical copies of the Register for Legal Entities and the Cadastre. After the reorganisation, the Division for Statistical Populations is responsible for managing three statistical base registers:

- The Central Register for Establishments and Enterprise, based on the administrative Central Coordinating Register for Legal Entities which is owned by the Brønnøysund Register Centre
- The Population Register, based on the administrative Central Population Register, which is owned by the Norwegian Tax Administration
- SN-Cadastre, based on the administrative Cadastre (the national register for ground properties, addresses, buildings, dwellings and digital maps). The Cadastre is owned by the National Mapping Authority

In addition the division manages a newly developed Database for Geographical Data (Hokaasen 2010). The development of a fourth statistical population register for employment is under consideration. In this context employment is identified by the conjunction of an employee's personal identification number and the employer's business identification number. The possible fourth statistical population register will be based on the administrative Register for Employers and Employees in the Norwegian Labour and Welfare Service.

In some cases, a statistical population is established by linking information from two or more registers. This is the case for the dwelling household register as described in part two of the paper. To facilitate the linkage of information from several registers, the coherence and possibility for linkage between the sources must be good. This is a special object of attention for the population manager. It goes without saying that the principle of re-use of basic data is well established in the Norwegian statistical system, as it is in the Norwegian public sector as a whole.

2 The duties of the population manager

A new statistical profession was introduced and referred to as "population manager", though not as an official job title. The primary duty of the population managers is to provide the best possible populations for the production of official statistics, by keeping track of the units in the statistical population registers. More specific the population manager is to integrate the statistical base registers into one database model (re-using common solutions), create a common user interface and to integrate information from base registers into new statistical populations. In the Norwegian case, integration is possible because the population registers have common identifiers for persons (PIN – personal identification number), businesses (BIN – business identification number) and addresses (the numerical address). By means of the common identifiers, information from one population register can

¹ http://www.ssb.no/english/about_ssb/statlaw/

be linked to other registers. The other register might be another population register, but also any other specific register (UNECE 2007).

Quality assurance is an important duty of the population manager. Facilitating successful integration of register based data is a very important aspect of quality assurance but there are other quality dimensions which currently are being explored and developed. Daas et.al. (2010) have made the following list of important quality dimensions for administrative data sources which are used as input for statistics:

- Technical checks (technical usability of the file and the data in the file)
- Accuracy (closeness of the objects and variables to the exact/true objects and values defined, and the extent to which data are correct, reliable and certified)
- Completeness (degree to which a data source includes data describing the corresponding set of real-world objects and variables)
- Time-related dimension (indicators that are time and/or stability related)
- Integrability (extent to which the data source is capable of undergoing integration or of being integrated)

The integration of population registers provides new areas of application and improved statistical base data. Integrating geodata contributes in that respect and opens for extended and improved ways of presenting and analyzing the statistics. Integrating employed persons and employers in a statistical base register for employment is another example of the integrated approach. SN uses the integrated approach also for the construction of the statistical population register for dwelling households. As described in the second part of the paper, the household register is created by linking information on persons from the CPR with information on dwellings from the Cadastre. The numerical address is the linkage key.

It is very important for the population manager to have a service minded attitude and to communicate with the owners of the administrative sources and the users of the data in SN. Again the quality issue is important. The population manager is a very important “centerfield player” enhancing quality awareness at both ends, at the source and at the end user. In many cases SN coordinates the interplay at the source between owners of the administrative base registers. Linking persons to dwellings to establish a household register uncovers many quality issues in the sources. Often the owners of the sources have to cooperate amongst themselves to improve quality. In this type of work, SN is a highly valued partner and very often the driving force. The end users in SN need information on the statistical population. They need guidance and training in using the data, and are encouraged strongly to report on the quality of the data.

3 Consequences for the production of statistics

The population manager makes the population registers and statistical populations available for common use in SN. Arranging the population registers for common use ensures consistency in the statistics as a whole. Consistency is based on a common frame of reference and on official statistical standards.

Within SN there is an obligation to use the population registers. This did put an end to any solution which may have existed locally at the statistical divisions. Previously statistical divisions might have had separate solutions for registering and maintaining separate statistical populations e.g. a specific sampling frame. SN as a whole must use the statistical base registers for sampling, administration of data capture and/or in direct statistical production. There is a common understanding that increased use of a data source improves the quality, which in turn increases the use of the source.

The Division for Statistical Populations provides populations for use directly in statistics, e.g. population by age, sex, marital status and citizenship, or the stock of dwellings. In addition the

division provides populations for statistics which are based on linked statistical populations and which are expected to be consistent, e.g. dwelling households and the stock of occupied dwellings. The Census of 2011 is an end-user with high priority, receiving the census populations for persons, households and dwellings. Sample based statistics use the statistical populations as a sampling frame, as tool for tracing interview objects, and whenever possible, to reduce response burden by minimizing the number of items on the questionnaire or in the interview.

SN coordinates the use of registers internally and towards the register owners. This is done by written agreements, on the use of base registers and specialized registers, between the register owners and SN. SN has centralized the storage of all the agreements, making it relatively easy to keep an overview on the available registers and the agreements on their use for statistical purposes. The agreement states the name of the contact person who will represent SN in the cooperation. The responsibilities are divided amongst the statistical divisions. E.g. the Division for Income and Wage Statistics is, on behalf of SN, responsible for the contact towards the Norwegian Tax Administration. The users of the data in SN must keep the contact person informed on the communication between SN and the register owner. To a certain extent this helps to coordinate the use of the data. There might be room for improvement in this field in SN, perhaps through a more centralised data capture from administrative sources, but this is not SN policy yet.

During the annual series of meetings to develop the operating plan for SN for the next year, the subject of registers and register use is on everyone's agenda. This also contributes to the coordinated use of registers.

Part 2 – Experiences and challenges from establishing a statistical population of households

4 The Norwegian Census of 2011 will be entirely register based

The Norwegian Census of 2011 will be produced from registers only. This will be the first time in the history of SN. The Division for Statistical Populations will produce census populations for persons, households and dwellings. Many census variables will be produced from the statistical population registers. This is the result of many years of dedicated register development and cooperation. The work started in 1964 when the Central Population Register (CPR) and the Cadastre were established, both originally in SN. The Cadastre was transferred to the National Mapping Authority in 1980. In 1992 the CPR was transferred to the Norwegian Tax Administration.

The CPR and the Cadastre have always been of relatively good quality, but they needed a last and major improvement to facilitate a fully register based census. The improvement was done in 2001 by the Dwelling Address Project, when the dwelling number was introduced as a part of the numerical address (Hendriks and Severeide 1998). The dwelling number makes it possible to link the smallest units from both registers, meaning persons from the CPR to dwellings in the Cadastre. Before the introduction of the dwelling number as a part of the numerical address, persons could be linked to buildings, but not to dwellings. By definition, a dwelling household consists of the persons living in a dwelling. In a register based census, a dwelling household is made up from persons linked to dwellings by means of the common identifier for numerical address. It seems a simple operation, linking persons to dwellings, but in reality it is the result of many years of hard work in SN and between SN and the register owners. The work also involved the municipalities who were responsible for indentifying the dwellings and entered them into the Cadastre.

In the case of Norway, the numerical address consists of the following elements:

	Digits	Example	
Municipality code	1-4	0301	Municipality of Oslo
Street code	5-9	17059	Storgata
House number	10-13	0045	House number 63
Entrance	14-17	9901	Entrance A
Sub number	18-21	0000	Not in use
Dwelling number	22-26	H0205	Dwelling number

The numerical address from the example is 0301 17059 0045 9901 0000 H0205. The notation of the numerical address is the same in the CPR and the Cadastre. This makes it possible to link information on persons, identified by the PIN-code, from the CPR to information on dwellings, identified by the numerical address, from the Cadastre. The process of linking persons to dwellings is referred to as the micromethod in the rest of the paper. Additionally the Cadastre provides a lot of other information, such as type of building, year of construction, number of rooms and other core variables for the census, including coordinates.

5. The micromethod – legal residence

We will now briefly present the main features of the micromethod where we link residents from the CPR and dwellings from the Cadastre using the address. The linkage is done at micro level, i.e. that all individuals are connected to a unique dwelling when possible. When we can not link directly we use address information from the CPR and the Cadastre, hoping to uncover and correct for errors and omissions in the registers. We break the population down in to many different subpopulations because they must be handled in slightly different ways. All depending on the type of error or omission we find. The subpopulations follow a hierarchical structure in which we handle the most secure first. Further on in the process we use less secure methods for finding links. This ranges from updating parts of the CPR-address to selecting a dwelling by imputation. The number of persons linked to dwellings in each subpopulation documents the changes due to adjustments in the method and/or changes in the registers. The micromethod constantly strives to improve the quality of the different subpopulations. Over time, this will serve as quality indicators where hopefully fewer and fewer persons will be handled in less secure methods.

Household formation based on address information from the CPR only, provides too few households in all. The micromethod approach provides an under-representation of small households and an over-representation of large households. A different approach, using updated household formation from the 2001 census, gives too many households in all, with too many small households and too few large households. Therefore we transform the registry information. If we only had used dwelling addresses from the CPR the number of households per 1st January 2011 would be 1 961 000 (the number of unique dwelling addresses in the CPR). So far, the micromethod has resulted in 2 047 000 private households (2 081 000 incl other households). But the work is not completed and the number of households is expected to rise.

6 Additional information on addresses – place of usual residence

According to the international recommendations for population and housing censuses (UNECE 2006):

“The place of usual residence is the geographic place where the enumerated person usually resides. This may be:

- (a) The place where he/she actually is at the time of the Census: or*
- (b) His/her legal residence: or*
- (c) His/her residence for voting or other administrative purposes.”*

So far the paper was concentrated on address for legal residence (b) from the CPR. SN has collected addresses from other sources than the CPR. This might be information on actual address (a) or address for administrative purposes (c). A person might end up with an original legal address from the CPR, an adjusted address from the micromethod, an address from a university, a postal address from Posten Norge AS, an address from The Norwegian Labour and Welfare Service etc. We then end up with a file containing all the available addresses on a person at a given time. This makes a solid foundation for the operationalisation of the concept of place of usual residence as defined above.

We expect to find addresses for place of residence, which differ from legal address, for students, long-distance commuters and institutional residents. The expected deviation population therefore controls and orders the priority of which addresses are used when more than one address is available. We will find several addresses for persons who for different reasons have a need for different addresses in external registers. This could be postal addresses or addresses used for administrative purposes. Missing or late notifications on change of legal address will in some cases also be detected using the sources for actual addresses. A major problem for the institutional population, is that the source does not contain any information other than that the person has visited an institution during the year. As a result, we can change the original household based on legal address, but we can not link the person to a new address.

Furthermore, we identify persons with one or more addresses from other sources that differ from the registered address in the CPR. This could be a student address and a job in the same municipality, but far away from the legal home address. In cases where many actual addresses are found on a person, we select the newest, the one with best quality and the one that follows the priority order of the expected deviation population. Registration dates are constantly compared to the last known date for change of legal address from the CPR.

Many of those identified with a new actual or administrative address may only be linked to a municipality or zip code. In some cases this is due to poor quality of the actual address. This is obviously a big problem because we want to link persons to dwellings. The method for linkage gradually broadens the basis for imputation (linking randomly) from vacant dwellings at the same entrance, via housenumber, street code to zip code area. Finally the method imputes from all vacant dwellings at municipality level when no better alternative is available.

By 1st January 2011 we had approximately 371 000 vacant dwellings after linking persons to dwellings using legal addresses. Approximately 25 percent of these were found in streets with fewer than 5 vacant dwellings. On average there are 5.6 vacant dwellings per street. This describes to a certain extent the basis we have for the imputation of dwelling-numbers for persons with incomplete actual addresses. There are of course more vacant dwellings available for imputation in large cities and fewer in the rest of Norway. This will limit the level of detail in the statistics.

By 1st January 2010 almost 116 000 persons were given a place of usual residence which is different from the legal address. That is about 2 percent of the Norwegian population. The table below shows the quality of the sources which were used and the level at which they can be linked to the Cadastre.

Source	Quality						Total	%
	0. Municip.	1. Zip	2. Street	3. Grnd prp.	4. House nr	5. Entrance		
EE	0	6 441	0	0	0	0	6 441	5,57 %
Edu1	0	15 672	0	957	1 019	34 585	55 413	47,90 %
Post	0	0	0	1 116	727	27 013	33 223	28,72 %
Soc.	0	0	0	409	284	9 724	12 486	10,79 %
Edu2	8 113	0	0	0	0	0	8 113	7,01 %
Total	8 113	22 113	0	2 482	2 030	71 322	115 676	
%	7,0 %	19,1 %	0,0 %	2,1 %	1,8 %	61,7 %		8,3 %

Table 1. Quality in the alternative sources for place of usual residence (EE = Register for Employers and Employees, EDU1 = Educational Institutions, Post. = Register for Postal recipients, Soc. = Social Security Register, EDU2 = Universities and Colleges Admission Service).

7 Actually occupied dwellings in the city of Trondheim

We try as far as possible with respect to quality, to identify a person's place of usual residence. This may be a legal, actual or administrative address. We then end up with a population of person according to their actual address, as far as possible, in addition to an already established population of persons according to their legal address. We can provide populations that will suit different user needs. It will be possible to generate statistics for legal address (option b from the international recommendations). In addition, it will be possible to generate statistics for place of usual residence for selected groups (option a). A population and housing census in accordance with international recommendations shall be initially conducted after "Place of usual residence". However, by using legal address only, we do not properly illustrate the actual distribution of persons in Norway. Furthermore the number of vacant dwellings will be overestimated. The following example from Trondheim shows that the distribution of vacant dwellings after linking persons by legal address (figure 1) is misleading, compared to the distribution of vacant dwellings after linking persons by actual address (figure 2).



Figure 1 Vacant dwellings in the city of Trondheim based on legal addresses.

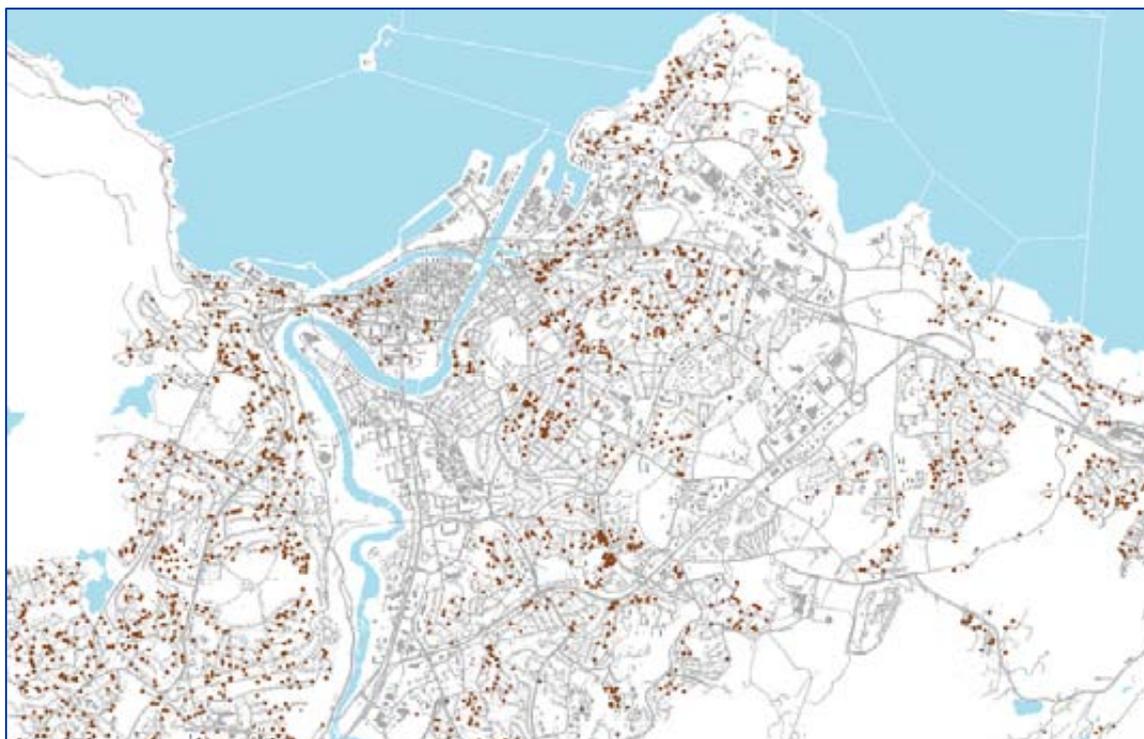


Figure 2 Vacant dwellings in the city of Trondheim based on actual addresses.

8 Final remarks

As pointed out in the first part of the paper, the population manager makes the population registers and statistical populations available for common use in SN. Arranging the population registers for common use ensures consistency in the statistics as a whole, which is a very important advantage of the approach. Population managers have viewpoints and can make recommendations on how to use the data. However, it is up to the statistical divisions in SN to determine which specific populations and statistical concepts are being used. This underlines the importance for the population manager to have a service minded attitude and to communicate with the owners of the administrative sources and the users of the data in SN.

REFERENCES

- Daas, P. , S. Ossen, M. Tennekes, Li-Chun Zhang, C. Hendriks, K. F. Haugen, T. Laitila, A. Wallgren, B. Wallgren, A. Bernardi and F Cerroni 2010: List of quality groups and indicators identified for administrative data sources. *Deliverable WP4.1 BLUE-Enterprise and Trade Statistics. European Commission European Research Area, 7th Framework Programme*
- Furseth, J. 2010: Om populasjonsforvaltning – Tema 2 Nye datamuligheter. *Paper in Norwegian to the Nordic Statistician Meeting (Copenhagen, Denmark, 11-13 August 2010)*
- Hendriks, C. and P.I. Severeide 1998: Establishing a dwelling register in Norway: The missing link. *Paper to the joint ECE-EUROSTAT work session on Population and Housing Censuses (Dublin, Ireland, 9-11 November 1998)*
- Hokaasen, E.O. 2010: Geodatabaseprosjektet – Tema 3 Statistikkproduksjon. *Paper in Norwegian to the Nordic Statistician Meeting (Copenhagen, Denmark, 11-13 August 2010)*
- Longva, S., I. Thomsen and P.I. Severeide 1998: Reducing costs of censuses in Norway through use of administrative registers. *International Statistical Review*. 1998, 66, 2, 223-234
- UNECE 2006: Conference of European Statisticians recommendations for the 2010 censuses of population and housing.

UNECE 2007: Register based statistics in the Nordic countries. Review on best practices with focus on population and social statistics. *Also available in Russian*

Abstract

Statistics Norway (SN) has been using registers for statistical purposes for decades. During the late 1990's SN agreed to establish common practices for developing, updating, and using statistical base registers. Gradually the concept of statistical population management was introduced as SNs common approach for the way the statistical base registers were being developed into statistical populations. In 2008 this approach was consolidated by an organisational change in SN. A separate Division for Statistical Populations under the Department of Data Collection was established. The paper will present the concept of population management, the duties of the population managers and the consequences of the approach for the production of statistics. To exemplify the general principles of population management, the paper will discuss experiences and challenges from the work on establishing statistical populations for persons and households. This will also include the use of information on place of usual residence from other sources than the Central Population Register. Even though the examples will be taken from the demographic realm, they will be of general interest. The principles, which will be discussed, are also representative for the work on other statistical populations.