STATISTICAL COMMISSION and UN ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

<u>UNECE Seminar on New Methods for Population Censuses</u>
Organized in cooperation with UNFPA
(Geneva, 22 November 2004)

Session 1 – Invited paper

THE DUTCH VIRTUAL CENSUS 2001:
A NEW APPROACH BY COMBINING DIFFERENT SOURCES

Submitted by Statistics Netherlands [*]

1.   SUMMARY

1.   Data from many different sources were combined to produce the Dutch Census tables of 2001. Since the last Census based on a complete enumeration was held in 1971, the willingness of the population to participate has fallen sharply. Statistics Netherlands found an alternative in the Virtual Census, using available registers and surveys. The Virtual Census is cheaper, comparable to earlier Dutch Censuses, and more socially acceptable. The Netherlands takes up a unique position in the European Census Round. The table results are not only comparable with the earlier Dutch Censuses but also with those of the other countries in the 2001 Census Round.

2.   In producing tables for the 1981 and 1991 Census rounds, the focus in the Netherlands was on combining population register data with results from the Labour Force and Housing Surveys. Less care was given to overall consistency of the table estimates. For the 2001 Census Eurostat and other international organisations required more detailed information than for earlier Census Rounds. Moreover, in the last decade Statistics Netherlands has acquired more and more experience in dealing with data of various administrative registers for statistical use. This enabled the development of a Social Statistical Database (SSD), which contains coherent and detailed demographic and socio-economic statistical information on persons and households. The Population Register forms the backbone of the SSD. Sample surveys are still needed for information that is not available from registers. Examples of variables that are not available in the Dutch registers are level of education and occupation. However, these two variables are available in the Dutch Labour Force Survey.

3.   Overall numerical consistency among all tables in the Census tables set of 2001 is required. This need stimulated methodologists at Statistics Netherlands to develop a new

---

[*] Paper prepared by Eric Schulte Nordholt. The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

estimation method that ensures numerically consistent table sets if the data are obtained from different data sources. The method is called repeated weighting, and is based on the repeated application of the regression method to eliminate numerical inconsistencies among table estimates from different sources. The newly developed method of repeated weighting guarantees that combining survey and register information leads to consistent estimates in the tables of the Census Programme. The Dutch Census tables 2001 have been estimated by making use of all available information and therefore the best possible quality of the estimates has been obtained.

Keywords: Census; consistent table estimates; repeated weighting

2.   INTRODUCTION

4.   In 2003 data were combined to produce the Dutch 2001 Census tables. In the Netherlands this was not done by interviewing inhabitants in a complete enumeration, but by using data that Statistics Netherlands already had available. This way, the Dutch tax payer got a much lower census bill. The costs for a traditional census would be about three hundred million Euros, while the costs made now are 'only' about three million. The estimate includes the costs for all preparatory work such as developing a new methodology and accompanying software. The costs of the registers are not included, but the analyses of the results are. Registers are not kept up-to-date for censuses but for other purposes. Saving money on census costs is only possible in countries that have sufficient register information. As an example we can compare the costs of the Dutch Virtual Census of 2001 with the costs of the traditional Census that was held in Canada. In Canada the census costs amounted to approximately 450 million Euros. Canada has about 31.6 million inhabitants, twice as many as the Netherlands. Statistics Canada justifies the huge census costs by pointing out the enormous implications of the census results for the distribution of money among regions. Moreover, a virtual census would be impossible in Canada because of the lack of sufficient register data.

5.   The 2001 Census relates to forty extensive tables. Twenty-eight are about the Netherlands as a whole, nine are at the COROP level (NUTS 3) and three at municipal level (NUTS 5). The forty tables fall into a number of groups. Eight tables concern housing, two tables concern commuting and the other thirty tables are demographic tables, relating to occupation, level of education and economic activity. Additionally, demographic, housing and labour figures are compiled at sub-city district level for ten large cities that participate in Urban Audit II (Statistics Netherlands, 2003). These ten large cities are Amsterdam, Rotterdam, The Hague, Utrecht, Eindhoven, Tilburg, Groningen, Enschede, Arnhem and Heerlen.

6.   Except the financial aspect, other important differences exist between a traditional census and the virtual census conducted in the Netherlands. In spite of the mandatory character of a traditional census, a certain part of the population will not participate (unit non-response) and the part that does participate will not answer some questions (item non-response). Correcting non-response by weighting and imputation techniques is well worth trying. A well-known problem with traditional censuses is that participation is limited and selective. Traditional correction methods fall short of the need to be able to publish reliable results. The last traditional Census in the Netherlands (in 1971) met with much privacy objections against the collection of integral information about the population living in the Netherlands. This increased the non-response problem and the expectation was that non-response would be even higher if another traditional census were held in the Netherlands (Corbey, 1994). There are almost no objections to a virtual census and the non-response problem only plays a role in the surveys of which the data are used.

If non-response can be corrected in a survey, it will certainly be possible to correct for the selectivity of that survey in the census where it is used.

7.   The Virtual Census in the Netherlands was off to a later start than in other countries where a traditional Census was conducted. It did not make sense to really start the 2001 Census Project until all sources were available; some registers were available relatively late. Nevertheless, the Netherlands was quicker with the compilation of the forty census tables than most of the other countries that participated in the 2001 Census Round. In fact, the Netherlands was one of the first to send the complete set of forty tables to Eurostat, which co-ordinated the contributions of all European Union (EU) member states, accession countries and European Free Trade Association (EFTA) member states. The Netherlands had the advantage that the incoming census forms did not need to be checked and corrected. However, one must realise that for some variables only sample information is available, which implies that it was impossible to meet the level of detail required in some Dutch tables.

8.   An interesting option for the future is to use small area estimation techniques to estimate the cell values that could not be estimated adequately. A theoretical framework for small area estimation can be found in Rao (2003). The ONS studied the application of this technique in the context of its Neighbourhood Statistics Programme. This is a major initiative to bring together and make widely available statistics on a small area level. In each case of implementation of indirect small area estimates particular attention was paid to model specification. Some experimental synthetic estimates were published in the United Kingdom and others are undergoing a process of evaluation. Possibly, the techniques of repeated weighting and small area estimation can be combined in the 2011 Census Round.

9.   Currently, the advantages of the virtual census in cost and non-response problems amply make up for the loss of some detail compared to a traditional census. Moreover, not all information required will always be available for the users in traditional censuses. This is because traditional correction methods such as weighting and imputation sometimes do not correct for limited and selective participation. This means no reliable results can be published for some of the cells in the set of tables. One may wonder why simply applying mass imputation (filling in valid values for all missing scores) is not taken into account to overcome these problems. An important advantage of mass imputation is that once the records are imputed, any user will be able to reproduce results when using the same imputed file. However, mass imputation is not a viable strategy for raising survey outcomes to population totals. There are not enough degrees of freedom to sustain a sufficiently rich imputation model accounting for all significant data patterns between sample and register variables. Only if the interest is in totals of subsets of the population defined by the explanatory variables in the model, the imputation approach leads to approximately design-unbiased and hence reliable estimates (at least if the variances are reasonably small) (Kroese and Renssen, 2000).

10.   The Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) have more variables available in registers than the Netherlands. So the problem of insufficient detail in the outcome does not play a major role there. Moreover, some Nordic countries conduct a (limited) enumeration for variables missing in the registers. Most of the other countries are in a similar position as the Netherlands where some variables relevant for the census can be found in registers, while other variables are available on a sample basis only. That's why much interest exists in the Dutch approach to combine registers and surveys and to use modern statistical techniques and accompanying software to compile the tables. Last months British, French, Australian and New Zealand colleagues visited Statistics Netherlands to learn more about the Dutch approach. It is of course crucial that statistical bureaus may make use of registers that are

relevant for the census. For Statistics Netherlands this was laid down in the new statistical law that came into force in the beginning of 2004. Nevertheless, in the years to come Statistics Netherlands will have to establish good contacts with register holders. Timely deliveries with relevant variables for Statistics Netherlands are crucial for statistical production.

11.   The reason why Statistics Netherlands has compiled the set of tables is a gentlemen's agreement. In 1991 the Census Act was rescinded, officially cancelling Statistics Netherlands' obligation to hold a census once every ten years (Corbey, 1994). There is no European obligation to supply census data, but it is almost inconceivable that the Netherlands would not compile census data for the international organisations just like all other European countries do. Eurostat has a co-ordinating role in collecting harmonised data on the EU and a duty to make international comparisons of the outcome.

12.   It will be several years before all countries participating in the 2001 Census Round have sent their final set of tables to Eurostat. Therefore, Statistics Netherlands took the initiative to compare the 2001 results of a limited number of European countries. The results of the Dutch 2001 Census were also compared to earlier Dutch Censuses. Such work has been carried out in the past as well. The data compiled on 1981 and 1991 were much less detailed than the set of tables of the 2001 Census. The 1991 Dutch Census was largely based on a register count of the population in combination with the Labour Force Survey 1991 and the Housing Demand Survey 1989/1990. Contrary to 1981 and 1991, Statistics Netherlands has published census information over 2001 on the municipal level.

3.   METHOD OF COMPILING

13.   The current virtual census relates to 2001. The backbone of the census is the central Population Register (PR), which is the combination of all municipal population registers. PR data of 1 January 2001 were used as the basis for the set of tables. The set of tables focuses on frequency counts and not on quantitative information. The SSD datasets on 2001 (the Social Statistical Database include integrated microdata on employees and self-employed) were not available on time for the 2001 Census. Therefore, we used datasets of 2000 that were available in the beginning of 2003 to deduce the individual data of the end 2000 as an approximation for the situation on 1 January 2001. Different variables, such as occupation and level of education, were obtained from the Labour Force Survey (LFS). The variable job size was obtained from the large Survey on Employment and Earnings (SEE). To obtain sufficient records, information on persons from the LFS 2000 and the LFS 2001 was combined. For the housing tables we used PR data of 1 January 2001, the Housing Register 2001 and the Survey on Housing Conditions (SHC) 2000. For the tables on commuting we used the PR data of 2000 and 2001, the SEE 2000 and the SSD datasets of 2000.

14.   Some variables of the PR and SSD datasets are available on an integral basis. Examples are age, sex, marital and employment status. Survey variables are only available for a part of the population. Examples are the highest level of education attained (LFS) and whether someone rents or owns the property they live in (SHC). We guaranteed consistency among the tables by using the technique of repeated weighting. The method of repeated weighting has been described extensively in Houbiers et al. (2003) and Houbiers (2004). It is based on the repeated application of the regression estimator and generates a new set of weights for each table estimated. The results of five simulation studies testing various aspects of repeated weighting can be found in Van Duin and Snijders (2003). When using repeated weighting, the weights of the records in the microdata are adapted in such a way that a new table estimate is consistent with all earlier table estimates.

15.   To apply the technique of repeated weighting we used the latest version of the software package VRD developed by Statistics Netherlands. The letters VRD stand for Vullen (Filling) Reference Database and the aim of the application is to fill and manage the reference database. The main functions of VRD are the estimating of tables via repeated weighting, adding these tables to the reference database, and withdrawing aggregates from the reference database. Under the condition of small, independent samples, the variances of the table values can also be estimated. The estimating of the tables does not occur in VRD itself, but takes place in Bascula 4.0 automatically without the VRD-user seeing this explicitly. Estimating the tables and the variances can be done in the batch or interactively.

16.   To be able to estimate every table as accurately as possible, every estimate is based on the largest possible number of records. Tables that contain register variables only, are counted from the registers. Tables that contain at least one variable from a survey are estimated from the largest possible combination of registers and surveys. The combination of registers and surveys form blocks from which the census tables have been estimated. By way of illustration six blocks have been displayed below on basis of which the census tables for the economic active population (employed and unemployed together) were estimated.

1. The register block.
2. The NACE block (all records from the register block for which the international code for economic activity NACE is known, and also the non-employed).
3. The SEE block.
4. The cross-section between NACE and SEE blocks.
5. The Economic Activity block (in fact this is the LFS block, supplemented with information on the employed and retired).
6. The LFS block.

17.   Blocks 2 up to and including 6 were compiled on the basis of survey data. To produce estimates for the complete population, weights have to be determined. These weights depend on:
– the precise composition of the block concerned (one or more surveys);
– the design of the survey(s);
– the non-response correction of the survey(s);
– the reduction of the variance by means of auxiliary information;
– the reaching of consistency.

18.   Complete consistency is not always possible, for example if too many restrictions were imposed. In some cases complete consistency is possible, but it leads to a very large variation in the weights and thus increases variance drastically. In those cases it is better to restrict the detail that is published.

19.   In compiling the census tables we adapted the weights of the blocks at every VRD turn by means of all relevant register counts and the tables estimated earlier from the blocks. This way, all tables are mutually consistent. Every table has to be calculated from the largest block from which the table can be determined. If all tables are estimated this way with the correct weights, the tables' results are mutually consistent. By starting every time from the largest block, the most detailed possible census tables have been achieved.

20.   The figures of the 2001 Census relate to persons living in the Netherlands on 1 January 2001 (counting unit persons). The persons who were living in the Netherlands at the beginning of that day according to the PR were 'counted' in the Virtual Census. Most of the Dutch population

lives in private households, the others are part of institutional households. The number of employees in the tables relate to the end of the year 2000 for which 22 December 2000 was used as reference date to fix the number of jobs of employees in the Netherlands. It was impossible to have a reference day in 2001 for the number of employees since the SSD datasets 2001 were not available on time to use in the 2001 Census. The SSD data used registers information on the jobs of employees. If an employee holds several jobs at the same time, he or she can appear several times in the employee register. In the set of tables the features of the main job are used, in which the main job of an employee has been defined as the job with the highest gross wage for the social insurances.

21.   The 2001 Census was compiled partly on the basis of sample data. Therefore, margins of inaccuracy have to be taken into account for some results of the 2001 Census. Because of the reliability of the results, rules of thumb are being applied for cell values that are based on a sample from the census population. The exact margins of inaccuracy cannot be given because of composing blocks from the surveys and the complex design of these surveys. The rules of thumb have been deduced on the basis of the assumptions that the two LFS datasets (for 2000 and 2001) form one sample and that the 'inclusion probabilities' for this sample were given by the block weights of the LFS block. The rules of thumb for records of observations from the LFS run as follows:
–   Table cells based on less than 10 persons are always suppressed.
–   Table cells based on 25 or more persons are always published.
–   Table cells based on 10–24 persons are only published if they form a part of a breakdown (by age or sex), in which no cells based on less than 10 persons occur, and at least 50 percent of the cells in the breakdown have more than 25 persons. The threshold of 25 persons corresponds to an estimated relative inaccuracy of at most 20 percent (i.e. the estimated margins amount to 40 percent at most).

22.   The rules of thumb for records from the SHC are of the same form. However, somewhat higher threshold values are applied because of the fact that the sample size of the SHC is somewhat more limited than the one of the LFS. For table cells with households or dwellings as counting unit, analogous rules of thumb are applied for the Dutch Census of 2001.

4.   KEY RESULTS OF THE 2001 CENSUS IN THE NETHERLANDS

4.1   Population by sex, age and type of household

23.   At the start of 2001 a total of 16.0 million people were living in the Netherlands, 7.9 million male and 8.1 million female. In the age categories 0-14 and 15-74 year there were some more males than females, but in the category 75 year and older there were almost twice as many women than men. Most people live in private households. More than 200 thousand people lived in institutional households, such as health care institutions and institutions for retired and elderly people. About 36 percent of this group was male and 64 percent female. Of the people in institutional households 57 percent was over 75. This group is dominated by women. More information about the population by sex, type of household and age group can be found in Table 1.

Table 1. Population by sex, type of household and age group

Sex and type of household	All ages	Age in years		
		0-14	15-74	75+
Total population	15,985,538	2,977,283	12,036,171	972,084
Male	7,909,052	1,522,811	6,047,425	338,816
Female	8,076,486	1,454,472	5,988,746	633,268
		0-14	15-74	75+
Population in private households	15,766,606	2,970,545	11,947,996	848,065
Male	7,829,914	1,518,611	5,998,189	313,114
Female	7,936,692	1,451,934	5,949,807	534,951
		0-14	15-74	75+
Population in institutional households	218,932	6,738	88,175	124,019
Male	79,138	4,200	49,236	25,702
Female	139,794	2,538	38,939	98,317

4.2   Population by economic activity

24.   At the start of 2001 just under half of the people living in the Netherlands belonged to the economically active population (labour force). The working labour force included 7.4 million people: 6.8 million are employees and 0.6 million self-employed. The unemployed labour force comprised almost 200 thousand people. In the organisational set-up of the census, employees, the self-employed and unemployed are mutually exclusive categories. Self-employed people who also work a number of hours a week for pay are counted as employees. Someone in the working labour force cannot be unemployed at the same time. The number of unemployed is estimated on the basis of sample information.

25.   Of the economically active population 58 percent was male, while of the economically inactive population 58 percent was female. The economically inactive include attendants at educational institutions, retired people and people engaged in family duties. The number of housewives is more than 18 times the number of househusbands. More information about the population by economic activity and sex can be found in Table 2.

Table 2. Population by economic activity and sex

Economic activity	Total	Male	Female
Economic active population	7,586,914	4,388,239	3,198,675
Working	7,394,777	4,287,967	3,106,810
Employed	6,786,511	3,883,813	2,902,698
Self-employed	608,266	404,154	204,112
Unemployed	192,137	100,272	91,865
Economic inactive population	8,398,624	3,520,813	4,877,811
15-74	4,449,257	1,659,186	2,790,071
Attendant at educational institutions	640,446	342,934	297,512
Retired	1,355,940	620,493	735,447
Engaged in family duties	1,270,420	65,821	1,204,599
Other economically inactive	1,182,451	629,938	552,513

Working population by branch of economic activity

26.   The 7.4 million members of the working population can be divided by branch of economic activity by means of the NACE (Nomenc lature statistique des Activités dans la Communauté Européenne) code. For an employee who has more than one job we took the features of his or her main job. In the context of the Dutch 2001 Census, the main job of a person has been defined as the job that yielded the highest wage for the social insurances in 2000. Counted this way, the Netherlands had 0.2 million people working in agriculture and fishing, 1.5 million in manufacturing and construction and 5.7 million in services at the start of 2001. Of those working in services, 3.5 million worked in commercial services and over 2.1 million in non-commercial services.

Employees by working hours

27.   An interesting phenomenon is how many hours a week employees work in their main job. Of the almost 6.8 million employees in the Netherlands 4.2 million employees work full-time (35 or more hours a week), 1.8 million employees have a long part-time job (less than 35 hours, but at least 15 hours a week) and 0.8 million have a short part-time job (less than 15 hours per week). Of those working full-time 77 percent is male, while of the part-timers 75 percent is female. More information about the working hours and sex of employees can be found in Table 3.

Table 3. Employees by working hours and sex

Employees by working hours	Total	Male	Female
Employees	6,786,511	3,883,813	2,902,698
Full-time (=35 hours a week)	4,222,228	3,236,504	985,724
Part-time total	2,564,283	647,309	1,916,974
Long part-time (15-<35 hours a week)	1,793,656	419,071	1,374,585
Short part-time (<15 hours a week)	770,627	228,238	542,389

4.3   Working population by occupation

28.   Working people can be classified by occupation by means of the International Standard Classification of Occupations (ISCO).

For men the most common occupation categories in 2001 were:
- professionals;
- legislators, senior officials and managers;
- craft and related trades workers.

For women the occupation categories were:
- technicians and associate professionals;
- clerks;
- service workers and shop and market sales workers.

More information about the working population by occupation and sex can be found in Table 4.

Table 4. Working population by occupation and sex

Working population by occupation		Total	Male	Female
Working population		7,394,777	4,287,967	3,106,810
1	legislators, senior officials and managers	926,631	695,563	231,068
2	professionals	1,205,163	705,357	499,805
3	technicians and associate professionals	1,248,759	607,819	640,939
4	clerks	841,219	271,862	569,358
5	service workers and shop and market sales workers	800,629	259,173	541,456
6	skilled agricultural and fishery workers	105,256	78,280	26,976
7	craft and related trades workers	712,093	677,256	34,837
8	plant and machine operators and assemblers	446,722	398,845	47,877
9	elementary occupations	522,901	272,435	250,467
0	armed forces	37,032	34,227	2,805
99	occupation unknown	548,374	287,151	261,223

4.4   Population by level of education

29.   The population living in the Netherlands can be classified by level of education by means of the International Standard Classification of Education (ISCED). Actually, it is the highest level of educational attainment that determines the category by which a person is classified in the ISCED. Of the 12.0 million people aged between 15 and 75 years the most common level of education is the secondary level. The number of people with a tertiary level of education is larger than the number of people with a primary level of education. For the group aged over 75 the secondary level is also the most common, but there are also considerably more people with a primary than with a tertiary level of education. More information about the population by level of education and age group can be found in Table 5.

Table 5.Population by level of education and age group

Population by level of education	All ages	0-14	15-74	75+
Total population	15,985,538	2,977,283	12,036,171	972,084
No education at all	1,244,031	1,244,031	0	0
Pre-primary education (ISCED 0)	1,370,511	1,198,580	154,832	17,098
Primary education (ISCED 1)	2,787,104	534,672	1,825,655	426,778
Lower secondary education (ISCED 2)	3,145,529		2,924,405	221,125
Upper secondary education (ISCED 3c)	2,711,384		2,566,372	145,012
Upper secondary education (ISCED 3b)				
Upper secondary education (ISCED 3a)	1,873,656		1,828,072	45,584
Post secondary non-tertiary education (ISCED4)	483,684		468,699	14,985
First stage of tertiary education (ISCED5b)	247,194		238,029	9,165
First stage of tertiary education (ISCED5a)	2,081,590		1,992,670	88,920
Second stage of tertiary education (ISCED6)	32,760		31,082	1,678
Education unknown	8,094	0	6,356	1,738

5.   THE 2001 CENSUS COMPARED TO EARLIER DUTCH CENSUSES

30.   The first Census in the Netherlands was held in 1795 for the purpose of establishing voting constituencies. At that time the united provinces of the Netherlands were still a republic and the borders were different from the current borders. After Napoleon the Netherlands became a kingdom and once every ten years a census was held. The first Census in the Kingdom of the Netherlands was held in 1829. Before Statistics Netherlands was established, another six Censuses were held in 1839, 1849, 1859, 1869, 1879 and 1889 under the responsibility of the Ministry of the Interior. In 1899 Statistics Netherlands was established and was put directly in charge of the eighth Census. In the 20th century six more traditional Censuses were carried out in 1909, 1920, 1930, 1947, 1960 and 1971. The three most recent Censuses (1981, 1991 and 2001) were not based on a complete enumeration but on registers and surveys available for Statistics Netherlands.

31.   Originally, the censuses had two aims. First, they were meant to correct errors in the municipal population registers. Second, they were used to get extra information about the socio-economic phenomena in the country. Since the Netherlands conducts a register-based census, the first aim no longer exists. Also, the quality of the central Population Register (PR), which unites all municipality population registers, has improved considerably over time. This is because the incentive for municipalities to keep their population registers up-to-date is the allocation of central government funds among municipalities. This is mainly based on the population size according to the local registers. Another reason is that it is extremely difficult to function in Dutch society without being included in the PR. So both municipalities and citizens have enough incentives to keep the PR of good quality. Recent actions in Rotterdam to improve the quality of the municipal population register for some old quarters prove this statement. The second aim is still valid and many census results are published in a historical or international context. Currently, census data are popular for comparisons among countries.

32.   Table 6 presents some key results of the Dutch Censuses in the period 1829-2001. Remarkable is the ageing of the Dutch population, especially in the post-war period.

Table 6. Population by age group in the period 1829-2001

Census		All ages	Age group		
Number	Year		0-19	20-64	65+
		× 1,000	in % of the total population		
1	1829	2,613.3	44	50	5
2	1839	2,860.6	45	50	5
3	1849	3,056.9	43	53	5
4	1859	3,309.1	42	53	5
5	1869	3,579.5	43	52	6
6	1879	4,012.7	44	50	5
7	1889	4,511.4	45	49	6
8	1899	5,104.1	44	50	6
9	1909	5,858.2	44	50	6
10	1920	6,865.3	42	52	6
11	1930	7,935.6	40	54	6
12	1947	9,625.5	38	55	7
13	1960	11,462.0	39	53	9
14	1971	13,060.1	36	54	10
15	1981	14,216.9	31	57	12
16	1991	15,070.0	25	62	13
17	2001	15,985.5	24	62	14

6.   THE DUTCH 2001 CENSUS COMPARED TO OTHER COUNTRIES

33.   More than fifty countries participated in the 2001 Census Round. Most countries chose a day in 2001 as their reference day, although they chose many different days. As it will take a long time before all countries finish the tables required by the international organisations, the Netherlands took the initiative to make some simple comparisons among nine countries that were relatively quick in compiling the set of tables for Eurostat and willing to join the comparison analyses.

34.   The calculations in this paper are the author' own and are based on the set of standard tables produced from census data for Eurostat by nine different countries. It is expected that there are definitional differences among the countries which will affect comparisons. Also the statistics produced by the author do not necessarily reflect the way countries usually choose to present their data. Furthermore, they may not be considered in some cases to be the definitive source of information on these topics. The nine countries are the Netherlands (NL), Norway (NO), Sweden (SE), Finland (FI), Estonia (EE), Switzerland (CH), Slovenia (SI), Greece (GR) and the United Kingdom (UK). The nine countries differ in size, but all except the United Kingdom have a fairly small number of inhabitants compared to France and Germany.

35.   The nine countries are members of the European Union (EU) or the European Free Trade Association (EFTA). The Netherlands joined the European Community at the start in 1958, the United Kingdom joined in 1973 and Greece in 1981. The European Community became the European Union in 1995 when Sweden and Finland joined. Estonia, the most northern of the ten accession countries, and Slovenia, the most northern part of the former Yugoslavia, joined the EU

in 2004. Norway and Switzerland are EFTA members and work closely together with the EU countries. Norway is also a member of the European Economic Area (EEA). The EEA agreement came into force on 1 January 1994. EEA countries are the EU 15, Norway, Iceland and Liechtenstein. Switzerland did not join the EEA, but works together with the EU countries on a bilateral basis. Statistics is one of the issues on which the EEA countries work together. The aim of the statistical co-operation in the EEA is to build a European Statistical System that gives a coherent and comparable description of the economic, social and environmental developments in the EEA countries.

36.   The nine countries that are compared have different census reference dates: 31 March 2000 (Estonia), 5 December 2000 (Switzerland), 1 January 2001 (The Netherlands, Sweden and Finland), 18 March 2001 (Greece), 29 April 2001 (United Kingdom), 3 November 2001 (Norway) and 31 March 2002 (Slovenia).

37.   Table 7 presents the estimated costs of the 2001 Censuses, and the population and the area of the nine countries. Estonia, Slovenia, Greece and the United Kingdom held traditional censuses; Switzerland used a combination of a traditional census and register information to produce the census tables. Norway relied largely on registers, but conducted a census for some missing housing variables. Sweden and Finland held entirely register-based censuses and the Netherlands performed a virtual census based on existing registers and surveys. The 2001 Census costs for Norway, Estonia, Switzerland, Slovenia, Greece and the United Kingdom include enumeration costs. In the Netherlands, Sweden and Finland such enumeration costs do not exist for the 2001 Census, so the costs presented in Table 7 for these three countries are rough indicators of the extra costs of producing census tables for the international organisations and of analysing and publishing the results. Table 7 shows that the costs per inhabitant in those countries that required completion of a census form for the census were much higher than the countries that did not have enumeration costs. In Table 7 the population densities among the nine countries can be compared. The Netherlands has the highest population density, followed by the United Kingdom and Switzerland. The population density in the Nordic countries (Norway, Sweden and Finland) and in Estonia is relatively low. Slovenia and Greece occupy a middle position.

Table 7. Comparison of nine countries according to the 2001 Census results

	NL	NO	SE	FI	EE	CH	SI	GR	UK
Cost of the Census (in millions of Euros)	3.0	14.6	1.0	0.8	10.2	99.1	8.0	49.7	367.4
Population (× 1,000,000)	16.0	4.5	8.9	5.2	1.4	7.3	2.0	10.9	58.8
Area (× 1,000 km²)	41.5	323.9	450.0	338.1	45.1	41.3	20.3	132.0	244.1
Cost of the Census per inhabitant (in Euros)	0.2	3.2	0.1	0.2	7.3	13.6	4.0	4.6	6.2
Population density (persons per km²)	386	14	20	15	31	177	99	83	241

38.   Table 8 presents some simple demographic comparisons of 2001 Census data for the nine countries. The data presented in Table 8 are calculated from the tables produced for the international organisations and sent to Eurostat. They were checked by the different countries that produced the tables. Remarkable differences exist among the nine countries. The percentage of women is by far the highest in Estonia. The percentage of the population that is single is high in the Nordic countries, low in Switzerland and especially low in Greece. For the percentage of singles the Netherlands, Estonia, Slovenia and the United Kingdom occupy a middle position.

Non-nationals form a high percentage of the population in Estonia and Switzerland, where it is relatively difficult to get the nationality. The population of Finland and Slovenia contain an extremely low percentage of non-nationals. In Slovenia the non-national population includes only those with stated foreign citizenship. Finland also has an extremely low percentage of people born outside the country. Nationality data is not available for the UK, because data on citizenship is not collected in the UK Census. In most countries the percentage of the population born abroad is greater than the percentage that are non-nationals. However, not all non-nationals are persons born outside the country, i.e. these groups are not nested.

Table 8. A demographic comparison according to the 2001 Census results

	NL	NO	SE	FI	EE	CH	SI	GR	UK
in % of the total population									
Percentage of women	50.5	50.4	50.5	51.2	53.9	51.0	51.2	50.5	51.4
Percentage of singles	44.7	48.4	49.8	47.1	44.1	42.1	44.5	39.7	44.3
Percentage of non-nationals	4.2	4.1	5.4	1.8	20.0	20.5	1.9	7.0	.
Percentage of people born outside the country	10.1	6.9	11.3	2.6	19.2	21.6	8.6	10.3	8.3

39.   Table 9 presents two indicators to make a simple economic comparison of the nine countries. Different definitions and ways of collecting the data may hamper the comparisons. The percentage of the population that is economically inactive is relatively high in Greece. For this indicator Switzerland has the lowest score. When we compare the unemployed, as percentage of the economically active population, distinct groups can be discerned. Finland, Estonia and Greece have relatively many unemployed people, whereas in the Netherlands, Norway, Sweden and Switzerland unemployment was low. Slovenia and the United Kingdom occupy a middle position.

Table 9. An economic comparison according to the 2001 Census results

	NL	NO	SE	FI	EE	CH	SI	GR	UK
in % of the total population									
Economically inactive population	52.5	48.5	51.8	50.8	53.3	45.8	51.7	57.8	52.1
in % of the economically active population									
Unemployed persons	2.5	2.3	3.8	12.5	13.8	4.0	6.7	11.0	5.7

40.   Table 10 presents two educational indicators to compare the nine countries. Different definitions and ways of collecting the data may hamper the comparisons as well in Table 10. All countries have difficulties fitting their national education classification into the ISCED codes. For all countries compared except Estonia and Slovenia the OECD (2003) study also gives some information about the education indicators in 2001. A somewhat different population is considered and therefore the absolute results are hard to compare. However, the relative differences in percentages of people with primary and tertiary education in the OECD study agree to a reasonable extent to the results of the censuses. Norway has a high quality education register and an extremely low percentage of the population aged 15-74 with a primary level of education or less. Also Slovenia has a very low percentage of primary educated people. Finland, Switzerland, Greece and the United Kingdom have a relatively high percentage of people with primary education. The percentage of the population aged 15-74 with a tertiary level of education shows less variation among the nine countries than the percentage with a primary level of

education. Estonia, Finland and Norway have the highest percentages and the United Kingdom, Greece and Slovenia the lowest. A more detailed comparison of the UK and Netherlands census data can be found in Chapter 12 of Schulte Nordholt et al. (2004).

Table 10. An educational comparison according to the 2001 Census results

	NL	NO	SE	FI	EE	CH	SI	GR	UK
in % of the total population									
Population 15-74 with a primary level of education or less[a]	16.5	4.1	15.6	37.2	11.3	33.5	5.5	38.1	37.0
Population 15-74 with a tertiary level of education	18.8	22.9	20.2	24.6	24.8	17.5	13.3	14.4	14.6

[a]: Including level of education unknown.

7.   CONCLUSIONS

41.   The virtual census has proved to be a successful concept in the Netherlands. It has many advantages compared to traditional censuses. The costs are now considerably lower and nevertheless data on the Netherlands have come available that could be compared to results of earlier Dutch censuses and to the results of other countries that took part in the 2001 Census Round. It was the third time that the Netherlands conducted a virtual census. However, the Dutch data that have been compiled on 1981 and 1991 were of a much more limited character than the set of tables of the 2001 Census. Moreover, they were largely based on a register count of the population in combination with the then existing surveys about the labour force and housing conditions.

42.   The technique of repeated weighting has been used successfully to produce a consistent set of tables for the 2001 Census. Every table was calculated from the largest block from which the table could be determined. All tables have been estimated this way for the 2001 Census with the correct weights, and therefore the tables' results are mutually consistent. By starting every time from the largest block, the most detailed possible census tables have been achieved. Before compiling tables with this new technique, micro integration of the different sources in the SSD remains important. In the micro-integration process the data are checked and incorrect data are adapted. It is strongly believed that micro-integrated data will provide more reliable results, because they are based on a maximum amount of information. Also the coverage of subpopulations will be better, because when data are missing in one source, another source can be used. Another advantage of micro-integration and repeated weighting is that there is no reason for confusion among users of statistical information anymore, because there will be one figure on each socio-economic phenomenon, instead of several figures depending on which sources have been used.

43.   It is possible to use the technique of repeated weighting in other countries as well. However, first one needs to have the possibility to use registers for statistical purposes. One must realise that in most countries not all census variables can be derived from register information. Additional surveying then remains a necessity, but a consistent set of census tables can be produced using the technique of repeated weighting.

44.   At the end of 2003 the complete set of forty census tables for the Netherlands was sent to Eurostat. Afterwards, a book about the Dutch Virtual Census of 2001 was written (Schulte Nordholt et al., 2004). This book provides a wide-ranging description of the socio-demographic

and socio-economic state of the Netherlands based on the 2001 Census results. It discusses differences in size and composition among households, economic activity of households, individual activity status by region, age, education level and branch of economic activity. There are separate chapters on the economic activities of young people and people of retirement age. The economic activities, levels of education and occupation of foreigners from various countries of origin are compared with each other and with the native Dutch population. Regional aspects are also examined, including commuting. The results of the 2001 Census are compared with the Census results of other European countries and with earlier Dutch Censuses. Lastly, the Virtual Census methodology used is described in some detail.

45.   The PDF version of the book can be found at the Statistics Netherlands website, at page http://www.cbs.nl/en/publications/recent/census-2001/b-57-2001.htm. Here also an extra Chapter (number 15) is ava ilable with an overview of the used data sources, methods and definitions. Hard copies of the book were sent to all authors of the book, to the management of Statistics Netherlands and to several libraries. The book was also offered to the Prime Minister, the Minister of Economic Affairs and the Minister of Education, Cultural Affairs and Science of the Netherlands and to Director-Generals of statistical offices in several countries. In August 2004 the book was publicly released at an official presentation in the Statistics Netherlands' office in Voorburg. The research process and the main findings were then presented to an audience of academics, press representatives, government officials, as well as Statistics Netherlands' employees. Since the publication of this book several articles were written in national and regional newspapers about the Dutch Virtual Census of 2001 and its results. Announcements and interviews appeared in several mailing lists, newsletters and journals. For further information on the Dutch Virtual Census of 2001 (including the set of forty standard tables for the Netherlands in Excel format), see http://www.cbs.nl/en/publications/articles/general/census -2001/census-2001.htm.

8.   GENERAL RECOMMENDATIONS FOR FUTURE CENSUS ROUNDS

46.   In this section, based on the experiences of the Dutch Virtual Census of 2001, some recommendations are made for future Census Rounds. It is very useful to have Census guidelines and a table program, but some errors in this program could be corrected and some aspects have to be described in some more detail for the next Census Round. It looks like the guidelines have been written with the idea in mind that all countries conduct a traditional census, but more and more countries have chosen other options. It would be useful to take that aspect into account when revising the guidelines.

47.   Some variables are partly described in the guidelines and for further information one is referred to other literature. This is a bit confusing. The risk is that different countries have interpreted these variable definitions differently. This would hamper the comparability among countries and therefore it is better either to describe a variable fully or to refer completely to other literature.

48.   The number of different classifications for one variable (e.g. age) is sometimes too large. This implies that sometimes cells in tables cannot be estimated consistently, whereas less estimation problems would be occurred if the number of different classifications would have been reduced. This problem is much more severe if non-nested classifications of the same variable are used.

49.   What to do with people who have different economic activities at the same time? In a traditional census one can ask the respondent about his or her main economic activity. In a virtual

census one has to find a criterion to choose one economic activity per person. If different countries use different criteria, the comparability of the results will be hampered. In the Netherlands we could not distinguish between employers and other employed people and therefore these two self-employed categories have been merged. People who were at the same time employed and self-employed were counted as employees. As we have no register of unemployed people, this group is the only part of the economically active population that has to be estimated. Therefore, sometimes the number of economically active people is an estimate, although we know exactly how many people are employee or self-employed. People who are at the same time both economically active and student are counted as economically active people. The population of retired people is not clearly defined. What to do with people who are partly retired? And what with people who live from their property instead of from their pension? We prefer to pay more attention to the variable country of birth and less to the variable citizenship in the next Census Round. Different countries have namely different policies towards changing nationality. Therefore, Table 8 in this paper shows remarkable differences between these two variables among countries. International more relevant comparisons can be made by distinguishing first and second generation foreigners on the basis of the variables country of birth of the person and his or her parents. This is done in the PAU (Population Activities Unit of the UN) tables and Chapter six of Schulte Nordholt et al. (2004). In the family nuclei it would be useful to add extra categories for same sex (married or cohabiting) couples. This is a growing group that has to be taken into account in the next Census Rounds. For the variable family status the rest category 'child with other family status' is missing.

50.   The NUTS classifications change over time. Therefore, it is crucial to provide the countries with the table lay-outs including the classifications of the Census year. The list of country codes has to be improved and used consistently in all tables. Now only a general unknown category is included. This implies that for countries not in the list all the records are counted as unknowns, which implies that the totals per continent become incorrect. What to do with no longer existing countries not in the list? This is especially relevant for the variable country of birth if the original country was split.

51.   For the NACE codes it is useful to add a separate category unknown. Now we have attributed the unknowns proportionally to the other categories. Also for the variable occupation a category unknown should be introduced. We have included the unknowns in the total, but that implies that the sum of the occupation categories is often smaller than the total. Finally, for the ISCED a category unknown is included in the table, but here we merged the unknowns with the categories no education at all and level of education pre-primary as we could not distinguish among these three categories.

Remarks about individual tables of the standard set

Table 3: Persons by country of citizenship which hold citizenship of a country which fell apart in different states (for example Soviet-Union became Russia, Ukraine etc) are classified by the formerly existing country of citizenship (Soviet-Union). The information available in the PR does not provide sufficient information in those cases to determine the country of citizenship according to actual countries of citizenship existing on 1 January 2001.

Table 4: Persons born in countries which no longer exist are if possible classified as being born in the country in which that country has changed. For example persons born in the German Democratic Republic are classified as being born in Germany. Persons born in the Dutch East Indies are classified as being born in Indonesia. For countries which have disintegrated this is not possible. In that case the person is classified by the former country of birth (for example,

Yugoslavia instead of Croatia). The information available in the PR in those cases does not provide sufficient information to determine the country of birth according to actual countries of birth existing on 1 January 2001.

Table 5: In determining the municipality of residence one year prior to the Census municipalities are classified according to the municipal boundaries on 1 January 2001. Thus the information on persons living on the same address on 1 January 2000 and 1 January 2001 are classified consistently. This principle is also applied for municipalities which stopped existing in the year 2000.

Table 6: The information to fill the category 'engaged in family duties' (part of economic activity) was derived from the Labour Force Survey by means of personal questions about the social status of a person. Because no decisive criterion could be used, the distribution between this category and the category 'other economically inactive' is sometimes rather fanciful.

Table 15: In this table of the set of standard census tables some errors appeared in the names of the occupation categories according to the ISCO:
– 'seniors officials' should be 'senior officials';
– 'Skill agricultural' should be 'Skilled agricultural';
– 'Craft and relative trades' should be 'Craft and related trades'.

Table 20 and Table 40 (a, b and c): The source of these tables is the Labour Force Survey in combination with the Survey on Employment and Earnings, while information about employers was derived from SSD data sets. The data refer to the end of the year 2000. No information is available for people with a fixed place of work outside the Netherlands. Employees without a fixed place of work are included in the category employees with a fixed place of work, with place of head office as place of work. Employees of which even the head office (of the business where they work) is not known are assumed to have the place of residence as the place of work. Those employees were mainly self-employed persons for which it is not unreasonable to assume that their place of residence and work coincide. Finally, we have to mention that NUTS 4 divisions do not exist (they would not make sense) in the Netherlands and that no data are available for work at home.

Table 38d: In part G (Economically active population by status of employment and industry (branch of economic activity)) of this table the column 'Unemployed' contains very few numerical values. This is caused by the fact that no register exists of unemployed persons in the Netherlands. For many municipalities not enough records in the Labour Force Survey are available to produce reliable estimates for the unemployed. For some of the smaller municipalities the 'Total economically active population' (part E) is equal to the number of 'Employed' (part G). In those cases zero unemployed were found in the Labour Force Surveys of 2000 and 2001. It is likely that for those small municipalities the unemployment is biased downwards.

Main adjustments made in the lay-out of the set of standard tables

Table 3, 4 and 5: In several cases the list of '21 Remainder of New Independent States' has been complemented. In the same Tables the following code numbers have been uniformed: 153 into 151, 155 into 152, 156 into 153 and 157 into 154 (all 'European New Independent States').

Table 7: The codes 14 and 15 of the variable 'country of citizenship' have changed places underneath 'males all' and 'females all'. The classification is all right underneath 'both sexes all'.

The classification is changed so that code 14 means 'Central Europe' and code 15 means 'European New Independent States' all over the Table.

Table 8: In the classification of 'living in a family' the column 'child' was missing. This code was mentioned in the Guidelines provided by Eurostat, so it is added to the table.

Table 9: In the description of this Table, provided by Eurostat, is written that this Table concerns the female resident population aged 15 and over living in private households. However, the phrase 'living in private households' was not mentioned in the title of Table 9 and therefore we adjusted the title.
In the part 'with partner of spouse' the column 'without child' is missing. It is added analogously to the part 'without partner or spouse'.

Table 11: The category 'unemployed' was only included underneath 'both sexes all'. It is added underneath the other parts as well. In the same Table underneath 'males aged 15-20' the category 'Legislators, senior officials and managers (ISCO-COM1)' was included twice. One of the lines has been removed.

Table 14: A new line is added to compensate for the cases in which the household type could not be stated: '13 with other family status'.

Table 16: The table outline for this Table used erroneously ISCO 88 titles and not those of ISCO 88 COM. The titles where changed into ISCO 88 COM titles.

Table 18: Underneath 'females' the NACE-codes 410 and 930 have been removed, because these codes do not occur underneath 'total' and 'males'. The codes 41 and 93 have been maintained.

Table 24: Underneath 'type of private household' a new line is added: '4 household type not stated'.

Table 30: The title for the age group '75-80' is changed into '75-79'.

Tables 29, 30, 31, 32, 33, 34, 35, 36 and 37: Two of the NUTS 3 areas are missing. 'Utrecht' (code 311) and 'Flevoland' (code 231) have been added to the tables.

Tables 38, 39 and 40: The NUTS 5 classification supplied by Eurostat is the classification of 1 January 1999. The tables have been adjusted to the classification of 1 January 2001.

References

1. Corbey, P., 1994. Exit the population Census. *Netherlands Official Statistics*, Volume 9, summer 1994, pp. 41-44.

2. Duin, C. van and V. Snijders, 2003. Simulation studies of repeated weighting. Discussion paper 03008, Statistics Netherlands, Voorburg / Heerlen. http://www.cbs.nl/en/publications/articles/general/discussion-papers/discussion-paper-03008.pdf.

3. Houbiers, M., 2004. Towards a social statistical database and unified estimates at Statistics Netherlands. *Journal of Official Statistics*, Volume 20, No. 1, pp. 55-75.

4. Houbiers, M., P.Knottnerus, A.H. Kroese, R.H. Renssen and V. Snijders, 2003. Estimating consistent table sets: position paper on repeated weighting. Discussion paper 03005, Statistics Netherlands, Voorburg

/ Heerlen. http://www.cbs.nl/en/publications/articles/general/discussion-papers/discussion-paper-03005.pdf.

5. Kroese, A.H. and R. H. Renssen, 2000. New applications of old weighting techniques, constructing a consistent set of estimates based on data from different sources. ICES II, Proceedings of the second international conference on establishment surveys, survey methods for businesses, farms, and institutions, invited papers, June 17-21, 2000, Buffalo, New York, American Statistical Association, Alexandria, Virginia, United States, pp. 831-840.

6. OECD, 2003. Education at a glance. OECD Publications, Paris, France.

7. Rao, J.N.K., 2003. Small area estimation. Wiley, New York, United States.

8. Schulte Nordholt, E., M. Hartgers and R. Gircour (Eds.), 2004. The Dutch Virtual Census of 2001, Analysis and Methodology, Statistics Netherlands, Voorburg / Heerlen, July, 2004. http://www.cbs.nl/en/publications/recent/census-2001/b-57-2001.htm.

9. Statistics Netherlands, 2003. Urban Audit II, the implementation in the Netherlands. Report, BPA no. 2192-03-SAV/II, Statistics Netherlands, Voorburg. http://www.cbs.nl/en/publications/articles/regional/urban-audit-II-Netherlands.pdf.

*****