Session 592 Uses of administrative record benchmarking in modern census-taking

Input data quality in register based statistics – The Norwegian experience

Hendriks, Mr. Coen
Statistics Norway, Division for Statistical Populations
Oterveien 23
N-2225 Kongsvinger
Norway
E-mail: hen@ssb.no

Acknowledgement
The author wishes to thank Mr. Jan Furseth for usefull comments on the draft of the paper.

Introduction

The three C's of register based statistics are Co-operation, Communication and Co-ordination. They contribute strongly to improved quality in register based statistics. The paper briefly describes the current practise in Statistics Norway (SN) for handling registers. Experiences from the first fully register based Census of 2011 and reviews according to Eurostat's Code of Practise show there is a need for improved quality of input data. Improvements can be achieved by several types of co-operation between SN and register owners. The paper describes three approaches: two types of agreements, on respectively data processing and co-operation, and thirdly the forums for co-operation between SN and register users. Common for all the types of co-operation is that they have a sound legal basis. In addition, they all have strong elements of communication and co-ordination within SN, between SN and register owners and among register owners.

The current practise in Statistics Norway for handling registers

Statistics Norway (SN) is using four administrative registers as base registers for the production of statistics. The base registers are:

- The Central Coordinating Register for Legal Entities (CCRLE)
- The Cadastre (Ground properties, addresses, buildings and dwellings)
- The Central Population Register (CPR)
- The Register on Employers and Employees

To facilitate the statistical use of these registers, SN has established databases with a statistical version of each register except for the Register on Employers and Employees. The databases are updated daily with notifications from the administrative sources. They are integrated with each other, allowing data inspection at micro level, browsing from one database to the other and extraction of combined data from different sources. We often use the numerical address as a common key for linkage, but naturally the personal identification number (PIN) and the business identification number (BIN) are also being used frequently. Access to these databases is given separately for each database on a strictly need to know basis. Managing the databases for statistical purposes is called statistical population management (Hendriks and Åmberg 2011).

In addition to data from the base registers, SN receives data from many other administrative sources. The data can be a time slice e.g. at end of the year or on Census

Day, or the data can cover all the units or events during a reference period. These registers are considered specialized administrative registers. Unlike administrative base registers, specialized administrative registers serve one specific purpose or a clearly defined group of purposes. The authority maintaining the registers is normally the main user of the information and the data content is naturally influenced by this fact. Specialized registers often receive information on the population and some basic data from a base register, but supply other data themselves (UNECE 2007).

Statistical population management is the main duty of the Division for Statistical Populations in SN. Specialised registers are currently handled by the statistical divisions which are in charge of the subject matter which is covered by the register. This has resulted in different practices for receiving, handling and storage of the data. Not in the least, this has resulted in different practices for getting the data ready for common use. This is not an ideal solution.

The Department for Datacapture, of which the Division for Statistical Populations is a part, has taken an initiative to professionalize and develop the co-operation on registers within SN, and between the register owners and SN. Some of the measures are:

- Professionalising the contact with the register owners
- A list was drawn up with information on 82 registers which are being used regularly for statistical purposes. The list provides updated information on registers, register owners, agreements and decisions on filing requirement
- Guidelines for the follow up of the register owners during and after datacapture
- Improved professionalism in receiving the data, dataflow to the users of the data, population management and facilitating for the use of data internally in SN
- Better use of common information in the population registers
- Ensure involvement from other statistical divisions in the management of statistical basic data
- Improve the quality in the administrative sources and develop ways to measure, document and communicate on quality of the registers

Some of the measures are in the planning stage, some are currently being developed and some have been in place for some time. Two very interesting measures are the first one on professionalising the contact with the register owners and the last one on quality improvement. These measures will be highlighted further down in the paper.

Experiences from the Norwegian register based Census

SN has been planning a register based census for decades (e.g. Longva et.al. 1998). The Census of 2011 is the first Norwegian fully register based census. In principle doing a register based census is about matching information from relevant registers and compiling the statistics. However this turned out to be rather complicated when it came to the production of register based census statistics on dwelling households, occupied and non-occupied dwellings and on housing conditions.

When matching the resident population from the CPR to dwellings from the Cadastre, one should expect to get a reasonable number of dwelling households. This exercise was done and compared to the established statistics on households. Even though the work was not finished entirely, some remarkable differences became clear. The number of private households from the matched registers was reduced by 6 percent compared to the established household statistics. The number of persons living alone was reduced by 17

percent, dropping to the level of the early 1990's. Overall the matched registers resulted in fewer, but larger households.

Errors occur in the register households whenever people who do not live together are grouped into the same household, and/or when people in the same household are divided into different households. Such errors are described as unit errors. A unit error problem involves more than just mismatching between two sets of fixed units. There are two reasons why the number of households is unknown. Firstly, the register on dwellings is not perfect. There are both missing and wrongly registered dwellings identification numbers in the dwelling register (which is a part of the Cadastre). There might also be delays in updating. Therefore the number of dwelling units at a given address cannot be known for sure. Secondly, even when the errors in the dwelling register are disregarded, it is not true that the number of dwelling households will always be the same as the number of dwelling units. Moreover, unit errors will almost certainly arise in a longitudinal perspective, because the updating of the dwelling identification number in the population register is not perfect (Zhang 2011a).

Unit errors also occur in the CPR. A notification of change of address, in the case of removal, might not be sent or it may be delayed. Resident persons may have left the country without sending a notification of emigration. In October 2011, SN did a postal survey among 218 000 resident foreigners and resident Norwegians born abroad, aged 20 years and older. The postal services returned 22 000 envelopes because they could not be delivered correctly to the addressee. An analysis of the information in the CPR on the returned addressees gives an indication on unit errors in the CPR. After four months seven percent of the addressees had emigrated. Six percent had residence and work permits which were overdue, indicating that they possibly have left the country. Eleven percent had sent a notification of change of address in Norway. This indicates that there is a considerable time lag between the actual emigration or removal and the registration of the event.

After much discussion on the methodology it was decided to base the census file for households on the households from the running statistics, linked to dwellings from the Cadastre. This ensures that the number of households, their distribution according to the number of persons per household and other statistical properties from the running statistics on households was maintained in the census. However, the running household statistics was revised in advance because it had not covered the recent development in housing of immigrant workers. Single immigrant workers tend to be registered on specific addresses, resulting in very large households from the register, while the actual situation might be quit different for this specific group. Persons in these types of register based households are counted as unspecified types of private households in the census.

Co-operation on quality - agreements on data processing

Even though there was much debate in SN on the methodological approach towards register based census statistics on households and dwellings, everyone agrees that the input quality of the data could be improved. Rather than repairing errors, they should be avoided in the source. This is the case for the register based sources for the Census, for every other register based source for statistics and by and large for every other use of administrative data.

SN uses the Code of Practise (Eurostat 2011) as a framework for systematic quality reviewing of statistical products and processes. By the end of 2012, every statistical

division in SN will have had at least one review according to the Code of Practise. So far the reviews have underlined the importance of improved co-operation on quality with the register owners. This will reduce the need for editing.

SN aims at improving the co-operation with register owners. SN has good skills in combining data from different sources and on quality assurance of register based sources in general. The register owners have good skills on the subject matter for the register, but are very often lacking skills in the practical work on registers. Combining these skills will without doubt lead to better quality in the sources.

In this co-operation SN can give feedback on quality issues at micro level to the register owners, as long as the errors are found within the source. When errors become apparent after matching information from two or more sources, SN can give feedback by means of aggregated data. Quality reports at micro level from one source are considered a complaint on the data quality in the source. Such reporting, using a single source approach, is well within the provision on secrecy from the Norwegian Statistics Act[1]. However sharing information from a dataset which combines information from different sources would be in conflict with the Statistics Act and is not practised in general. In some cases SN has received a request from a register owner to assess data quality. Provided the register owner is allowed to use other administrative data, an agreement on data processing can be drawn up between SN and the register owner. An agreement on data processing regulates the matching of the sources for quality improvement purposes and the reporting of errors at micro level (the multiple source approach).

Such an agreement has been drawn up between SN and the Norwegian Tax Administration. The Tax Administration is in charge of the Central Population Register and has access to the Cadastre from the Norwegian Mapping Authority. The Cadastre is being used by the Tax Administration to improve the quality in the CPR, e.g. by checking information from a notification of change of address against type of building, the number of dwellings in a building, or by checking the validity of an address.

After the agreement on data processing was signed, SN can do these and other checks electronically and on a much larger scale than the Tax Administration can do by itself. After matching persons from the CPR against the Cadastre, SN has identified groups of persons with incomplete or illogical data and transferred these to the Tax Administration for follow-up. On a population of 5 000 000 persons, approx. 61 000 persons are registered without dwelling numbers, while information from the Cadastre shows that these persons should have been registered with a dwelling number. Approx. 34 000 persons are registered on addresses without any dwellings at all, which is very unusual in Norway. 61 000 addresses are having too many persons registered. The number of errors is extensive and there are various types of them. These errors cause many problems for register based household statistics, and indeed for other administrative use of the data. Therefore SN is very happy to cooperate with the Tax Administration.

Co-operation on quality – agreements on co-operation

The Code of Practise reviews showed that there is a potential for improved co-operation with the owners of the administrative registers. Co-operation on quality with register owners has been SN-policy for many years, but the approach was renewed recently.

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

The management of SN initiated the process, starting in November 2011. It was decided to establish general agreements on co-operation with external register owners in order to improve input quality in the registers which SN uses in the production of statistics. The approach has become a routine during the annual updating of next years operating plan. So far SN has received many positive reactions from the major register owners.

The purpose is to improve and standardise the contact with register owners, in order to improve the quality of administrative data sources before they are transferred to and checked by SN. This approach will be beneficiary for SN, other users of the registers and for the register owners. In all this will contribute to a reduction in public spending.

A coordinating group was established in SN with representatives from the statistical departments, from the Divisions for Statistical Populations and Statistical Methods and Standards. The legal advisor in SN participated also and the person who hade done the Code of Practise reviews. The coordinating group is chaired by the Director for the Department of Datacapture. SN has invited the major register owners to enter into top level agreements with SN. SN has made a template for the agreements. The template can be adjusted to the specific needs and characteristics of each register and it owner. It was pointed out that the public sector as a whole will benefit from improved data quality and each data owner will experience improved quality in its products and services. After the agreements are signed, they will be reviewed annually at top level.

The agreement on co-operation covers the following items:

1. General conditions
    - 1.1. Background
    - 1.2. Purpose
    - 1.3. Appendixes to the agreement
        - 1.3.1. Which registers are covered by the agreement
        - 1.3.2. Contact persons for the day-to-day contact
        - 1.3.3. A description of the quality improvements
        - 1.3.4. Practical arrangements for transferring the register
        - 1.3.5. Possible other issues
    - 1.4. Duration
    - 1.5. Legal basis for extradition of the data
2. Duties of the parties
    - 2.1. General demands
    - 2.2. The duties of Statistics Norway
        - 2.2.1. Data quality
        - 2.2.2. Quality reports from Statistics Norway
    - 2.3. The duties of the register owner
        - 2.3.1. Data quality
        - 2.3.2. The duty to coordinate
3. Delivery of the data to SN
4. Meetings
5. Secrecy
6. Costs
7. Contact persons for the agreement

A template for the agreements can be obtained from the author.

The agreements are supplemented by a quality report for each register. The quality report is based on the quality indicators from the Blue-Ets Work Package 4 (Daas and Ossen 2011). While the Blue-Ets Work Package 4 will use a quantitative approach for the quality indicators, the quality report will use a simpler descriptive approach to describe the quality indicators. The indicators were grouped in the following five dimensions of quality: Technical checks, Accuracy, Completeness, Integrability, and Time-related dimensions.

The quality report includes the following items:

1. Purpose/background
2. Updating
3. Co-operation
4. Contact persons
   4.1. For the register owner
   4.2. For SN
5. General quality indicators (at unit and variable level, specific problems are highlighted)
   5.1. Technical checks
   5.2. Accuracy
   5.3. Completeness
   5.4. Time-related checks
   5.5. Integrability
6. Transfer of the register data to SN
7. Estimated number of hours used on
   7.1. Data capture
   7.2. Automatic micro checking and editing of the data
   7.3. Manual editing
8. Feedback at micro level
9. Feedback from matched data, at aggregated level
10. Proposals for other types of feedback
11. Proposals for quality control at the source by the register owner
12. Proposals for other measures which will improve register quality
13. Proposals for other measures which will improve co-operation on register quality
14. What can be done to simplify reporting from respondents to the administrative register, to reduce response burden
15. Proposals for further development of the register content

An example of a quality report can be obtained from the author.

Forums for co-operation between SN and register owners

SN is very active in the Co-operation Forum and the User Forum for the administrative Central Coordinating Register for Legal Entities (CCRLE). The legal basis for this type of co-operation is the Statistics Act:

> *§ 3-2. Administrative data-processing systems*
> *(1) Statistics Norway shall have the right to use administrative data-processing systems in the state administration and in nationwide municipal organisations as the basis for official statistics.*
> *(2) When state bodies or nationwide municipal organisations are to establish or modify a major administrative data-processing system, notice thereof shall be*

> *sent in advance to Statistics Norway. Statistics Norway may seek additional information. Statistics Norway may also put forward proposals concerning the manner in which data-processing systems should be designed in order to safeguard considerations for statistics.*

In short: SN has the legal 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. In practise, when SN wants to develop a new source for statistical purposes, SN refers to § 3-2 and to § 2-2 on the obligation to provide information.

With regards to the CCRLE, SN is in a special position. In section 2 of the Act on the CCRLE[2], SN's Business and Enterprise Register is specifically mentioned as one of the affiliated registers to the CCRLE. Furthermore the act lists the variables which are being shared openly among all the affiliated registers. This facilitates the transfer of a limited range of individual based information from SN to the CCRLE, not violating the conditions on secrecy in the Statistics Act. Such information might come from a questionnaire, from a telephone conversation or some other source.

To organise the data exchange between the CCRLE and the affiliated registers (including SN's Business register) in practise, two bodies where established:

- The Co-operation Forum – to facilitate the data production and exchange between the affiliated registers
- The User Forum – a broader forum where a variety of users of the data from the CCRLE meet. SN participates among other things to inform on new statistical use of the register

SN participates actively in both, as a member and occasionally in the chair. The participation from SN is appreciated highly because statisticians tend to have the best general overview, and to have the skills, methods, tools and time for quality assessment and checks.

The approach with a co-operation forum and a user forum is proposed for the other national base registers as well (Cadastre and the CPR). Currently there is a user forum for the CPR which meets twice per year. SN has two representatives in the CPR user form.

There is an example of how SN is involved in the development of administrative registers. During 2006 and 2007 the Ministry of Finance initiated a working group to formulate proposals for the modernization of the administrative CPR. SN was invited to participate in the working group. Someone in the working group proposed to exclude SN from a future co-operation forum. This idea was given up immediately when the representative for SN mentioned § 3-2 from the Statistics Act. SN is currently a very active and highly appreciated participant in the Program for Modernization of the Population Register under the Tax Administration.

Conclusion

Extensive use of registerbased data will inevitably improve data quality, especially when the data is integrated with data from other sources. Statistical offices can highlight quality issues, helping the data owners to take measures to avoid such problems. This will be for the benefit of statistics, but also for other users of the data. An important conclusion is

[2] http://www.brreg.no/english/acts/

that communication on quality issues between data owners and statistical offices will improve quality for the benefit of statisticians and administrative users.

Communication is the key word for good quality statistics from registers. The success of the register based approach in Norway is the result of good communication, co-ordination and co-operation, internally in SN, between SN and register owners, and among the owners of administrative registers. The co-ordinating group in SN which is described in the paper is working closely together, combining different types of qualifications and expertise. Note that there is a solid legal basis for the approach and the legal adviser is a central person in the approach.

Improvement of input data quality is a never ending story when it comes to the production of register based statistics. However, there will always be a need for good methodological solutions to bridge the gap between good quality data and good quality statistics from registers. For various reasons, the available or observed values may differ from the ideal measures that are of statistical interest for the target units. Zhang (2011b) explores topics of statistical theory for register based statistics to help understand and deal with these quality issues.

REFERENCES

Daas, P. and S. Ossen 2011: Report on methods preferred for the quality indicators of administrative data sources. *Deliverable 4.2 BLUE-Enterprise and Trade Statistics. European Commission European Research Area, Seventh framework programme.*

Eurostat 2011: European Statistical Code of Practice for the National and Community Statistical Authorities. *Adopted by the European Statistical System Committee, 28th September 2011*

Hendriks, C. and J. Åmberg 2011: Building and maintaining quality in register populations. *Paper to the 58th ISI Congress (Dublin, Ireland, 21-26 August 2011)*

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 2007: Register based statistics in the Nordic countries. Review on best practices with focus on population and social statistics.

Zhang, Li-Chun 2011a: A Unit-Error Theory for Register-Based Household Statistics. *Journal of Official Statistics, Vol. 27, No. 3, 2011, pp. 415–432*

Zhang, Li-Chun 2011b: Topics of statistical theory for register-based statistics. *Paper to the 58th ISI Congress (Dublin, Ireland, 21-26 August 2011)*

Keywords: Register based statistics, Co-operation on register quality, Improved input data quality