

**Federal Statistical Office, Germany**

**June 2000**

**Dieter Bierau**

## **Testing a register-based Census**

Against the background of the experience acquired with the last census in the 1980s, "milder means of data collection" were looked for in the 1990s, but the estimated costs for the developed 12 alternative census models ranged from 1.5 billion DM to about 5 billion DM.

The expected costs and the fear of political decision-makers that a complete census covering all citizens might prompt boycott movements like that in the 1980s lead to the decision not to conduct a traditional census with enumerators again, but only to evaluate existing data sources for a new census. This decision was expanded later to include a survey of buildings and dwellings which should be conducted by mail, contacting the owners of the buildings.

In this framework the Statistical Offices of the Federation and the Länder developed a new census model which should in the end result in census-like data files for any person which should be suitable for evaluation in any combination of variables. The proposals included rather complicated methods of checking and linking the data, which have hardly undergone any practical tests yet. Data linkage has to be realised without personal identification numbers, as such ID numbers would not be permitted in Germany for legal reasons.

In March last year, it was therefore decided that the proposed new methods and the suitability of registers for statistical evaluation purposes be tested intensively before actually conducting a census. At present, nine working groups of statisticians are preparing such tests which are scheduled for December 2001.

## **New methods for the Census**

The most important source for the new census model are the community based population registers.

Population registers in Germany emerged as early as in the last century as part of tasks performed by the police. In this century, the functions of such registers have been defined more generally and their contents have been largely standardised. Contrary to the former GDR, there is no central population register in the Federal Republic of Germany; keeping such registers is a task of the communities. Today, each of the about 14,500 communities in the Federal Republic maintains a population register, and every citizen living in the community has to register there. The data stored comprise basically name and first name, sex, date and place of birth, marital status (single, married, etc.), citizenship, place of residence, and address. In addition, the register contains a number of other characteristics, for example on voting rights and former addresses. One of the main functions of the censuses conducted before 1987 was to adjust the communal population registers by means of the statistical results of the census, so that communities and statistical offices were then able to use the same basic figures.

As regards the error size in the population registers and especially the error variance between the communities, estimations range from a deviation of less than 1% to deviations between 3% and 4%. However, the data records to be supplied by the communities from their population registers should not be taken over without checking. It was planned to perform various checks for double entries and other plausibility checks.

Other administrative registers are maintained at the Federal Institute for Employment, which can access the social insurance files for wage earners and salaried employees: One of those registers is a file of persons subject to social insurance contributions. It contains name, date of birth, sex, and address. Although it does not include self-employed persons and part of employees with other types of jobs, it covers about 85% of all persons engaged in economic activity. That register might also be evaluated at a detailed regional level.

For buildings and dwellings, which in the past were covered together with most population censuses, there are no nation-wide registers. That means, such data must be collected as primary statistics (by questionnaire). A very important part of the proposal is therefore a survey of buildings and dwellings which, for cost reasons, should be conducted by mail, contacting the owners of the buildings. Nevertheless, this part of the model is the most expensive one, costing an estimated 250 million DM. The owners or managers of the buildings should provide information on the age, size, and equipment of the buildings and dwellings and (that is new!) the names of the tenants or owner-occupiers.

Although the population registers contain the address, that is they indicate the building in which a person lives, they do not indicate the specific dwelling within the building. On the other hand, the registers contain not only the basic variables such as name, age, sex, and address, but also information on who is the person's spouse, what children belong to the family, and when - and from where - a person moved to the current address. Using information like that, statisticians of individual towns have been trying for years already to artificially create an image of the existing households. Computer programmes have been developed which use the available variables to group persons together to form households. For smaller buildings or in rural areas, this works quite well. For large buildings in cities, such methods regularly overestimate the number of households: Too many small households are formed because for some persons no characteristics have been found which would allocate them to other households. A problem that appears impossible to solve is the automatic allocation of persons living in communes or consensual unions. This means that, for many larger buildings, the number of households generated is larger than the actual number of dwellings existing in those buildings.

It would be an obvious choice here, when interviewing the owners of buildings, to collect not just technical data on the dwellings but also the names of the residents living in the dwellings. As these persons are registered in the population registers, they could provide the link between dwelling data obtained from a primary survey and data from population registers. With the number of tenants or owner-occupiers being known, the number of dwellings is known, too. This means that the maximum number of households possible can be integrated into the household generation programmes.

As mentioned earlier, there are different assumptions with regard to the quality of the population registers. They are used for many administrative purposes, so we assume that errors would be detected rather quickly. Although many communities make great efforts to keep their registers up to date, not all of them are equally committed. Therefore, the statistical offices have been looking for opportunities to check and improve the quality of data from the registers. By interviewing the owners of buildings and by using household data generated from the population registers, we get a tool for mutual data checking. Inconsistent data should be verified by directly interviewing the households.

It is also planned that all data received should be linked up on a person-related basis. In addition to linkage in the context of household generation, this refers mainly to data obtained from other administrative files, especially employment data. Such employment data are available in files on wage earners, salaried employees, officials, soldiers, unemployed persons, and persons undergoing vocational training. These files should be linked with data from the population registers through the characteristics: name, date of birth, and address.

Even if all person-related linkages should be feasible in legal and technical terms, employment data would not be available on all persons engaged in economic activity. For them the model includes a supplementary 10-percent sample survey. The basis for that sample should be all persons at working age to whom no employment data have been allocated during the linking activities.

For education issues, which were covered in most previous censuses, no reliable and up-to-date source is available. They are not included in the model.

The costs of that model have been estimated at nearly 400 million DM. When compared with the costs of previous censuses, however, this is a rather small amount.

## **Scope of the Census Tests**

What I have presented now is the overall scope of proposals made for a new census which we presented in August 1998. Political decision-makers in Germany have generally

opted for a methodological change-over to a register-based census, but decided that before realising them, comprehensive testing shall be performed on the basis of the proposals submitted.

The tests will cover not only the technical opportunities regarding the various kinds of data linkage and household generation; they will also extend to the question of whether the results actually correspond to reality or to what extent they differ from reality. Also, the quality of population registers and other data sources will be tested. To compare results with reality, it will be necessary to send interviewers to selected buildings who will interview all persons living there. For the tests, too, a separate law will be required, for example, to create an obligation for the communities to provide population data to the statistical offices or to create other obligations to provide information.

At present, nine working groups of the statistical offices are preparing the law and the various parts of the test. The test will comprise different parts:

## **1. A Birthday-sample of all communal registers**

- **Checking for multiple entries in the population registers:**

As described before, the population registers are not kept centrally but at community level. So, errors may gradually slip in if persons registering or deregistering in different communities are not correctly recorded. This might lead to a situation where persons are registered in several communities. This should be detected after some time because letters mailed by community agencies would be returned undeliverable. But it is assumed that not all communities make sufficient efforts to clear up such cases. Therefore, we intend to obtain from all communities the data of all persons whose birthday is on either of three specific days of the year which have been chosen at random and the data of persons whose data are incomplete. That will be about 1.2 percent of the population. The data records will be combined centrally at the Federal Statistical Office and checked there for double entries. Such double entries would then have to be cleared up at the statistical offices of the Länder.

- **Checking data channels:**

The population registers of the about 14,500 communities are maintained at some 1,500 computer centres. Although there is a uniform standard for data records, some programming work will become necessary in practice to actually standardise the records. As far as we know, there are about forty different computer programs used for the management of the population registers. Besides that, we are not quite sure, if every community keeps its records according to the federal standard or if there are deviations with some of the characteristics. As we get records from every municipality we can check the data transmission channels and data records differing from the standard.

**2. Checking the accuracy of population registers when compared with field work in a sample of buildings:**

By means of a two-stage sample, some 38,000 buildings with some 250,000 dwellings and about 550,000 persons will be selected in about 560 communities of different size. We know that the problems of the new model are concentrated in the bigger cities and especially in the bigger buildings as regards register quality or the difficulties in generating households. Therefore the sample is not representative: it focuses on large communities and on large buildings.

To get informations on the persons living in those buildings, the register records will be demanded from the relevant communities. Interviewers of the statistical offices of the Länder will check by field work whether and how the registers reflect reality.

**3. A sub-sample of buildings for testing different procedures**

In the selected communities mentioned above (2.) we will draw a sub-sample of some 16,000 buildings in about 230 communities. With that sub-sample the following procedures will be tested:

- **Checking the opportunities of obtaining information from the owners of buildings:**

The owners of the buildings will be asked by mail to provide information on the buildings and dwellings and to supply the names of the tenants or owner-occupiers. As part of the field work in the buildings, as mentioned in part 3, the persons living there will be asked to provide data on the dwelling (equipment,

heating, etc.). Then it will be tested to what extent the owners are able to provide correct information on the individual dwellings and whether they are able to indicate all the names of tenants or owner-occupiers required for household generation.

- **Checking and improving household generation:**

For the buildings selected in part 3 the data records from the population registers will be combined per building to form households and, through the tenant's names, will be allocated to specific dwellings. Through the field work in the buildings, it will be checked in how far the generated households reflect reality.

This is the very core of the model and the most complicated part in it – and we do not get any data records now from the population registers to develop the algorithms. The law on the census test will give us two years after the reference day to improve the methods. In these two years we can work with the register records and the results of the survey in the selected buildings to compare them and to improve the algorithms.

It will also be tested, how often there are inconsistencies between the two data sources, which have to be checked by field work.

- **Checking record-linkage:**

From the administrative registers of the Federal Institute for Employment, which were described earlier, personal data will be demanded for the addresses of the selected buildings. By means of address, name, and date of birth, these data records will be linked with the personal data records from the population registers, which will have been grouped to households. As part of interviewing the persons living in the relevant buildings, those persons will also be asked for employment information. These data will then be compared with the linked records mentioned before.

#### **4. Evaluating the results**

Hopefully before 2002 all statisticians involved will discuss the results of the different parts of the tests, describe them in a summary, evaluate them and make recommendations as to the feasibility of the new methods.

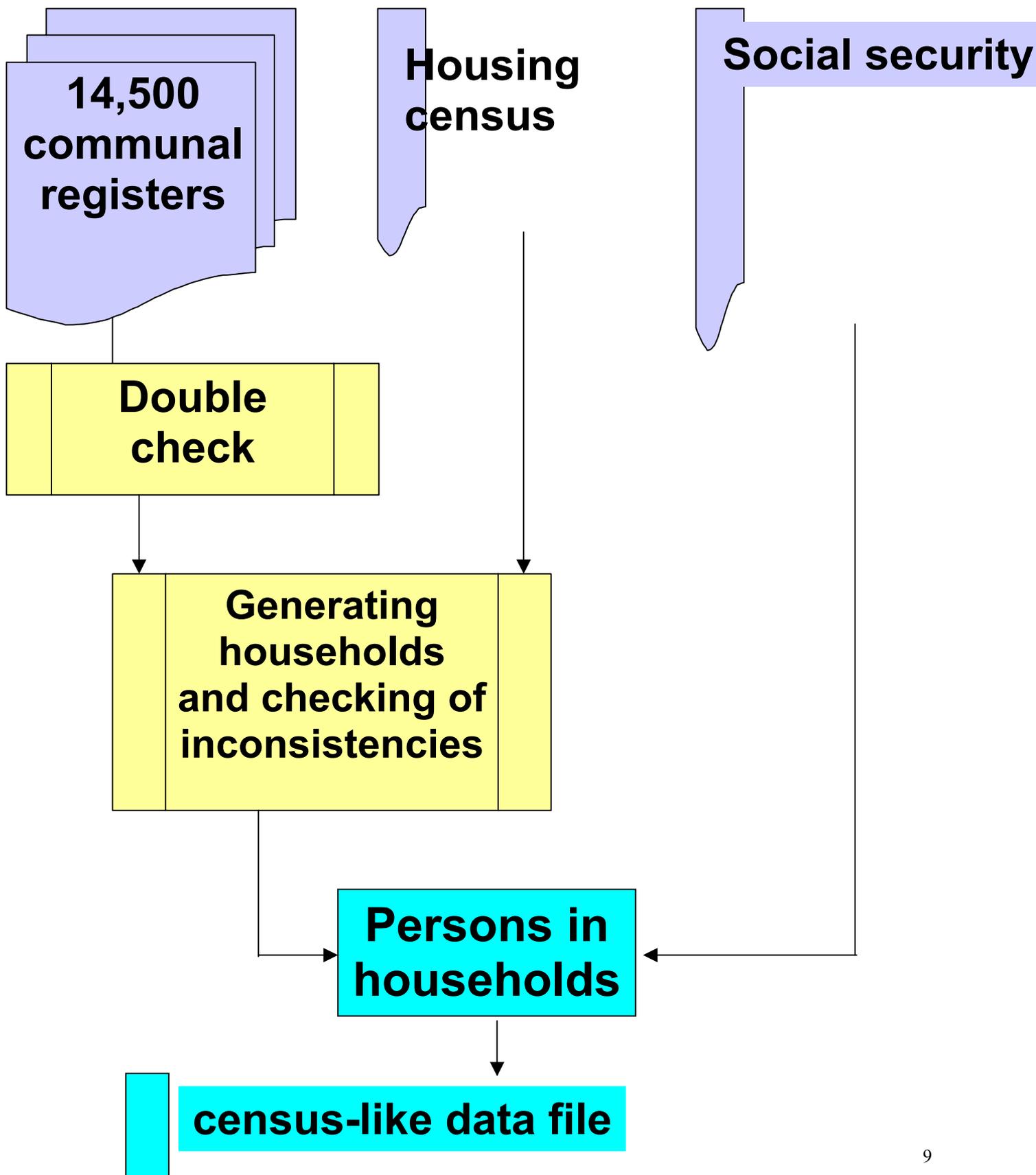
The law on the census test is still in discussion at federal ministries. Presumably in autumn 2000, the law will be introduced in our parliament, the Deutscher Bundestag. It is planned to conduct the test survey in March next year, but if there are any further delays, the reference day will be postponed to September.

Regardless of when the law will have been adopted, the statistical offices are making any efforts for the required organisational and technical preparations. According to our schedule, we hope to finish evaluating the tests by the end of this parliamentary term, that is before summer 2002. However, it is too early to speculate now about what conclusions the next Bundestag will draw from those analyses for actually conducting a new population census.



# Simple model of a register based

## census





# 1. Birthday sample of all communal registers

## All inhabitants born on

- 1st January,
- 15th May,
- 15th September or
- with incomplete birth-data

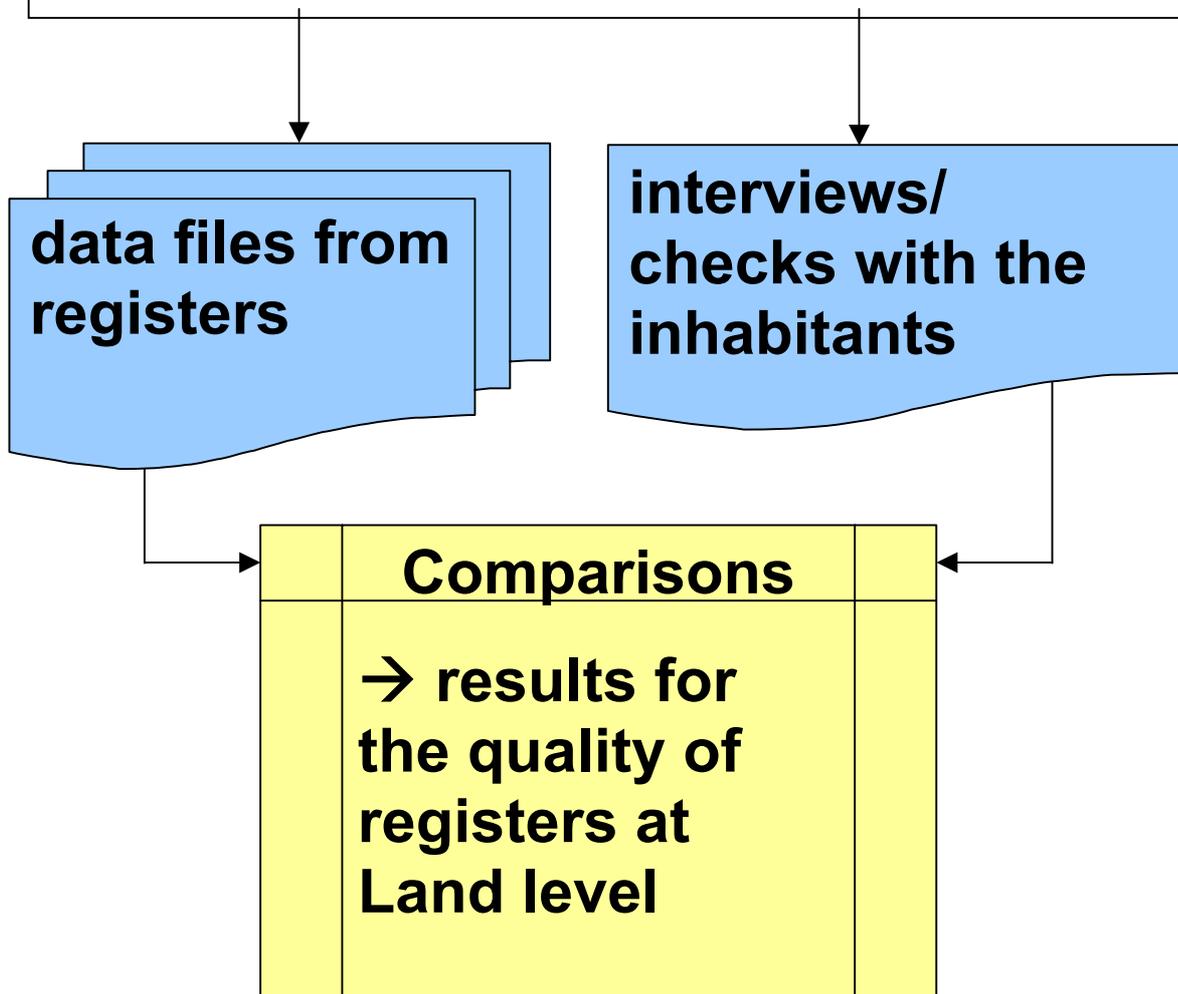
→ about 1.2 % of the population  
(  $\approx$  970,000 records)

- Checking for multiple entries in the population registers
- Testing data channels and register software



## 2. Sample of 38,000 buildings/ addresses

**Two-stage sample of  
38,000 buildings/ addresses  
in 570 communities  
with about 250,000 dwellings  
(about 550,000 persons)**





### 3. Sub-sample of buildings/addresses for testing different procedures

**Two-stage sample of  
16,000 buildings/ addresses  
in 230 communities  
with about 120,000 dwellings  
(about 220,000 persons)**

