STATISTICAL COMMISSION and
UN ECONOMIC COMMISSION FOR
EUROPE

CONFERENCE OF EUROPEAN
STATISTICIANS

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

Session 1– Supporting paper

THE NEW METHOD OF THE NEXT GERMAN POPULATION CENSUS

Submitted by Federal Statistical Office, Germany[*]

1.   INTRODUCTION

1.   After World War II, 4 population and housing censuses were carried out in Germany: in 1950, 1961, 1970 and 1987. This was done in a traditional way by interviewing inhabitants in a complete enumeration. The last census scheduled for spring 1981 had to be postponed twice. The first time, it was for cost reasons: the Federal States (Länder) demanded a contribution of the Federal Government towards the expenses of the municipalities for the field operations. The second time, the census law was cancelled by the Federal Constitutional Court in spring 1983. In autumn 1982 a politically motivated boycott movement against the population census started and resulted in numerous (about 1 200) complaints of citizens against the constitutionality of the census law. Unexpectedly the Supreme Court cancelled the census law. The main reason for this verdict was that the envisaged transmission of census microdata to the municipalities for the adjustment of the population registers was considered to be an infraction of the general right of privacy. This verdict is still in force. It means that administrative data may be transmitted from the administration authorities to the statistical offices, but it is not allowed to transmit data on individual persons, which are collected in a statistical survey, back to the municipalities e.g. for the purpose of adjusting incorrect entries in the municipal population registers.

2.   The fear of the political decision-makers that a new complete enumeration of the citizens

_____

[*] Paper prepared by Johann Szenzenstein.

might prompt boycott movements like that in the eighties and the estimated high costs of a traditional census (about 1 billion Euro) led to the decision in the late nineties not to conduct a traditional census with enumerators any more, but to evaluate existing administrative registers for the next census. The Statistic al Offices of the Federation and the Länder were charged with the task of developing a model of a register-based census which should make it possible to produce for every citizen a data record containing all typical core variables of a census.

2.   THE NEWLY DEVELOPED MODEL OF A REGISTER-BASED CENSUS

3.   The most important element in the new approach is the use of a combination of administrative registers and surveys (complete enumerations) as data sources. Geographic, demographic and family characteristics of persons can be drawn from population registers. Economic and educational characteristics of employees can be obtained from registers maintained by the Federal Employment Agency and from other administrative registers. Information on the characteristics of dwellings and buildings has to be collected by a (postal) survey from the house owners.

2.1   The main data sources

a) Population registers

4.   The backbone of a register-based census are the population registers (PRs) which are kept decentrally by the municipalities. Each of the currently about 13 500 municipalities maintains a PR and every person living in the municipality (as his/her main or secondary place of residence) is legally obliged to register there. Normally the citizens have a major incentive to be registered in the PR of their current place of residence as this e.g. is a prerequisite for getting an ID card (the possession of which is obligatory for all adult citizens) or an income-tax card (which the employees have to provide to their emplo yers) or to be registered in the voters register. Even though the PRs seek to record every person, they are by no means perfect. People may remove (to another town or inside the town) and forget to notify the removal to the registration authorities. Another example of an improper registration in the PR is when the registration office cannot cancel a person in the PR because the name indicated in the de-registration form cannot be found in the PR (e.g. owing to a clerical error of the name). Therefore PRs are not always up-to-date.

5.   An important population group not registered in the PR are people who live in Germany without the authorities' knowledge, most of them staying illegally. These persons are not registered, but it is very unlikely that they would be enumerated in a traditional census.

6.   The German census programme is mainly based on the "de jure" concept. This means that people have to be counted at the place where they usually reside. The PRs meet the requirements of the "de jure" concept because they register the citizens at their main place of residence, that is that place of residence where they live for the most days of a week. There is only one exception from this definition: a married person who has a secondary place of residence from which this person e.g. goes to work for 5 days a week is registered with his/her main place of residence at the place where his/her family usually resides.

7.   For each person residing in the municipality, the PRs store the following information:

- ♦ residential status: sole, main or secondary place of residence
- ♦ basic demographic variables: sex, age (date of birth), marital status, country of citizenship; place/ country of birth, religious affiliation;
- ♦ geographic variables: place(s) of residence (at the reference day of the census),
- ♦ information which can be used to group persons together to form families and households: surnames (family name, name at birth, previous family name), first names, name of the spouse, names of the children, name of the mother (only for persons aged under 27), name of the father (only for persons aged under 27), the address of the previous place of residence, date of moving into the current dwelling.

8.   The characteristics and their items stored in the municipal PRs are largely standardised by federal regulations.

b) Employee registers

9.   The Federal Employment Agency (FEA) maintains registers covering:

- ♦ all persons in employment (wage earners and salaried employees) subject to obligatory social insurance ,
- ♦ all persons registered as unemployed at the labour administration and
- ♦ all persons attending a vocational training programme of the labour administration.

10.   These registers serve as sources for the following economic variables:

- ♦ activity status on the reference day,
- ♦ status in employment (apprentice, wage earner, salaried employee),
- ♦ full/part time employment,
- ♦ occupation (currently practised),
- ♦ industry (branch of economic activity),
- ♦ place of work,
- ♦ highest educational level achieved.

11.   In addition the registers contain information that is indispensable for linking data records to the PR records (name, sex, date of birth and place of residence).

12.   Employment data on civil servants, army personnel and judges, who are not covered by the registers of the FEA, can be obtained from other administrative registers.

13.   But there are no registers for self-employed persons. That means, with the evaluation of the employees' registers employment data would be available only for about 90% of all persons in employment. To get employment data on self-employed persons it is proposed to collect that information by a supple mentary sample survey.

c) Housing census

14.   Following the international recommendations that a census should provide not only statistical data relating to the demographic and economic situation of the population but also information on their housing conditions, it was de cided to include a housing census in the new model. As there are no nation-wide registers for buildings and dwellings, the housing variables have to be collected by a postal survey contacting the owners of the buildings.

15.   The owners (or the managers of the buildings) have to provide the following housing variables:

- ♦   for buildings: period of construction, number of dwellings, type of building (residential building, non-residential building with dwellings)
- ♦   for dwellings: occupancy status, tenure status, number of rooms/housing space, kitchen, bathing/toilet facilities, type of heating, monthly (net and gross) rent

16.   The housing census has to serve two additional purposes, namely to provide information which can be used for:

- ♦   grouping persons together to form households and
- ♦   checking and improving the data from the PRs.

17.   For these purposes the following variables are collected from the owners/ managers: number of persons (tenants or owner-occupiers) residing in a dwelling, the names of one or two occupants and the date of their moving in (the way in which these variables are used for generating households is depicted in the following chapter).

2.2   Combining data sources

18.   Like the former traditional census, the next register-based census will comprise statistical data relating to the demographic, economic and social characteristics of persons and households and to the characteristics on housing at national, regional and local levels. For this purpose data from the different data sources have to be combined. In the German model of a register-based census, the following types of combining data have to be carried out:

- ♦   merging the decentrally kept municipal PRs to form a Central Population Register (CPR).
- ♦   linking data records of the CPR to the data records of the employees registers.
- ♦   atching data of the housing census with the data of the CPR to group persons together to form households.

19.   This process should finally result in one data record for every person comprising all census variables which have been identified as necessary and essential for the users (core census variables) and a data recor d for every household. A major problem of these combining processes is the fact that linkage of microdata has to be carried out without a uniform linkage key in the form of a personal identification number being available. In Germany such ID numbers are not permitted for legal reasons. The data records have to be linked through the characteristics: name, sex, date of birth/place of birth and address.

a) Merging the municipal population registers

20.   The PRs of the about 13 500 municipalities are maintained at some 1 500 computer centres. Although there are uniform standards for data records and the stored variables and their characteristics are standardised by federal regulations, each municipal PR has to be checked if the records are in line with the federal standards before merging PRs to form a CPR. As the PRs are kept decentrally by the municipalities, it cannot be ruled out that some persons are registered in two (or even more) municipalities (with their sole or main place of residence). With regard to such double (or multiple) entries, two types have to be distinguished:

♦   A person is registered in two PRs only for a limited time period. This is always the case when a person moves to another municipality. According to the registration regulations, a person has to fill in the registration form only at the new place of residence; this registration authority at the new place of residence must automatically transmit the person's data needed for deregistration to the registration authority at the previous place of residence. Currently it can take some time (up to several weeks) before the whole registration/deregistration process is finished.

♦   A person is registered permanently in two PRs. This may occur, for example, when a person moves to another municipality and this municipality does not transmit the data of that person to the previous municipality or does not transmit them correctly (e.g. with a clerical error of the name). In these cases a person can be registered both in the new municipality and in the previous municipality.

21.   As in a mobile society like the German one, the number of (permanent and temporary) double entries can be considerably large, the CPR has to be checked for double entries and those detected have to be cleared up.

22.   For this check on double entries, the register authorities have to provide their data stocks for two reference dates: for the reference date of the census and for a reference date three or four months later. The check itself will be conducted in a two stage process. In the first stage, the CPR built from the data stocks for the reference date of the census will be checked for double entries by a nation-wide comparison of all the about 82 million data records (inhabitants at their main place of residence). In the second stage, the double entries found in the first stage will be matched with the data stock for the second reference date by an address-related comparison to identify the temporary double entries. Usually, they can be cleared up without further queries with the citizens concerned; for the decision whether a temporary double entry has to be counted at the current or at the previous main place of residence, the date of moving to the current place of residence, which is stored in the CPR, can be used.

b) Generation of (private) households

23.   As already mentioned in chapter 2.1, the CPR and the housing census provide information which can be used to group individual persons to form private households. The statistical offices have developed an algorithm with which one can artificially create an image of the existing households by using this information.

24.   The underlying household definition is the one referred to as the household-dwelling concept; that means, the aggregate number of persons occupying a housing unit (dwelling) form a household. The alternative household concept - the so called housekeeping unit concept - cannot be applied, although it would be preferable, as there is no information on economic relationships

between members neither in the PRs nor from the housing census. However, in Germany, the use of one concept or the other does not lead to greatly differing household struc tures as there is only a small number of multi-occupied dwellings.

25.   With the newly developed household-generation algorithm, households are created in a multi-stage process. There are -roughly speaking- 4 stages:

♦   1st stage: persons are combined to form family nuclei of the types "married couple with or without children (who may be the couple's common children or one spouse's children)" and "single parent with children" by using variables stored in the CPR: name of spouse, names of children, name of father/mother.

♦   2nd stage: family< nuclei created on the first stage and single persons are linked to their dwellings by matching the names of the occupants of a dwelling as re-ported by the house owner with the names of the persons as stored in the CPR.

♦   3rd stage: additional information on persons either stored in the CPR (family name/name at birth, date of moving into the current dwelling, address of the previous place of residence, age, sex and marital status) or reported in the housing census (name of second occupant) are used to group persons together to households. This stage consists of 10 sub-stages. The most important are:

a)   Combining persons to cohabiting couples (with or without children)
♦   when the names of two (not married) occupants are reported for a spe cific dwelling in the housing census and their age difference does not exceed a fixed threshold
♦   when a cohabiting couple has a common child (in these cases information on the relationship between this child and his/her mother and his/her father can be used to link them)
♦   when the dates of their moving into the current dwelling and the addresses of their previous dwelling coincide.

b)   Linking grown-up children to their parents when family names coincide and the age difference exceeds a fixed threshold and the dates of their moving into the current dwelling coincide.

c)   Linking elder single persons and elder couples to their descendants (children or grandchildren) when their family names or names at birth coincide and the age difference exceeds a fixed threshold

♦   4th stage: On this stage, the household generation is finished for all buildings for which there are still persons or families that could not be linked to any dwelling up to this stage.

26.   The distribution of these persons/families to the dwellings is carried out by using statistical criterions (e.g.: the largest family is linked to the largest dwelling or single persons are linked to households by using the information on the number of residents living in a specific dwelling).

3.   RESULTS OF TEST SURVEYS

27.   Political decision-makers in Germany have generally opted for a change-over to a register-based census, but decided that the newly developed model of a register-based census should be tested intensively before applying it in a real census. There were 3 main targets of the comprehensive test surveys, which were carried out on the reference date 5th of December:

- ♦ making an assessment of the quality (overcounts and undercounts) of the PRs (register test)
- ♦ testing the efficiency of procedures to identify and to clear up double (or multiple) entries in the CPR (double-entry check).
- ♦ testing the various kinds of data linkage and the algorithm for generating households (procedural test).

28.   For these test surveys, which were conducted on a sample basis, the municipalities and the German Federal Employment Agency had to provide the register records for all persons which had their place of residence (main or secondary place of residence) at the selected addresses, while the housing variables were collected by a postal survey from the owners of the buildings. Simultaneously a traditional household survey with enumerators was carried out at the sampled addresses, and the results of this household survey were compared with the results of the evaluation of the register data and with the results of the algorithm for generating households.

3.1   Results of the register test

29.   The objectives of the register test were to get reliable estimates of the rates of entries not cancelled in the PRs (overcount rates) and the rates of entries missing in the PRs (undercount rates) for the entire Federal Republic, for all sixteen Länder and for four municipality size classes (at the federal level).

30.   By means of a two-stage sample, in about 550 municipalities, some 38 000 buildings (with some 250 000 dwellings and about 550 000 resident persons) were selected by a selection proportional to size. This selection method made it possible to focus the sample on the larger buildings in the bigger cities. It was supposed that the problems of the new census model – with regard to the quality of the PRs and to the difficulties in generating households – are concentrated in these places.

31.   This supposition has been confirmed by the results of the register test. Great differences in PR quality were found between large cities and small municipalities. The overcount rate for municipalities with 800 000 and more inhabitants (7.6%) was almost 3 times as high as that for municipalities with less than 10 000 inhabitants (2.8%). Similar differences were found for the undercount rates, but at a significantly lower level (3.0% for the municipality size class of under 10 000). The average overcount rate of the unadjusted PRs was 4.1% and the average undercount rate 1.7% (see table 3.1).

Table 3.1
Overcounts and undercounts of the population registers for persons at their main place of residence by municipality size classes.

Municipality size class (inhabitants)	Persons registered*)	Including			
		undercounts		overcounts	
	1 000	1 000	%	1 000	%
under 10 000	22 947.5	303.6	1.3	634.6	2.8
10 000 – 49 999	26 112.7	348.4	1.3	900.0	3.5
50 000 – 799 999	23 944.5	509.3	2.1	1 175.7	4.9
800 000 and over	6 980.5	207.1	3.0	527.2	7.6
Total	**79 984.9**	**1 368.4**	**1.7**	**3 237.5**	**4.1**

_____

*) Excluding persons living in institutional households.

3.2   Results of the double -entry check

32.   As described before, the CPR can contain a considerable number of persons who are registered in two or even more municipalities with their sole or main place of residence. Although some of these double entries are discovered by the registration authorities themselves because letters mailed by municipal authorities (e.g. electoral notifications) are returned undeliverable, it has to be assumed that there is still a large number of undetected double entries in the CPR. One of the major goals of the test surveys was to get reliable estimates both on the number and structure of the (undetected) double entries and on the efforts needed to clear up these cases.

33.   Another objective of the double -entry check was to compare the results of different techniques applied to identify double entries. To detect such duplicates in the CPR, so-called object identification techniques have to be applied, i.e. for each person registered at a specific address it has to be checked by a nation-wide comparison whether that person is registered at another address. This identifica tion process can be performed by using the variables name, sex and date of birth/place of birth. Unfortunately, register data can be erroneous (e.g. misspell ings of names or incorrect dates of birth) or incomplete (e.g. missing date of birth or place of birth). Consequently, the identification of double entries cannot be guaranteed to be fault -free: there will be persons whose double entries are not detected and there are false double entries.

34.   For these purposes a birthday sample was taken from the PRs. All municipalities had to provide register records for all persons born on either of three specific days of a yea r (1st of January, 15th of May and 1st of September) and for all persons with an incomplete date of birth; that was a total of about 970 000 records (1.2% of the overall population).

35.   Those data records were checked for double entries by applying a rather simple identification technique: For each record the variables birth name, surname(s), date of birth and place of birth were first standardised in technical terms and after that combined to form a string; a double entry was found, where two data records showed identical strings. Table 3.2 presents the expanded results of this double-entry check technique by municipality size classes.

36.   Table 3.2 shows that the overcounts caused by double entries constitute only about one fifth of total overcounts (e xcluding "temporary" overcounts) and the rates of overcounts caused by double entries do not vary by municipality size classes. This means that the adjustment of the

CPR for double entries will reduce the total overcount rate (2.3%) by only about 0.6% (for Germany and for three of four municipality size classes).

Table 3.2
Overcounts in the CPR after checking for (permanent) double entries by municipality size classes (persons at their main place of residence).

Municipality size class (inhabitants)	Persons*) registered	Including					
		Net overcount**)		cleared up by double-entry check		remaining overcounts	
	1 000		%	1 000	%	1 000	%
under 10 000	23 071.0	459.5	2.0	149.9	0.7	309.6	1.4
10 000 - 49 999	26 928.1	643.4	2.5	153.3	0.6	490.1	1.9
50 000 - 799 999	24 839.1	801.6	3.4	139.3	0.6	662.3	2.8
800 000 and over	7 342.0	416.3	6.0	43.0	0.6	373.3	5.4
Total	**82 180.0**	**2 320.8**	**2.9**	**485.5**	**0.6**	**1 835.3**	**2.3**

_____
*) Persons living in institutional households are included. **) Excluding "temporary" overcounts.

37.   In the second part of the double-entry check, commercial software for identifying double entries was applied to the data records of the birthday sample. Although those computer programmes apply more sophisticated identification techniques (e.g. phonetically comparison, fuzzy-logic procedure), they have not proved superior to the self-developed programme (which works with a simple one-to-one string comparison). Indeed most of the commercial software products detected more double entries, but the share of false double entries was considerable (their shares ranged from 7.5 to 33%). That would not be acceptable in a complete census. It was therefore decided to give up the original plan to apply commercial software products for tracing double entries at the next census and, instead, to further develop our own computer programme (e.g. by including the variable place of birth into the string comparison).

3.3   Results of the procedural tests

38.   From the sample described in chapter 3.1, a sub-sample of some 16 000 buildings in about 230 municipalities was drawn. This sub-sample served mainly two objectives:
- Checking to what extent micro data of the FEA registers can be linked with the CPR records.
- Checking and improving the algorithm for generating households.

3.3.1   Linking microdata of FEA registers to the CPR data

39.   As shown in table 3.3.1 only 89.6% of employees registered in the registers of the FEA could be linked to persons registered in the CPR, although in both registers all variables necessary for an exact one-to-one second linkage (name, sex, date of birth) were contained. The main reason for these results – which at first seem disappointing - was that a relatively large percentage of the addresses stored in the FEA registers were out of date (the employees have to report address changes to the FEA only once a year, namely with the annual report at the end of

the year). Further investigations by combining data of the sample for the procedural tests with the data of the birthday sample showed that about 90 % of the employees which could not be linked to the CPR on an address-related comparison can exactly be linked by matching the two data sources in a nation-wide comparison. Therefore it can be expected that in a complete census about 98 % of records stored in the FEA registers can be linked exactly to the CPR records in an automated process.

3.3.2   Evaluation of the algorithm for generating households

40.   This part of the procedural test served the following two targets:

- ◆ It should give an answer to the question of whether the results of the developed algorithm for generating households correspond to reality or to what extent they differ from reality
- ◆ It should be tested in how far inconsistencies detected by comparing the data reported by the house owners (number of occupied dwellings, number and names of residents living in these dwellings) with the CPR data (the number of persons registered in the CPR at the same address) can be used to cancel overcounts in the CPR. To clear up such inconsistencies, the concerned persons have to be queried either by a postal survey or by a direct interview.

41.   Compared to the results of the household survey, the number of household is slightly underestimated by the household generation (s. part of table 3.3.2.1: households with incomplete records in the household survey questionnaire were excluded). This is mainly caused by the undercounts of the CPR (dwellings which were reported as occupied in the housing census could not be filled with persons from the CPR); there were only few cases where persons were falsely combined to form a household, although these persons actually resided in different dwellings. Table 3.3.2.1 (part a) shows also that large households (especially households with 5 and more persons) are significantly overestimated. This is a consequence of the relatively high overcount rates in the CPR.

42.   These test results indicate that the accuracy of the results of the household generation algorithm depends strongly on the quality of the CPR data. To prove this supposition a simulation was made: persons which were stated as an overcount or an undercount were excluded from the household generation process. The results of this simulation are presented in part b of table 3.3.2.1: the structure of the households by household size produced by the household generation algorithm would be very similar to the corresponding results of a traditional household survey if we succeed in reducing the overcounts and the undercounts in the CPR.

Table 3.3.2.1
Number of households by household size

Household size	Household generation	Household survey	Differences	
			1 000	%
a) Overcounts and undercounts in the CPR are not excluded				
Households total together including with...	**29 612.4**	**30 068.8**	**-456.4**	**-1.5**
1 person	9 006.5	9 517.1	-510.6	-5.4
2 persons	10 138.3	10 376.8	-238.5	-2.3
3 persons	4 760.3	4 690.3	70.0	1.5
4 persons	3 945.0	3 901.4	43.6	1.1
5 or more persons	1 762.3	1 583.2	179.1	11.3
b) Overcounts and undercounts in the CPR are excluded				
Households total together	**29 484.8**	**29 641.6**	**-156.8**	**-0.5**
including with...				
1 person	9 393.8	9 440.4	-46.6	-0.5
2 persons	10 220.5	10 176.0	44.4	0.4
3 persons	4 527.0	4 629.5	-102.5	-2.2
4 persons	3 779.4	3 843.6	-64.2	-1.7
5 persons	1 564.1	1 552.0	12.1	0.8

43.   As these results of the simulation show that the structure of the households will be biased with regard to household size (too few households with 3 or 4 persons and too many households with 5 or more persons), even if the overcounts and undercounts in the CPR can largely be eliminated, the household generation has to be further developed (particularly stage 4 has to be improved).

45.   Simulations were conducted to examine if the overcount rates in the CPR can significantly be reduced by clearing up inconsistencies between housing census data and CPR data. They showed that this adjustment procedure has an acceptable efficiency only for the sub-population living in one or two-family houses. In these parts about 55 % of all overcounts can be detected by interviewing only 7% of all households living in these type of buildings. For the sub-population living in multi-family houses, this relation proved to be inadequate (to reduce the over count rate for this sub-population by 50 %, about 20 % of all households living in this type of building have to be interviewed). It was therefore decided to apply this adjustment procedure in the next census only for one and two-family houses. By applying this adjustment procedure, the overcount rate could considerably be reduced for smaller municipalities, as there a big part of the population lives in one or two-family houses.

46.   Table 3.3.2.2 presents the correction effects of the double-entry check and the clearing up of inconsistent cases in the household generation process on the overcount rates by municipality size classes. For small towns with less than 10 000 inhabitants, the overcount rate can stepwise be reduced to:

- ♦ 1.3 % by subtracting double or (multiple) registered persons,

♦ 0.7 % by additionally clearing up cases with inconsistent data for household generation.

47.   As for small towns an average undercount rate of 1.3 % has been established, the application of these two adjustment procedures will result in a slight underestimation of the population in small towns by about 0.6 % on average.

Table 3.3.2.2
Correction effects of double-entry check and household generation on the overcount rates of the CPR by municipality size classes.

Municipality size class (inhabitants)	Overcount rate (excluding temporary overcounts)	Correction effect			Resulting overcount rate	Under-count rate
		double-entry check	household generation	combined effect (estimated)		
	%	percentage points			%	
less than 10 000	2.0	0.7	0.9	1.3	0.7	1.3
10 000 – 49 999	2.5	0.6	0.7	1.1	1.4	1.3
50 000 – 99 999	2.4	0.7	0.3	0.9	1.5	2.1
100 000 and more	4.3	0.6	0.4	0.9	3.4	2.4
Germany	2.9	0.6	0.6	1.1	1.8	1.7

3.4   Evaluation of the results of the test surveys

3.4.1   Linking microdata without identification number

48.   The tests validated that an exact one-to-one record linkage of microdata from different data sources without a uniform identification number is feasible and sufficiently accurate if the variables required (family name/birth name, surnames, sex and date of birth) are available. It can be assumed that about 98 % of persons in employment registered in the files of the FEA can be linked correctly to their records in the CPR in an automated process.

3.4.2   Quality of PR data

49.   In Germany, a major task of a population census is to determine the official number of inhabitants for each municipality and each "Land". The official inhabitant numbers serve as a measure for the distribution of tax revenues between the Federation and the Länder, among the Länder themselves (the financial resources of Länder with large revenues are reallocated to those with smaller revenues) as well as between the Länder and the municipalities. For this purpose, census results of high accuracy are needed.

50.   The tests showed that the overcounts in the CPR can be significantly reduced by applying the tested adjustment procedures (comparing data stocks for two reference dates in order to discover and to clear up "temporary" overcounts, double-entry check, interviewing households living in one or two-family houses, for which inconsistencies between data from the housing

census and CPR data were stated in the household generation process). However, the remaining errors in the CPR, that is in particular the remaining relatively big variance of overcount rates be-tween the municipalities, cannot be accepted as a census result.

51.   The Statistical Offices of the Federation and the Länder have therefore suggested to extend the tested model of a register-based census with a sample survey.
A random sample of addresses from each municipal PR will allow limiting the size of register errors, at least by probability. The size of overcounts, undercounts and the balance between overcounts and undercounts, and therefore the population of each municipality enumerated from the register, can be estimated by a suitable random sample.

52.   It was assumed that the number of overcounts, undercounts and their balance should be estimated with an accuracy of at least $\pm$ 1 %, given a probability of 95 % (confidence interval) on the average of the municipalities of the respective municipality size class. This should be the guide-line to calculate the necessary sample size.

53.   The supplementary sample survey can also be used to collect further information on variables usually not available from administrative registers, such as the economic variables for self-employed persons or the educational variables for economically inactive persons.

54.   Providing sufficiently accurate information on the additional variables requires another criterion for the necessary sample size. Assuming categorial variables with a quota of 0.5 to be estimated (worst case) and given an estimated cluster effect of 2.0 (for the proposed sample design), it can easily be calculated that 5 000 persons will be needed to attain an (absolute) standard error of not more than 1%, which should be the limit for the publication of regional results.

55.   For an average town of at least 50 000 inhabitants and given the sample design described above, 550 addresses with a mean population of 9.4 persons should be surveyed to fulfill the criterion of an (absolute) standard error of less than, or equal to 1%.

56.   Due to the extension of the sample size - motivated by the second goal, that is to get precise information on additional variables -, the original goal to achieve a sufficiently accurate estimation of the number of inhabitants of a municipality will be achieved with a higher accuracy than originally demanded, especially for smaller towns.

57.   A nation-wide sample size of about 20.4 million persons would be necessary in order to meet the required accuracy criterions. Small municipalities would however experience an extremely high sampling fraction, up to total enumeration. On the other hand the procedure of generating households as a tool for discovering register errors has shown to be more effective for one and two-family houses, which is the usual kind of settlement in small towns; with the

application of this tool, the overcount rate for municipalities with less than 10 000 inhabitants can be reduced to 0.7 %. For these reasons, the Statistical Offices of the Federation and the Länder recommend to the political decision- makers to apply the supple mentary sample survey only for municipalities with 10 000 and more inhabitants. This would reduce the total sample size to 5.9 million persons.

58.   An important aim of the next census is to create a single, person-level database which will be used to generate all statistical outputs from the census. This will be accomplished by a correction procedure, where information on the numbers of overcounts and undercounts obtained by the supplementary sample survey is used to cancel or to add an according number of individual persons from/to the CPR. It is planned to use probability information on being an overcount/under count obtained in the process of generating households for this correction proce- dure. An appropriate procedure has still to be developed.

4.   RECOMMENDATIONS FOR THE NEXT CENSUS

59.   On the basis of the above described results of the test surveys, the Statistical Offices of the Federation and the Länder have recommended to the political decision-makers a census model for the next census which should consist of the following components:

- ◆ Demanding the population register data from all municipalities (about 13 500) for two reference dates; each of the two data stocks will contain about 88 million individual data records (about 6 million people are registered with a secondary place of residence);
- ◆ Demanding register data from the Federal Employment Agency (for about 34.5 million employees subject to compulsory social insurance contributions) and from the public administration (for about 2 million civil servants, army personnel and judges);
- ◆ Conducting a postal housing census by interviewing the about 17 million owners of the buildings.
- ◆ Collecting census data for persons living in institutional households (by interviewing the managers) and for students living in student hostels (by direct interviews).
- ◆ Checking the central population register for double (multiple) entries; it can be expected that about 90 % of the discovered double entries can be cleared up without further inquiries with the persons concerned.
- ◆ Combining individual persons to form private households by matching the records of the population registers with the records of the housing census; by this means, about 38.5 million households will be created; this generation of households will also be used as a tool to discover erroneous entries in the population registers; for this purpose about 1.7 million persons (living in one or two-family houses) residing in municipalities with less than 10 000 inhabitants will have to be interviewed (by phone, mail or direct interview).
- ◆ Conducting a supplementary sample survey in (about 1 500) municipalities with 10 000 and more inhabitants (overall sample size: 5.9 million persons).

5.   CONCLUDING REMARKS

60.   The new census method will be a much cheaper alternative to the costly census projects of the past. The costs of the recommended model of a register-based census are estimated at about EUR 336 million, that is only one third of the expected costs of a traditional census (about EUR 1 020 million). Although a combined method of data collection (the use of registers and surveys) will have to be applied in order to ensure the complete coverage and a sufficient accuracy of the census results, the new census method will involve a much smaller response burden on the citizens (in total about 27 million respondents) than a complete enumeration of the population (about 82 million respondents).

61.   The main disadvantage of the new approach is that it cannot guarantee full census information and reliable census results at any detailed regional level:

- ♦ There will be an overall surplus of undercounts to overcounts (by 0.6% on average) in municipalities with less than 10 000 inhabitants, which means a systematic underestimation of their population.
- ♦ There will be no information on variables which cannot be obtained from registers or from the housing census (e.g. the economic variables for self-employed persons or the educational variables for persons who are not economically active) for small municipalities (less than 10 000 inhabitants).
- ♦ The census results for local units below the municipality level may be biased, as the errors in the population registers are corrected by the results of the supplementary sample survey, which can guarantee unbiased estimates only for the municipality level.

62.   An interesting option for the future is therefore to use small area estimation techniques to estimate the values of the additional variables for small towns and for local units within large cities.

63.   At the end of 2003, the report on the main results of the test surveys, which included the recommendations of the Statistical Offices on the new census method, was submitted to the competent Ministries of Interior (Federal Ministry and Ministries of the "Länder"). It is planned that they will make their decision on the method and the date of reference of the next census at their next regular meeting in November 2004. We hope that they will decide in favour of the proposed census model and they will decide that Germany will join the next census round in the European Union, which is scheduled for 2011.

64.   There is a general agreement between producers and users of census data that a new census is urgently needed, particularly in order to calibrate the "integrated system of population statistics", which provides up-to-date monthly, quarterly and annual data on the population. The population census is the core of this statistical system. Based on the municipality results of the census, the population figures are continuously updated (in a breakdown by sex, age, marital status and citizenship of persons concerned) by registration of births and deaths, external and

internal migrations, marriages and divorces and by registration of naturalisations; these updated population figures serve also as the estimation frame of a 1% sample survey ("microcensus"), with which a large scale of demographic and socio-economic variables are collected annually. In the course of time, however, incorrect registrations may lead to deviations from the real situation in the updated results. For this reason, regular stocktaking is required like in commercial accounting. In population statistics the function of stocktaking is fulfilled by the population census. As the last census was conducted in 1987 a new stockstaking should be carried out in 2011 at the latest.

References

Bierau, D. (2001). *Neue Methode der Volkszählung. Der Test eines registergestützten Zensus.* Statistisches Bundesamt, Wirtschaft und Statistik, vol. 5/2001, pp. 333-341.

Bihler, W. (2002). *Das Stichprobenverfahren der Testerhebungen für einen registergestützten Zensus.* Statistisches Bundesamt, Methodenberichte, vol. 2/2002.

Michel, B. (2004). *Ergebnisse der Mehrfachfallprüfung des Zensustests.* Statistisches Bundesamt, Wirtschaft und Statistik, vol. 9/2004, pp. 979-991.

Rimmelspacher, B. and von Roncador, T. (2004). *Haushaltegenerierung im Zensustest.* Bayerisches Landesamt für Statistik und Datenverarbeitung, Bayern in Zahlen, vol. 09/2004, pp. 340-352.

Statistische Ämter des Bundes und der Länder (2004). *Ergebnisse des Zensustests.* Statistisches Bundesamt, Wirtschaft und Statistik, vol. 8/2004, pp. 813-833.

*****