Methodological Experiences from a Register-Based Census

Claes Andersson[1], Anders Holmberg[2], Ingegerd Jansson[3]
Karin Lindgren[4], Peter Werner[5]
[1]Statistics Sweden, 701 89 Örebro, Sweden
[2]National Institute of Statistics, 386 Monivong Blvd, Sangkat BKK 1, Chamkar Mon, Phnom Penh, Cambodia
[3]Statistics Sweden, Box 24300, 104 51 Stockholm, Sweden
[4]Statistics Sweden, Box 24300, 104 51 Stockholm, Sweden
[5]Statistics Sweden, 701 89 Örebro, Sweden

Abstract
In 2011, the first completely register-based census was conducted in Sweden. Several registers, such as the Total Population Register and the Real Property Register, have been matched to enable the results. Being the first of its kind in Sweden, the Census means that some methodological questions have been studied in more depth than previously. This paper summarizes the experience from the methodological work, the choices that have been made and the trade-offs. We particularly focus on the measures taken to evaluate and report the quality of the final statistical register, including an evaluation study. We also briefly describe measures taken to ensure the confidentiality of the published statistics.

Key Words: Register-based statistics, Evaluation study

1. Background

During 2011, national censuses have been carried out within the European Union. Census data from all member states will be published by Eurostat, the statistical office of the EU. Populations, variables, levels, and categories have been defined and specified in detail by Eurostat. The data is compiled in large multidimensional tables, pre-specified by Eurostat, and must be delivered to Eurostat no later than the end of March 2014, including extensive documentation on the quality of data. Results from all member states will be available through a common interface using a hub solution, developed and maintained by Eurostat. Many national statistical offices will have additional dissemination of results, but not Statistics Sweden.

In Sweden, the census was for the first time fully based on registers. In order to facilitate the production of census statistics, a system of registers has been built that will be an important part of the future production of official statistics on households and housing.

Sweden has a long tradition of keeping track of the number and demography of the population. It dates back to 1688, when the church was first required to keep books of births, deaths, and households in all parishes. In 1749, Tabellverket was published, the first continuing census in the world. Twenty years later, Tabellkommissionen was founded, the first statistical agency in the world and the predecessor of Statistics Sweden (Jorner 2008).

Between 1960 and 1990, traditional censuses were conducted every fifth year, with data being collected by a self-administered mail-out mail-back questionnaire. In 1995, the Swedish parliament took the decision that the next census should be completely register-based. For several reasons, among them political concerns of privacy, the necessary legal regulation was not in place until more than ten years later. At that time, a fully register-based census was not feasible due to the lack of a link between the Total Population Register (TPR) and the Real Property Register (RPR). In 2007, the Swedish parliament passed a bill on the creation of a new dwelling register, including multi-dwelling buildings and single houses, which would be a link between the TPR and the RPR and facilitate the creation of households.

Although vastly experienced in register-based statistics, Statistics Sweden faces new challenges with a completely register-based census. Register-based statistics differ from traditional surveys in some respects. A register-based survey utilizes a register created and maintained for administrative purposes. In a register-based system, a 'statistical register' is first created where the administrative data are edited and transformed to best meet multiple purposes and the aims of multiple surveys. The system of statistical registers relies on three core registers kept by Statistics Sweden: the Business Register, the TPR, and the RPR. The core registers are linked to various subject matter registers such as registers of employment, occupation, education, and buildings.

Comprehensive identification of persons, businesses, buildings, and dwellings is an essential part of the system. Unique identification of persons and businesses has been in use in Sweden for many years for various administrative purposes. However, addresses had to be updated as a preparation for the construction of the new unique dwelling identification key. For single houses, the address is unique. For apartment buildings, the address until recently would only tell the entrance and possibly the floor, but not the apartment. In order to identify each apartment, the property owners were responsible for labeling apartments in a block of apartments according to specific rules, since the numbers carry information about the ordering of apartments on a floor. The property owners were also responsible for informing the residents of their dwelling numbers. This was followed by a mailed questionnaire to every adult in the country where the Tax Agency (where the administrative register of all individuals in Sweden is kept) asked for the address of residence, including the dwelling number, and the TPR was updated accordingly. As residents move to a different apartment or single house, they are required to inform the Tax Agency. This is on the whole a well-functioning procedure since almost all public services utilize the administrative register and there is an incentive for individuals to keep the information updated in order to get information, social benefits, and other types of public services.

The registration of the population on addresses including dwelling numbers started in September 2010 and the last forms where sent out by the Tax Agency in March 2011. The process of preparing and compiling the census statistics, and methodological aspects thereof, has been described in more detail Axelson et al (2010) and Hedlin et al (2011a, 2011b).

2. Evaluating the quality

Eurostat requires the member states to submit a description and evaluation of the quality of the census data. A thorough description of the data collection process and the processing of data will be part of the documentation.

2.1 Households

The forming of households and families is entirely based on the information from the registers. Individuals registered at the same dwelling form a dwelling household (not necessarily the same as a household in terms of a housekeeping unit). From the information in the registers, household and family variables are derived, such as size of family or household and type of family or household. This requires some information in addition to where persons are registered (i.e. legal marital status or child and parent relations), in combination with some basic rules (i.e. there has to be at least two people to make up a family, two married couples living together count as two families, children with divorced parents can only be counted as members of one household, etc.). There can be more than one family in a household, but never more than one household in a family.

Evaluation studies have been carried out in connection with all previous Swedish censuses (Andersson and Holmberg 2011). The purpose of the current evaluation study is to estimate the error size of some of the error sources present in the individual and household statistics of the census. In particular, focus has been on how complete and how accurate the population registration on dwelling number is, and how this may influence household type and household size.

A stratified sample of 15 000 individuals was drawn from the TPR, targeted towards groups where register data is missing or low quality is expected (i.e. large cities or missing dwelling number in the TPR). A total of 108 strata were defined by five variables; Dwelling number exists in the TPR, Municipality of residence, Age class, Type of dwelling, and Number of families in dwelling according to the TPR. About 1/3 of the sample was allocated to the group with missing dwelling number. Within both groups, a sequential scheme with proportional allocation with constraints was applied.

Data collection started in January 2012 and ended in May the same year. Respondents were given the opportunity to answer on line or by a mailed questionnaire. The first approach, sent by mail, was an invitation to respond by web only. A second invitation was sent by mail to those not yet responding, including a printed copy of the questionnaire. Finally there was a third possibility to respond by a telephone interview.

The respondents were asked to confirm the address they were registered at on 31 December 2011 (reference date of the census). If the address given was not correct, they were asked for a correct address. Other questions concerned if the dwelling is owned or rented, how many dwellings there are at the same address, and how many other persons were living at the same address on the reference date. For all others living at the same dwelling, name, sex, year of birth, and whether the person in question was living together with parent or spouse/partner was asked for.

The total weighted response rate was 65 percent. In all strata, the response rate was at least 40 percent. As anticipated, nonresponse is higher among those who are not registered with a dwelling identification number in the TPR. Nonresponse is also higher among younger people (18-34 years). As can be expected, nonresponse is higher when there is more than one family registered at the same dwelling. This is probably an indication of a higher amount of incorrect information in the register for those records.

The evaluation of the register had itself to be evaluated. If household size or household type differed between survey and register, there was a re-contact by telephone in order to establish the "true" value. Almost 3000 individuals were contacted and 85 percent of them agreed to confirm their answers from the initial data collection. The re-interview focused on household size and type of household. In 65% of the cases, the initial data collection gave the true value, in 25% the register gave the true value, and in the remaining 10% of the cases neither the initial data collection nor the register was correct.

The design of the evaluation study is described in more detail in Axelson et al (2012).

Table 1 shows some results on household size. The number of correctly classified households is the number of households where the register and the evaluation study agree. Gross error is the number of households wrongly included in the category plus the number wrongly excluded. Net error is the number of households according to the register minus the number according to the evaluation study. The proportion of correct households and the relative net error are calculated relative to the number of households according to the register.

The results indicate that the numbers for smaller households are underestimated and the numbers for larger households are overestimated by the register.

Table 1: Results from evaluation study: Household size

Size of household	1	2	3	4	5	6-10	10<	Total
Number correct	1 410 283 ± 15 366	1 212 330 ± 13 768	396 320 ± 16 524	426 208 ± 15 092	126 055 ± 10 684	35 237 ± 6 386	143 ± 121	3 606 574 ± 32 662
Proportion correct	98.0 ± 1.1	95.5 ± 1.1	78.4 ± 3.3	83.9 ± 3.0	71.4 ± 6.1	45.2 ± 8.2	7.3 ± 6.2	90.7 ± 0.8
Gross error	256 571 ± 45 277	306 984 ± 35 894	199 400 ± 24 748	146 163 ± 20 854	67 824 ± 12 387	45 666 ± 6 787	1 814 ± 121	---
Net error	-199 233 ± 45 233	-192 778 ± 35 853	18 352 ± 24 801	17 755 ± 20 934	33 210 ± 12 431	39 760 ± 6 782	1 814 ± 121	-281 119 ± 33 799
Relative net error	-13.8 ± 3.1	-15.2 ± 2.8	3.6 ± 4.9	3.5 ± 4.1	18.8 ± 7.0	51.0 ± 8.7	92.7 ± 6.2	-7.1 ± 0.8

Table 2 is the corresponding table for household type. Results are similar to Table 1, i.e. the numbers for larger households (multi-person, 2 or more families) are overestimated and the numbers for smaller households are underestimated. One explanation to the figures could be that younger people and students fail to register their new address as they move away from their parents. Note however that the larger types of households are less common and thus the number of such households in the sample is low.

Table 2: Results from evaluation study: Household type

Household composition	1 adult, no children	1 adult with children	2 adults no children	2 adults with children	Multi-person households	2 or more families	Total
Number correct	1 410 281 ± 15 365	263 768 ± 13 814	1 037 597 ± 9 798	949 246 ± 18 511	38 687 ± 10 014	11 038 ± 3 754	3 710 616 ± 31 328
Proportion correct	98.0 ± 1.1	88.1 ± 4.6	97.7 ± 0.9	93.6 ± 1.8	34.3 ± 8.9	21.3 ± 7.2	93.3 ± 0.8
Gross error	267 860 ± 46 619	64 927 ± 17 389	197 325 ± 30 473	118 991 ± 21 557	91 124 ± 13 304	44 465 ± 4 450	---
Net error	-210 517 ± 46 625	6 362 ± 17 335	-149 317 ± 30 465	9 786 ± 21 561	56 890 ± 13 311	37 325 ± 4 462	-249 471 ± 32 734
Relative net error	-14.6 ± 3.2	2.1 ± 5.8	-14.1 ± 2.9	1.0 ± 2.1	50.5 ± 11.8	71.9 ± 8.6	-6.3 ± 0.8

2.2 Dwellings

2.2.1 Unoccupied dwellings

Dwellings that appear to be unoccupied pose a particular problem with the dwelling register. There are likely to be unoccupied dwellings, but also possible that dwellings will falsely appear as empty in the register. In order to investigate this further, data from a survey on unoccupied dwellings in multi-dwelling buildings was matched to the register. The survey targets rental units on the open market and has two parts, a total survey of municipal housing companies and a sample survey of private bodies. The first step was to evaluate the number of dwellings according to the survey and according to the dwelling register. The survey and the register do have slightly different reference dates, 1 September 2011 and 31 December 2011.

The results show that the number of estimated dwellings is higher in the survey. The number of dwellings in the register is underestimated by approximately 10 percent for the municipal housing companies. The main reason is that in the survey some respondents have included dwellings that according to the register are meant for special housing such as dorm rooms and nursing homes. This tendency was not as strong for the private bodies. The comparison between numbers of unoccupied dwellings in the survey and dwellings appearing as unoccupied in the register shows that the register overestimates the unoccupied dwellings to a high extent, which was an expected result.

2.2.2 Type of ownership

The census variable on type of ownership differs in categories and definitions from what is commonly used in Sweden and supplied by the registers. Thus this variable called for a rather complicated derivation, and it is necessary to evaluate its quality. A question on type of ownership was added to the evaluation study, and the data from the survey was compared to the census data. Table 3 shows results from the comparison. The relative net error rate indicate that the quality of the data is suffcient, but the results should be interpreted with caution since some assumptions have been made when translating the Swedish situation to fit the definitions supplied by Eurostat.

Table 3: Results from evaluation study: Type of ownership

Type of ownership	Owner-occupied dwelling	Dwelling in cooperative ownership	Rented dwelling
Number correct	1 631 015 ± 24 254	797 256 ± 20 036	1 421 220 ± 31 590
Proportion correct	96.5 ± 1.4	94.4 ± 2.4	95.3 ± 2.1
Gross error	99 994 ± 27 865	132 585 ± 28 793	227 599 ± 41 964
Net error	18 794 ± 27 253	-37 932 ± 28 883	-87 318 ± 42 062
Relative net error	1.1 ± 1.6	-4.5 ± 3.4	-5.9 ± 2.8

3. Other quality aspects

3.1 Missing data

If there is no dwelling key registered for an individual in the TPR, this individual cannot be linked to a dwelling and included in a household, and household or family variables cannot be derived. Missing keys can have several causes; individuals failed for some reason to register their dwelling number with the TPR, or the property owner failed to register dwelling numbers or inform the households residing in the building of their dwelling number. 307 000 persons, or 3.2 percent of the persons in the TPR, had no registered dwelling key as of December 31 2011.

The rate of missing dwelling keys differs geographically. For two thirds of the municipalities (290 in total), the rate is at most three percent of their population. The largest rate for a single municipality is 14 percent. It is anticipated that the problem of missing keys will decrease with time as people move to a new apartment and need to register their new address with the Tax Agency.

It was decided that no imputation of households should be carried out. One reason is that there is no useful auxiliary information from other sources to aid the imputation (recall that the previous Swedish census was carried out in 1990). Matching persons without

dwelling keys to apartments that appear to be unoccupied is another option. However, 48 percent of the persons with missing dwelling keys have addresses in buildings with no apartments according to the RPR, implying that the numbering of apartments failed for the whole building. The remaining 52 percent are registered at addresses where the buildings have at least one registered apartment. The registered apartments might be used for matching, bearing on the assumption that the relationship between size of apartment and size of household is similar for individuals with and without registered dwelling keys, but this would require a large effort with doubtful quality of the result and no such calculations have been carried out.

Considering the above, and the fact that compensation for missing data is not required by Eurostat, it was decided that missing data will be documented and reported, but not adjusted for.

Item missing data occurs to a small extent. The Census relies on the work carried out at the different register holders within Statistics Sweden. Information on item missing data is reported to Eurostat and will be available as the census data is published.

3.2 Coverage

A register based census is not immune to coverage problems, but the causes of these problems are different than those of a traditional census.

There has not been any evaluation of the coverage of the census. However, the coverage of the TPR is investigated regularly at Statistics Sweden and its source, the administrative register kept by the Tax Agency, has also been evaluated.

3.3 Disclosure control

Eurostat has taken the decision that each country should decide on feasible measures for risk assessment and methods to protect the individuals and households from disclosure. Data that are delivered to Eurostat are assumed to be properly protected and no measures for statistical disclosure control will be taken by Eurostat. This will possibly affect the comparability between countries, but is deemed necessary due to country specific regulations on confidentiality and secrecy.

Risk assessment for the Swedish census data focuses on a few particularly sensitive variables in combination with units that are particularly easy to locate in the data. The main measure taken to protect the data is record swapping of random pairs of households on a geographical variable (see for example Shlomo et al 2010 for a description of the method, and Jansson 2012 for a discussion preceding the choice of methodology for the Swedish census). This method has the advantage that all tables produced from the protected data are consistent, and the method can be targeted so that households most at risk for disclosure are more likely to be swapped. However, some individuals are unique at a national level when country of birth or citizenship is combined with age and gender, and extra measures had to be taken in order to protect this group from disclosure. For this group, the country of birth/citizenship is changed within continent according to a probability matrix based on the country frequencies in the population.

4. Future use of the system of registers

The system of registers that now is in place will be an important part of the future production of official statistics on households and housing. An up to date and coherent system of registers makes it possible not only to improve the official statistics on households and housing, but also to make statistics on demand with improved longitudinal quality at low cost, increase the possibilities of producing statistics on smaller domains and special populations, and provide standardised register variables and populations. Great investments have been made due to Census 2011, but the usefulness of the system of registers stretches far beyond a single census.

References

Andersson C. and Holmberg A. (2011). Methods and design to evaluate the 2011 Census and future register based household and housing statistics. Discussion paper to the Advisory Scientific Board of Statistics Sweden.

Axelson M., Hedlin D., Holmberg A., and Jansson I. (2010). Methodology in the Swedish register based census. Paper presented at the 2010 International Methodology Symposium, Statistics Canada, Ottawa, October 26-29.

Axelson M., Holmberg A., Jansson I., Werner P,. and Westling S (2012) Doing a Register-based Census for the First Time: The Swedish Experiences. Paper presented at the 2012 Joint Statistical Meetings, San Diego, 28 July – 2 August 2012.

Hedlin D., Holmberg A., Jansson I. and Lorenc B. (2011a). The first fully register-based census in Sweden. Paper presented at the 2011 Joint Statistical Meetings, Miami Beach, 30 July – 4 August 2011.

Hedlin D., Holmberg A. and Jansson I. (2011b). Combining registers into a fully register-based census – some methodological issues. Paper presented at the NORC U.S. Census Bureau conference: Utilizing Administrative Data: Technical, Statistical and Research Issues, Washington D.C., 27-28 October 2011.

Jansson, I. (2012) Issues and plans for the disclosure control of the Swedish Census 2011. Paper presented at the Workshop on Statistical Disclosure Control of Census Data, Luxembourg, 19-20 April 2012.

Jorner, U. (2008) Summa summarum. SCB:s första 150 år. Statistiska centralbyrån.

Shlomo, N., Tudor, C., and Groom, P. (2010) Data Swapping for Protecting Census Tables. In Privacy in Statistical Databases, Ed. J. Domingo-Ferrer and E. Magkos. Proceedings of the CENEX-SDC Project International Conference, Corfu, September 2010. Berlin: Springer.