

Working Paper No.2
9 November 2004

ENGLISH ONLY

**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– Invited paper

THE NEW FRENCH ROLLING CENSUS

Submitted by INSEE, France*

ABSTRACT

To answer the increasing demand for fresher data and spread the burden of conducting the census over a longer period, the French National Institute of Statistics and Economic Studies has conducted a large program to redesign the census, based on the “rolling census concept” proposed by L. KISH. In this rolling census, every commune under the threshold of 10 000 inhabitants will be surveyed once within a five year period ; larger communes will be divided into five rotation groups of addresses, each rotation group being surveyed in one of the five years. This paper presents the principles and the methodology of this operation and the situation at the end of 2004.

I. REASONS FOR THE REFORM

1. Census has been regularly conducted in France, every five years since 1801 until the 2nd world war. Since the beginning, the census has two main goals : first, determine the official population of all administrative districts, and second provide sociodemographic characteristics of any geographical level of the country, from quarters of cities to France. The 32th general census was conducted in France in 1999 according to the usual framework : everybody enumerated at the same time.

* Paper prepared by Jean-Michel Durr.

2. However, some facts led France to reform the census. First, the periodicity has enlarged since the 2nd world war : 7 years between the 1975 and 1982 censuses ; 8 years between 1982 and 1990, and 9 years between the last two censuses. The last census was postponed from 1997 to 1999, because of budget considerations. Actually, there was no legal basis for the census in France, as in certain countries where the census is mentioned in the constitution, and only a decree was necessary to implement the census. As the international recommendations are to make a decennial census, the trend was clearly to reach this periodicity according to financial considerations. Conversely, there is in France a great trend of decentralization for more than twenty years. More power is given to local authorities, and these new players need more and more fresher local information to conduct their policies. The census, which is the major local information source, had to adapt.

3. Another reason is the parts casting between the national institute of statistics (Insee) and the communes. The commune is in France the smallest level of administrative organization and the mayor is both the representative of the State in the commune and also the representative of the commune as an independent unit. Since the beginning, mayors has been involved in the census as the representatives of the State, in charge of hiring the enumerators, but those were managed by officers of Insee. It was necessary to clear the roles.

4. Finally general census represents a huge burden for Insee and the communes every 8 or 9 years, with many one-shot investments as data processing programs, not re-usable for the next census. In large cities, enumerators faced growing difficulties to get in touch with the population with the consequence of increasing non-response rates. The big number of enumerators makes them hard to manage and to obtain a high level of quality in data collection. Spreading the burden over time was also an objective for Insee.

5. For those reasons, Insee initiated a project for redesign the census at the end of the 90's. The constraint was not to increase the budget of the operation. As France has no population register and French population is very sensitive to privacy, it was decided to explore others ways. Leslie Kish [Kish, L., 1981 et 1990] had proposed in a few papers a new method to conduct censuses, based on the principle of rolling surveys, that is to say doing regular surveys in order to accumulate information. In 1995, Michel Jacod and Jean-Claude Deville laid the foundations of the reform in a report to the director general of the Insee. Michel Isnard was charged to take the study further and to deepen certain items like the sample base. At the beginning of 2001, a special unit named "programme de rénovation du recensement" was created to conduct the redesign of the census with the objective of implementing the new method in 2004.

II. METHODOLOGY

6. The first principle of this new methodology is to collect information over a quinquennial rolling cycle to produce every year significant data for the medium year. Every year Y, data are produced with the data collected in Y-4, Y-3, Y-2, Y-1 and Y and are significant on average for Y-2.

7. The second principle is to have recourse to sample surveys to respect the budget constraint. As a seven year period between two censuses was considered as a good distance, the budget must be every year the same as one 7th of a general census. The consequence is to enumerate 5/7th of the population over 5 years, and one 7th every year.

8. According to the very heterogeneous size of the almost 37 000 french communes, a special sampling strategy was elaborated. For the smallest communes, sampling would be

inefficient because of the minimum sample size. For larger communes, the problem is to build a sample base of dwellings.

9. Two sets of communes are taken in account : under and over the threshold of 10 000 inhabitants. The small and medium-sized communes, with a population under 10 000 inhabitants, house half of the French population (about 30 millions). They are surveyed once within a five year period, and all their dwellings are visited. On the other side, all the communes over 10 000 inhabitants are visited annually, but only a fraction of their dwelling (8%) are surveyed. Every year : half of the population at 20% + half of the population at 8% give 14% of the total population, that is to say about one 7th.

10. First, consider the set of communes under the threshold of 10 000 inhabitants. In each region (regions are parts of the French territory with average population of 1.5 million inhabitants, there are 22 regions in Metropolitan France), five groups of commune are formed using data from the 1999 population census. They will consist of balanced samples [Deville, Tillé, (2000)] of the distribution of different characteristics of the communes. This method is a generalization of the stratification sampling, and consist in selecting reference pattern and building samples close to this pattern. For each region, the five groups represent about the same number of each characteristic, for example the same number of men and women, or persons aged less than 20. Variables taken in account are age (five groups), sex, type of dwelling (individual houses or collective building). Those variables were selected after many simulations, and we also test their stability over time by comparing the results given with the data from the 1990 and 1999 censuses. They assure a good correlation with most of the different variables of the census. The five groups are spread all over each region, thus each group is representative for the region.

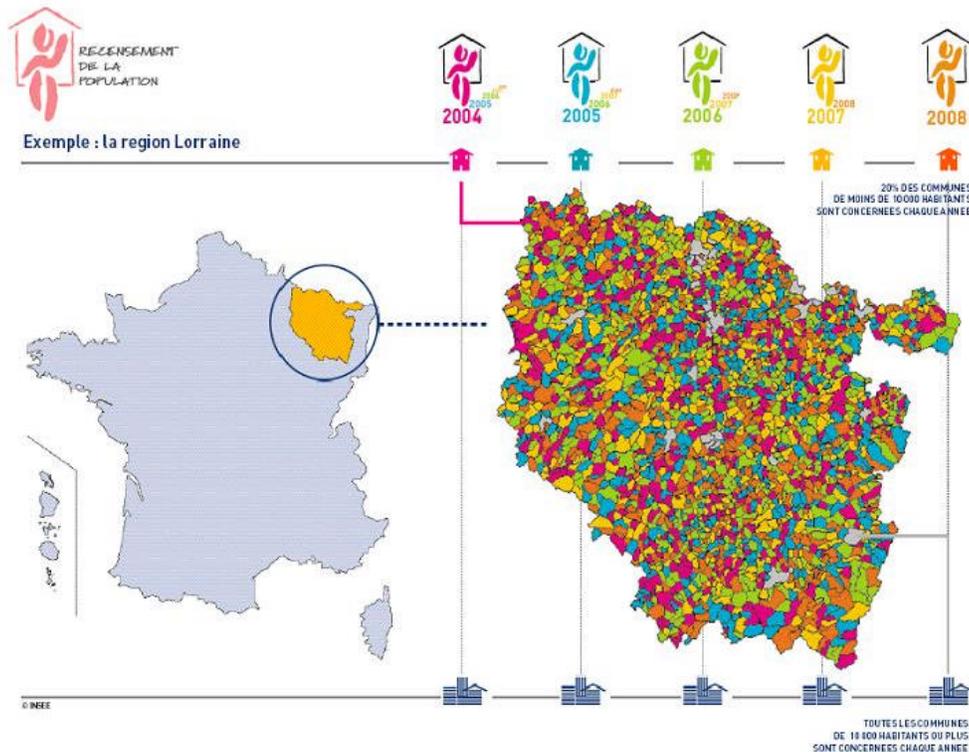


Fig. 1 : the five groups for the region Lorraine

11. The “large communes” sample is based on the “répertoire d’immeubles localisés” (RIL) (inventory of buildings). The RIL is a list of buildings (residential, institutional or commercial) identified individually so as to generate a digitized map. Initially, the RIL will be populated with data from the 1999 census, which provided a statistical portrait of each residential building. The RIL is updated with information from the French post office, the building or demolish permissions, the local tax file. Actually, the elements in the RIL are addresses, because it is more consistent among the different operators as the French post office, the tax department and the communes. Several buildings can be grouped at the same address. Insee and communes exchange information to update the RIL. For example, communes are asked to confirm possible new addresses given by administrative sources.

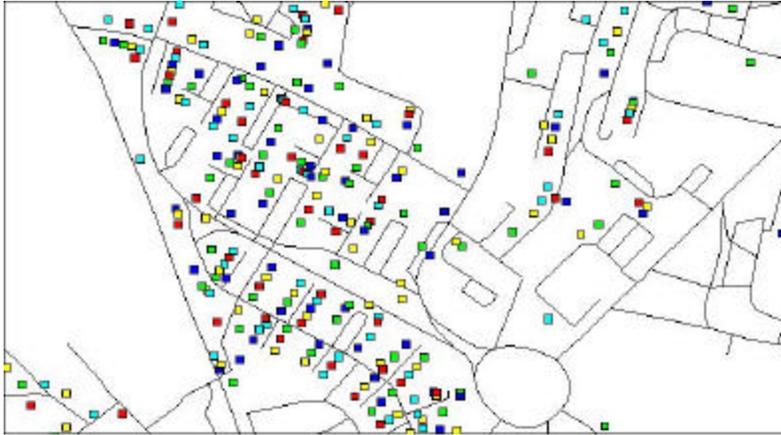


Fig. 2 : the five groups of addresses

12. Five groups of addresses are constituted with the same method used for the groups of communes under 10 000 inhabitants (balanced groups and same variables). Every year, a 40% sample, balanced on age, sex and number of dwellings, is drawn in one group. The sample rate is to be understood in terms of number of dwellings. 40% of the dwellings of the group represent 8% of the total number of dwellings in the commune ($40\% \times 1/5$). Thus, we avoid surveying the same dwelling more than one time over a period of five years. Every dwelling of a sampled address is surveyed, what is easier for the enumerators and more understandable for people.

13. The main problem faced is the heterogeneity of the number of dwellings per address. An address can be either a very large building with more than a hundred dwellings or a sole house. This involves cluster effects and possibly affect estimates. For this reason, a special stratum is constituted with very large addresses. After simulations, it was decided this stratum will group the largest addresses of the commune which represent 10% of the total number of dwellings. This stratum is divided in five groups and is exhaustively enumerated over five years (one group each year). Thus, there is no variance due to sampling for this stratum. This strategy improves accuracy for population estimates in quarters with large buildings, but in return, the sample rate is a little bit lower for other addresses. However, there is an escape clause to ensure the sample rate will not fall lower than 25% for the “non-large” addresses.

14. Every year, the RIL is updated in partnership with the communes. Demolished buildings are removed and new constructions are introduced with their supposed number of dwellings given by the building permission. This number is very important for estimates as an extrapolation variable, but building permissions sometimes give vague information. More, the characteristics (age and sex) of the population living in these new buildings are used for sampling. Therefore, we have to survey those buildings as soon as possible. Therefore, new addresses are exhaustively

surveyed in the next five years, and poured immediately after into one group of the “other addresses” stratum.

15. Every summer, the five groups in the sample base are updated. Each one comprise three strata : large addresses, new addresses and other addresses. Every addresses of the first two strata are exhaustively enumerated, and the third is sampled. The sample rate in this stratum is adjusted to survey 40% of the full group’s dwellings.

III. ESTIMATES

16. From the end of the first five years, in 2008, three sorts of results will be produced every year Y :

- official population of every communes, whatever the year of their last census;
- data for every geographical levels : based on the combination of the data collected over the last five years; they are significant on average for the year $Y-2$;
- data for France and regions, based on the survey of year Y .

Official populations :

17. How to estimate population every year? The solution varies according to the size of the commune.

18. Communes over the threshold of 10 000 inhabitants are surveyed every year at a 8% sample rate. The annual sample is composed of three kinds of dwellings : new addresses dwellings, large addresses dwellings, and other addresses dwellings. Average population over the last five years can be estimated by the mobile mean :

$$Pop_{Y-4..Y} = \sum_{i=Y-4}^Y Pop(LA_i) + Pop(NA_i) + w_i Pop(OA_i)$$

Where LA_i means Large addresses of year i , NA_i New addresses and OA_i Other addresses, and w_i the sample weight the year i for other addresses. Large and New addresses are exhaustively surveyed over five years, so their weight is 1.

Population in $Y-2$ is estimated by :

$$Pop_{Y-2} = Pop_{Y-4..Y} \times \frac{Nbdwellings_{Y-2}}{Nbdwellings_{Y-4..Y}}$$

Where $Nbdwellings_{Y-2}$ is the number of dwellings the year $Y-2$, given by the RIL, and $Nbdwellings_{Y-4..Y}$ the average number of dwellings over the last five years.

19. For communes with a population under 10 000 inhabitants, their last census took place one year comprised between $Y-4$ and Y . According to this date, the population number observed at this moment will be extrapolated or interpolated to $Y-2$. Extrapolation will use the trend of the number of dwellings observed in the local tax file. The rule is following :

Year of census	Population estimates for Y-2
Y-4	Use the trend Y-4 to Y-2 of the number of dwellings observed in the local tax file to extrapolate the population of Y-4
Y-3	Use the trend Y-3 to Y-2 of the number of dwellings observed in the local tax file to extrapolate the population of Y-3
Y-2	Keep the Y-2 census number of population
Y-1	Interpolate Y-2 on the Y-3 estimate and Y-1 census population trend
Y	Interpolate Y-2 on the Y-3 estimate and Y census population trend

Table 1 : population estimates for communes under 10 000 inhabitants

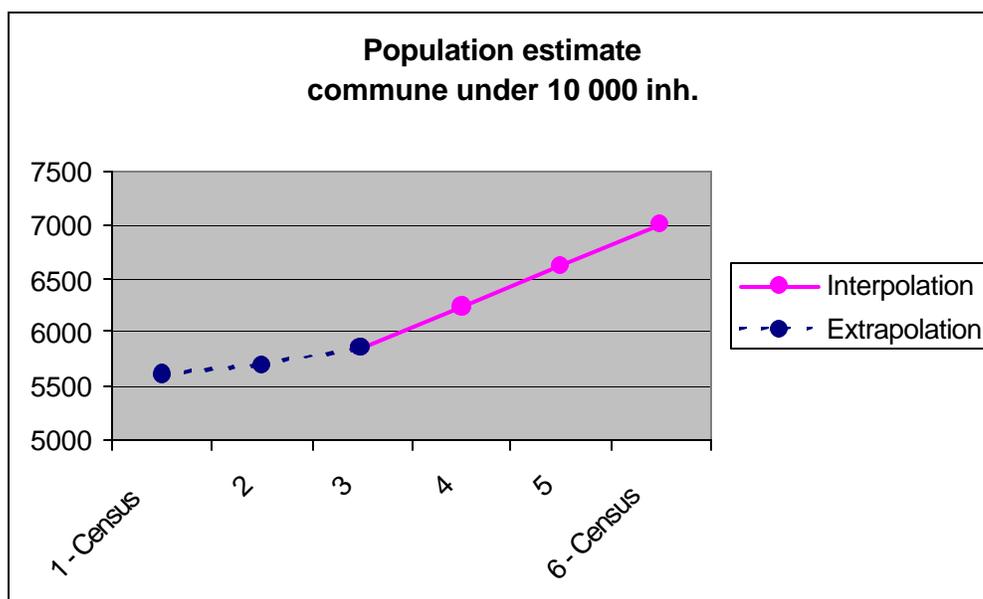


Fig. 3 : population estimates for communes under 10 000 inhabitants

NB : estimates are made two years after (in Y for Y-2)

20. Several comments :

- interpolations are not made between two censuses but between the last estimate and the next census in order to avoid splits;
- if the case of bad quality information from the local tax file, a simple extrapolation of the past trend between the last two censuses is done;
- estimating Y-2 in Y gives a two years maximum gap between date of the census and date of estimate. This method is safer, especially at the beginning. On the opposite, the population number of a commune given by the census will be officialized only two years later.

Detailed results

21. Every year, a file will be constituted with the data collected over the last five years. This file will contain records for every dwellings and persons surveyed during this period. Each record is properly weighted so the file will enable any tabulation at any geographical level.

22. Individual records have the same weight as their dwelling, because the sample is done in terms of dwellings.

Weight for data collected in communes over 10 000 inhabitants is :

$$w_i = \frac{1}{(\text{sample rate})_i} \times \frac{Nb \text{ dwellings}_{Y-2}}{Nb \text{ dwellings}_{Y-4..Y}}$$

where *sample rate* is 1 for large and new addresses and the real sample rate for other addresses.

The coefficient $\frac{Nb \text{ dwellings}_{Y-2}}{Nb \text{ dwellings}_{Y-4..Y}}$ roughly estimates the situation in *Y-2*.

23. Weight for data collected in communes under 10 000 inhabitants is 1 but it is also possible to use the coefficient of “expansion” given by the official population estimate. The accuracy is estimated by simulating many samples. The following tables shows the results according to the size of the estimated figure :

Size	Precision (std/mean)
>50 000	<1%
20 000-50 000	1,50%
10 000-20 000	2,00%
6 000-10 000	2,50%
3 000-6 000	3,00%
2 000-3 000	3,50%
1 000-2 000	4,50%
500-1 000	6%
250-500	8%
<250	>8%

Table 2 : precision of detailed results

24. This is to compare to the accuracy of non updated data many years after the census. For most uses, it will provide a considerable gain.

Overall estimates

25. Every year, more than 4 million dwellings and 8 million people are surveyed, and the sample is representative for France and regions. The national and regional results of the survey conducted at the beginning of the year will be published before the end of the year. In addition, the results for each “small commune” visited during the year collection campaign will be published at the same date.

IV. HOW TO USE THOSE DATA?

26. The general census provided a snapshot of the population, but the picture was not updated for a long time. The rolling census will not provide a movie, but rather a “slide show”. The detailed results represent a rolling average over the last five years and the overall estimates a view of the situation at the beginning of the year.

27. For most demographic criterions, with slow trends, measuring them with a five years rolling average is quite appropriate. The advantage is to get data updated every year. Whatever the date users ask for data, they get recent figures.

28. Some examples at different geographical levels :

- France and regions : information will be updated every year with the figures of the beginning of the year. For example, major migrations flows between regions or with other countries can be observed every year;
- Urban areas : communes under 10 000 inhabitants have been visited exhaustively during the past five years and 40% of the dwellings in communes over this threshold have been surveyed. For example, in the urban area of Montpellier (almost 380 000 inh.) is composed of 3 communes over 10 000 inh. and 29 smaller communes. The map above present the year of census for the different communes. Every year, 45 000 persons, 12% of the population of the urban area, will be surveyed. This rolling sample will well update every year the information available, for example about commuters.

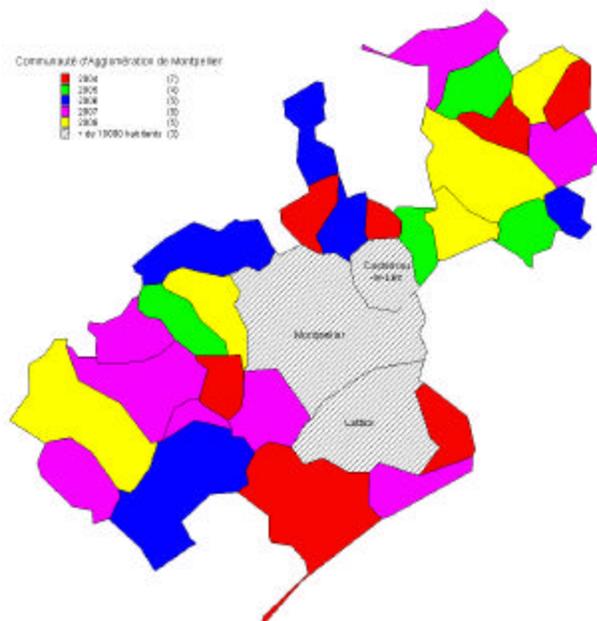


Fig. 4 : census frame in the Montpellier urban area

- Data linked to economic fluctuations require special treatment. Actually, rough comparison of the unemployment rate of communes may be very sensitive to the situation of employment the year they were visited. The good manner is to use the annual survey to calculate the ratio : unemployment rate of the commune divided by the unemployment rate of the region or the country at the same date. Thus, we can properly compare unemployment ratios among communes.
- Urban quarters : large communes are composed of quarter with about 2 000 inhabitants named IRIS. Census data were used for municipal management as schools, sports facilities... Users were afraid about the accuracy of the new data, based on a sample. But they didn't realized that the more the area is small, the more the data are moving. For example, what about 50 unemployed people in a quarter five years after the census ? On the contrary, rolling average is a perhaps a better and

stronger measure. Some examples were built to prove to users the interest of these new data. For example, sampling simulations proved the stability of a typology of IRIS based on characteristics like age or type of household. More, new buildings are exhaustively visited, so information about new inhabitants are very accurate. Thus, municipalities may plan the type of facilities required, as crèche or swimming pool, for example.

29. As every year only the oldest fifth is renewed, data may not be compared to last year but to five years before, and so on every year. It is important to understand how the rolling census is “neutral”. In the past, changes in socio-demographic characteristics were only observed between censuses, and unconsciously seen as linear. Now, the new rolling census will provide a better image of the society shifts.

30. Users will have to learn how to use these new data, because they are collected over a period of time and no more at a single date.

V. HOW THE ANNUAL CENSUS CAMPAIGN IS ORGANIZED?

31. A census campaign is prepared since the previous spring. In May, the updated RIL is sent to the communes for reactions before the end of June. Their remarks are taken in account in July and the sample base is updated. Then, samples are drawn and sent to the communes in September. Communes under 10 000 inhabitants involved in the following census campaign are contacted before summer to get the name of their census manager.

32. During the 4th quarter, forms are printed, with the name of the commune and a barcode (different for every questionnaire), and sent to the communes. Insee deploys almost 600 supervisors to organize and control the operations. Their role is to give advice to the communes and to control the course of the operations, but also to train the enumerators.

33. An advertising campaign (poster advertising and TV commercials) is launched at the beginning of the data collection. Data collection begins the third thursday of January and lasts four weeks in communes under 10 000 inhabitants and five weeks in the others. Before data collection, the enumerators have to recognize their census tract and note the rough number of dwellings it contains. Then, from the reference day, they go to each dwelling, leave the form to the household and take appointment to get them back.

34. In the communes under 10 000 inhabitants, homeless persons and people living in caravans are surveyed the first two days. In the other communes, the survey for this category of people will be conducted every five years, and the data kept unchanged during the interval. Every five years, bargemen will be surveyed by the Insee.

35. People living in institutions as military quartering, hospitals... are surveyed by Insee. The institutions in communes under 10 000 inhabitants are surveyed the same year as the households of the commune are. Institutions in communes over 10 000 inhabitants are surveyed at the rate of one fifth every year.

36. At the end of February, collected forms are gathered by the municipalities and sent to the regional bureaus of Insee. They are controlled, and if necessary, control surveys are conducted, for example to check non response or population number. Then, forms are optically captured by a service provider. Automatic codification and clerical operations follow to code variables as

profession, place of work and economic activity. The data are weighted and finally checked before dissemination.

VI. LEGAL ASPECTS

37. A law had to be voted to allow this reform. The “démocratie de proximité” law was voted in February 2002 by the parliament to lay the basis of the new system:

- Respective roles of Insee and communes;
- Five year cycle;
- Sample for communes over 10 000 inhabitants;
- Official populations : every year for each commune, fixed by using the census surveys and administrative sources.

38. A few texts (decrees) complete the legal structure to specify some points as the duration of data collection, the amount of the state subsidy to the communes or confidentiality requirements...

VII. REQUIREMENTS FOR IMPLEMENTATION

39. Many consultations and meetings have been held to launch the reform. As the two principles of the traditional census, exhaustiveness and simultaneity, were abandoned, municipalities and users were dubious about the project. First, they feared that the sample in large communes will damage the accuracy of the data, especially for city quarters. Second, they were concerned by the combination of five years data.

40. Insee insisted to keep the word “census” to preserve an important response rate of the population and because of the legal implications. The parliament mostly considered that this technical reform was liable to provide a considerable improvement in socio demographic knowledge. The situation was quite different before the vote of the law and after. Of course, the project still faced opposition, but it was now lawful. The major complaint from communes was about the update of the RIL, considered as an imposed burden with no compensation, and about the amount of the census subsidy.

41. In 2003 and 2004, a workgroup with representatives of municipalities and local governments, researchers, statisticians dealt with the question of dissemination and use of the new census data. This group built examples of use from the uses of the former census (some are presented above). Many discussions between statisticians and users occurred and help to better understand each others. For example, some users recognized that they used census data at very small geographical level, even several years after their collection. Statistician tried to explain the instability in time of very small populations.

VIII. The situation at the end of 2004

42. The first campaign took place at the beginning of 2004. We are now processing data collected. Despite it will take five years to reach cruising speed rate for dissemination, we decided to release statistical results as soon as possible. Actually, it will be the best way to prove the advantages of this new census method.

43. Population estimates for communes over 50 000 inhabitants will be disseminated at the end of 2004. First results at national level will be presented in January 2005, just before the

second edition. Before summer 2005, keys figures will be published on the web site for regions and every commune over 50 000 inhabitants and for commune under 10 000 inh. surveyed in 2004. At the end of 2005, the threshold will be lowered to 20 000 inh. because of the combination of two annual surveys, and to 10 000 following year. Results for large communes will also be more detailed year after year.

44. With the preparation of the 2005 campaign, it appears that on the one hand there is no slack period, one campaign being prepared when processing the previous one, but on the other hand, it is possible to capitalize on problems faced one year and to improve. For example, forms have been printed far quicker than last year.

BIBLIOGRAPHY

1. Deville, J.C., Tillé, Y. (2000), «Echantillonnage équilibré par la méthode du cube et estimation de variance » , Journées de méthodologie, décembre 2000, INSEE, Paris.
2. Dumais, J, Durr, J.-M. (2001), «La rénovation du recensement français », Actes du symposium 2001 de Statistique Canada.
3. Durr, J.-M. (2003), «La rénovation du recensement français, principes et méthode », Actes de la 54^{ème} session de l'Institut international de statistique, Berlin, 2003.
4. Kish, L. (1990), «Recensement par étapes et échantillons avec renouvellement complet», Techniques d'enquêtes, Vol 16, N° 1, pp. 67-86, Statistique Canada, Ottawa, juin 1990.
