UNITED NATIONS

E

Economic and Social Council

Distr.
GENERAL

ECE/CES/GE.41/2007/3
22 March 2007

Original: ENGLISH

ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Group of Experts on Population and Housing Censuses

Tenth session
Astana, 4-6 June 2007
Item 3 (a) of the provisional agenda

CENSUS PLANNING AND MANAGEMENT

Census planning and quality assurance strategy in a register-based census

Submitted by Sweden*

This meeting is organised jointly with Eurostat.

Summary

The Bureau of the Conference of European Statisticians (CES), at its meeting held in Washington, D.C. (United States) on 19-20 October 2006, approved the renewed terms of reference for the Steering Group on Population and Housing Censuses and the plan for future CES activities on population and housing censuses. The CES Bureau also agreed that the Steering Group would coordinate the work on the diverse types of meetings. The present paper was prepared on request by the Steering Group on Population and Housing Censuses, for presentation and discussion at the Joint UNECE/Eurostat Meeting on Population and Housing Censuses in Astana (Kazakhstan), 4-6 June 2007. The paper provides substantive basis for the discussion in the session of the meeting dedicated to "Census planning and management".

* This document has been prepared by Statistics Sweden.

I. BACKGROUND

A. Before Census 1990

1. Sweden has taken Population and Housing Census every fifth year since 1930. Before that it took a Census every tenth year. The first Population Census was taken in 1850. The Population and housing censuses were produced in Sweden before 1990 by combining data collection from administrative and statistical registers with mail-out, mail-back questionnaires to all persons over 16 years old and to owner of multi-dwelling houses. The questionnaires were used to complete the census topics with those topics, which were not possible to collect with registers.

2. The first time Sweden used any administrative register was in 1975. Before that it only used questionnaires. The amounts of questions in the questionnaire have been reduced for every census after that, because more and more data have been possible to collect with registers. Statistics Sweden always try to find possibilities to collect data with administrative or statistical registers to reduce the response burden.

B. After Census 1990

3. The last time Statistics Sweden used any questionnaires in a Population and Housing Census was in 1990. Since then no Population and Housing Census has been taken or will be taken.

4. It is important to know that the Swedish statistical system have all census-like statistics, as populations statistics, education statistics and statistics on the economic activity and can produce them yearly and even more often for the users. The only census-like core statistics that is lacking today is statistics for households and dwellings.

5. That means that instead of doing Population and Housing Censuses every fifth or tenth year, Statistics Sweden can produce census-like statistics every year for the users and do not need to put them together at special occasions and call them a Census.

6. The Government and Parliament then took a decision in 1995 that the next census must be totally register-based. Administrative registers were still lacking for the households and dwellings. The plan was to develop them until Census 2001, but that was not possible because no decision was possible to take in the Parliament.

7. The plan is today to develop a Dwelling register and create households by using administrative register for the next Population and Housing Census in 2011. The law was taken by the Parliament on 17 May 2006 and the ordinance is now expected.

II. CENSUS PLANNING

A. National needs

8. To take a census or more precisely to establish a new Dwelling register needs a new law that has to pass the Parliament. To do a census planning when you have to relay on direct

political decisions is not very easy. Especially when the political parties are not agree on that a census is really needed and the majority in the Parliament has changed over time as four elections have taken place after 1990.

9. Statistics Sweden has changed the plans many times since 1990 and made a lot of investigation. Many stop's and go's have passed and all the times Statistics Sweden tried to change the content and methods to be able to convince the politics that this is the most efficient way to collect and produce the most needed census data.

10. A census planning takes many years to come through. The international organisations start many years before the census date to make up their plans and recommendations.

11. The Census planning normally should start with contacting the users to see what needs they will have in the future. In the case of Sweden, it is more likely that it will first be investigated which registers will be in place when the census will be taken, and from that what kind of needs will be covered with data from these registers.

12. Sometimes Statistics Sweden need to create its own register, but normally it re-use an administrative register from another Authority.

B. International needs

13. It is quite complicated, as a register country, to meet all requests from among others the international institutions. A register system based on administrative registers is rather stable and it is normally not possible to change that content to meet the needs of the National Statistical Institute (NSI).

14. It is more likely that Statistics Sweden as a National Statistical Institute needs to change its data collection because the producer of that administrative register wants to make some changes in the register because something has changed in their environment.

15. That means that a lot of new topics that might come up in the discussion during international meetings, that might be rather easy to put in a questionnaire, could be very complicated or perhaps impossible to collect with an administrative register as the producer of the administrative register do not see the need to collect that topic.

16. That is way Sweden, as a register country, at some international meetings seem to have very little interest in developing new topics or change some topics to cover more details.

III. QUALITY ASSURANCE STRATEGY

17. Today Statistics Sweden does not have a common quality assurance strategy for registers that is used for all statistical registers or for the Population and Housing Census, but it has some works going on. The qualities in the registers are today described in many different ways.

18. Statistics Sweden is in the position today to have organised a special unit "Register coordination and Microdata" that will have the responsibility for developing a quality assurance strategy for the registers. The way of working is also changing and becoming more process-

based (instead of stow-piped), and that will put focus on among other thing the quality in the registers.

19. Work has also started to create a Register-system that will keep and create connections with all registers at Statistics Sweden and soon there will be a structured input data warehouse where all data from the registers will be loaded. That data warehouse needs a quality assurance strategy to make the warehouse usable for internal users and in parts for external users.

20. Below some extracts are presented of a chapter from a new book that has been published in March 2007. The authors are Mr. Anders Wallgren and Ms. Britt Wallgren, working in the Unit of Register coordination and Microdata at Statistics Sweden. The title of the book is "Register-based Statistics, Administrative Data for Statistical Purposes", and it is published in Wiley Series in Survey Methodology. The chapter discuss Quality of Register-Based Statistics and it is reproduced below with the permission of Mr. and Ms. Wallgren.

A. Specific quality issues for register-based statistics?

21. What is statistical quality when it comes to register-based statistics? Platek and Särndal (2001) state that the statisticians lag behind in building a theory of accuracy assessment for statistics based on registers, and that a theory is needed.

22. Holt (2001) points out that there are important differences between statistics from surveys with their own data collection on the one hand, and statistics from administrative sources on the other. Holt maintains that the most important aspect of quality when it comes to register-based statistics is not accuracy, but relevance.

23. Nanopoulos (2001) maintains that countries like Denmark that have well integrated register systems need a conceptual apparatus regarding errors in statistics, which will be different from that required by countries that mainly carry out sample surveys and censuses.

24. The conclusion is that it is important to consider the following when discussing the quality of register statistics:

(a) It is necessary to distinguish between surveys with their own data collection and ones that are register-based. Otherwise, there is a risk of uncritically using the traditional error models developed for sample surveys and censuses;

(b) There should also be a distinction between the quality of a register-based survey and that of a statistical register, as a register has many possible uses.

25. Sample surveys and censuses are carried out with *one* particular use in mind and quality issues generally focus on the estimates carried out. In the case of a statistical register, *many* different uses are possible – such a register may serve not only current surveys but also future ones.

26. Similar to other surveys, the quality of a register-based survey also relates to *one* specific use of the register and also focuses on the quality of the estimates, particularly their relevance and accuracy in relation to the purpose of the survey. Describing quality is here a question of indicating whether the quality of the survey is good or bad.

27. However, the quality of the statistical register itself is not related to one particular use and, when describing quality in this respect, it is important to indicate what characteristics the register has, thereby implying the uses to which it may be put.

28. The quality of the register will affect the quality of the surveys based upon it, and is determined by three factors:

(a) the administrative systems on which the register is based;

(b) the possibilities offered by the system of statistical registers with regard to improving coverage, content of variables and consistency; and

(c) the processing done to produce the register.

29. The first factor that determines the quality of the register is the administrative systems upon which it is based. Administrative systems are generally unique – the administrative data collection in the case of population registration, for instance, is altogether different from that for enterprises' statements of earnings. Data collection within the administrative systems is also usually different from the data collection of a statistical office. Even though the administrative systems involve respondents filling in forms, the reporting of information has its own conditions, and is governed by administrative rules and regulations.

30. The second factor that determines the quality of the register relates to the possibilities that the register system offers. The register should be co-ordinated with the rest of the register system, and the system as a whole should function efficiently.

31. The third factor that determines the quality of the register is the processing performed when the register was created. How was the register population defined, how was the content of variables determined and how were the data edited?

32. The administrative system and the internal processing of data have hitherto been perceived as specific to each register, which has hampered the exchange of experience and the development of methodology, but the quality of one particular register affects other register-based surveys that have a use for that register.

33. Thus, for example, missing values of the variable industrial classification in the Business Register constitutes a problem for all the registers and surveys that include this variable. Furthermore, all others are dependent on this problem being solved by those responsible for the Business Register, because otherwise statistics from different registers will not be consistent.

B. Detailed knowledge of a register's characteristics

34. The purpose of quality assurance is to investigate and remedy quality defects, which can occur in different parts of the work on a register. The traditional error model follows the different steps of a sample survey. In corresponding fashion, the quality of register-based statistics is described with the aid of groups of indicators for the different parts of the work of creating a register.

1. The phase of determining research objectives – its effect on the register

35. What statistical needs and requirements is the register supposed to fulfil? Have contacts been established with advanced users or researchers? Will the statistical office carry out its own advanced analyses and reports? If it is known what is required of a register, it is then possible to build up an understanding of the uses to which the register may be put. Advanced users of statistics are an important group when looking at the development and application of a register. They often have valuable experiences and ideas, which ought to be documented.

2. The inventory phase – how has it affected the register?

36. The inventory phase looks at the different sources that have been used to produce the new register. This can consist of administrative sources and statistical registers that already exist within the system. Do these sources, on the whole, have a rich content? If several sources have been integrated into the register, this will be of advantage to new users. The inventory should also investigate whether there are other sources connected to the area of study that have not been used. This can be an opportunity to carry out an active search for new sources.

3. Are any changes planned?

37. Are there plans for any changes or improvements regarding the register population, the object definitions or the content of variables? If so, it is a sign both of defects in the present register and of the existence of development work designed to increase the usability of the register.

4. Supplier contacts and editing

38. How does contact with the supplier occur? These contacts should be deep enough for the supplier to understand the needs of statistics and for the recipient to understand the conditions governing the administrative system. The contacts should be regularly renewed, so that the recipient gets information about coming changes.

5. The integration phase – how has the object set been created?

39. In this phase, it should be analysed how existing sources have been processed to ensure that the new register contains the desired object set. This involve the editing of administrative data, the matching of different sources and selection of objects or the processing of time references to produce the object set for the designated point in time or period.

6. The integration phase – how have the objects been defined?

40. What processing has been done in order to check and change the object definitions? For example, have the administrative data been checked and adjusted so that the definitions are those that are required? Have derived objects been formed in the new register? Is the quality of the object definitions checked? Are register maintenance questionnaires or evaluations conducted? How large are the errors?

7. The integration phase – how have the variables been created?

41. This phase should look at what processing has been done to produce the intended variables. This can include looking at whether the variables in the administrative sources have been edited and also at the scope and treatment of missing values. The various sources, from which the variables have been imported, should also be documented.

42. The scope of any possible errors should also be investigated, such as measurement errors or errors of classification in the spanning variables. The methods used to detect errors could include sample surveys or special evaluation surveys. Focus groups and cognitive interviews analysing the administrative forms could also be conducted to discover sources of measurement error.

8. Documentation as a part of quality assurance

43. Documentation work is also an important part of quality assurance. Incorrect and uncritical use of administrative data can be prevented using metadata that provide information about problems of comparability. Changes in the administrative system cause these kinds of problems, and must therefore be documented. There is otherwise a risk of arriving at incorrect conclusions.

44. Since it is possible for a statistical register to be used by various users working with the register system, the documentation of registers must be done in such a way as to make it accessible to all.

45. Metadata play a major role in the work on register-based statistics. In the case of the integration of different registers, it is necessary to know the definitions, and what the comparability problems are. It is important that the processing methods should be documented as well, to facilitate the development of methodology and the exchange of experience.

C. Overall appraisal of quality

46. Register-based statistics can sometimes be used through the calculation of one summarising value: “Employment in the manufacturing industry in the municipality has increased by 1.6 per cent”.

47. But often a number of large tables with hundreds of estimates are studied in an effort to discover patterns of relationship and of time series patterns. The patterns found give rise to conclusions: “The labour structure in the municipality has become worse with respect to age and level of education, hampering the high-tech sector”.

48. Such a conclusion is based on many comparisons with other municipalities and over time. Here it is impossible to conduct a discussion in terms of numerical errors in all of the hundreds of estimates on which the conclusion is based. With every error being divisible into many components each of which is to be judged on its own, the situation appears more daunting than ever. What the user wants to know is whether the conclusion in question can be drawn or not. Are there any statistical pitfalls that the user has not thought of?

49. On the basis of detailed knowledge of the characteristics of a register, an overall appraisal can be made of the quality of the register and the quality of the register-based survey.

50. The quality of the register should be described in general terms, so that potential users can see whether it suits their purposes. The description should relate to the various areas of application that may be of interest. Three ways of using registers and the corresponding quality aspects can be distinguished:

(a) Cross-sectional quality: what comparisons can be made within the register?

(b) Time series quality: what comparisons can be made over time on the aggregated level?

(c) Longitudinal quality: what comparisons can be made at micro level over time?

51. The quality of a register-based survey should be described for one particular use of the register. Is the quality of the estimates good or bad for this intended use? The relevance and accuracy of the estimates should be described.

1. An overall appraisal of the quality of a register

52. An overall appraisal of the quality of a statistical register should contain the following:

(a) Definitions of the register population and variables used must be available and easy to understand;

(b) What comparisons can be made within the register?
Example: If, in the area of regional statistics, there is a desire to compare municipalities, is the item non-response roughly the same in all municipalities or is it possible to make comparisons using imputed values?;

(c) What comparisons can be made over time on an aggregated level?
Example: Is the item non-response roughly the same for all years or can comparisons be made over time using imputed values? Have there been changes in administrative data that make comparisons over time more difficult? Has the register processing been carried out in order to increase the comparability between years?;

(d) What comparisons can be made on micro level over time? Correct longitudinal comparisons place the highest demands on a register. What checking and processing has been done in order to make this possible?

53. Such an overall appraisal calls for thorough knowledge of the register, which means that detailed documentation is necessary.

2. An overall appraisal of the quality of a register-based survey

54. An overall appraisal of the quality of a register-based survey should contain the answers to the following questions:

(a) How is the survey's target population defined? Is the definition adequate in view of the purpose of the survey? Are there any important differences between the target population and the register population?;

(b) How are the survey's variables defined? Are the definitions adequate in view of the purpose of the survey? This discussion is directed towards the variables that are of greatest importance for the survey, the most important aggregating and spanning variables on which the analysis is based;

(c) What comparisons are to be made in the survey? Is the register of sufficient quality for these comparisons?;

(d) Are the estimates made in the best way? For the sake of both cross-sectional and time series quality, it is important that multi-valued variables and missing values are dealt with using appropriate methods of estimation. In the case of time series quality, it is also important that level shifts in time series are linked;

(e) Are the results interpreted reasonably in the light of the uncertainty of the estimates? Tables based on statistical registers are not to be regarded as offering exact information – there can be random variation and other sources of error, so it is important to be careful of not over-interpreting the results.

55. Such an overall appraisal calls for knowledge of the quality of the register and for thorough subject matter knowledge. This means that the person carrying out the appraisal must be well acquainted with the research objectives in question.

D. Main quality issues in different kinds of surveys

56. How well the data collection process works, primarily determines the quality of a sample survey or census. This means that sampling errors, measurement errors and non-response errors are important quality issues here. As frame populations generally are based on early available but less reliable sources, coverage errors are a more serious problems in surveys based on data collection than in register-based surveys.

57. In register-based surveys administrative data and registers are used for statistical purposes. Administrative registers are processed so that objects sets, object types and variables meet statistical needs. The definitions of register population, objects and variables in a statistical register determine the relevance errors of the register-based surveys that use the register. These relevance errors can be judged from the documentation or the register. Many different sources are integrated when statistical registers are created. The quality of the linkage variables, the sources and the methods used, determine the integration errors in the new register.

58. In the opinion of the authors, the main quality issues for register-based statistics are relevance errors and integration errors.

IV. CONCLUSION

59. In addition to these coverage errors, relevance error can also occur if the definition used is not adequate. The difference between the population of interest and the target population is one of the survey's relevance errors. There are likely between 25 000–50 000 persons registered in the population register in Sweden that do not live permanently in Sweden. It is judged that 4–8 per cent of immigrants from outside the Nordic countries have left Sweden without reporting this. This relevance error affects statistical estimates describing the death rates, average income, etc. for immigrants from outside the Nordic countries so that the estimates become misleading.

60. In the example above, it can be seen that Statistics Sweden's population statistics use an administrative definition, the registered population, when defining the survey's target population. Administrative concepts always give definitions that are functional. It is sensible statistical practice that these administrative concepts are used to define the target population if the relevance errors are small. However, the basic rule is that the population definition should attempt to meet the demands of the statistical survey. If the administrative concepts are not sufficiently relevant or adequate, it is necessary to develop own definitions and carry out the register processing required so that the register's object set reflects the defined population as closely as possible.

61. Is it possible to leave the present administrative definition of Sweden's population? This is discussed within Statistics Sweden and there are ways to improve the definition of the target population:

(a) Include foreign students studying at Swedish universities, administrative data is available. They are not registered by the Tax Board, but by the universities;
(b) Exclude Swedish students studying abroad, they are registered by the Tax Board, but do not live permanently in Sweden. Administrative data is available for many of them;
(c) Swedish university students can in many cases be registered by the Tax Board where they lived before going to the university. Their present address is registered by the universities. This would result in a geographical reallocation of many university students; regional population statistics will be different.

62. This means that it is possible to leave the administrative definition of the target population and introduce a new definition, which will be more adequate.

*****