

**UNITED NATIONS SECRETARIAT**  
**Department of Economic and Social Affairs**  
**Statistics Division**

**ESA/STAT/AC.98/9**  
**19 August 2004**

**English only**

---

**United Nations Expert Group Meeting to  
Review Critical Issues Relevant to the Planning of  
the 2010 Round of Population and Housing Censuses  
15-17 September 2004  
New York**

**The 2008 Israel Integrated Census of Population and Housing\***

**By**

**Charles S. Kamen**  
**Israel Central Bureau of Statistics (ICBS)**

---

\* This document is being reproduced without formal editing.

1. The following note describes the principal components of the 2008 Israel Integrated Census of Population and Housing, to be conducted by the Israel Central Bureau of Statistics (ICBS).

## **BACKGROUND**

2. The Israeli census has four principal functions, which are similar to those characterizing censuses everywhere:

- (i) The demographic information on the population is used by the Central Bureau of Statistics to compute the weights applied to estimates based on our sample surveys of the population in order that they reflect the characteristics of the population;
- (ii) Enumeration of the population in localities provides the Ministry of Interior and other ministries with the information they require for budget allocations;
- (iii) Data from the 20% sample is the sole source of detailed socio-economic information for the entire population by geographical division and sub-groups;
- (iv) Making data available to users for a wide range of purposes: government, commercial, academic, educational, research.

3. Israel has conducted five population censuses, in 1948, 1961, 1972, 1983 and 1995. The first census, in 1948, had the primary function of establishing the Population Registry. There is no legal requirement in Israel for conducting a population census at predetermined intervals, nor one requiring that a census be conducted at all. That is the reason for the irregularity of the census dates. A specific government decision is required in order to conduct a census, and in practice it is the ICBS that initiates the process resulting in this decision. The arrangement leads to uncertainty in census planning and hinders the establishment of an ongoing census unit with a multi-year work program. For example, our forthcoming census was originally scheduled for November, 2006, but the budget division of the Finance Ministry decided to postpone it for two years.

4. In 2008, we plan to move from a "conventional" to an "integrated" census. In a conventional census the goal is to physically enumerate every person, either directly or by proxy through a member of their household. The questionnaire for the 100 percent enumeration usually contains a small number of basic demographic questions. A longer questionnaire is usually administered to a sample of those completing the short form. The main reason for restricting the long form to a sample is to reduce costs. All Israeli censuses except for the first used a short questionnaire for everyone and a long questionnaire for a 20 percent sample of households. The 2008 integrated census will enumerate only a 20 percent sample of households; there will be no complete enumeration of the population.

5. Conventional censuses have two major shortcomings: They are expensive, and are very vulnerable to lack of cooperation by the public. Not only are they expensive: their cost can be expected continually to increase, for three reasons. First, the population increases, and since a

substantial portion of the cost of a census goes for enumeration, the increase in the size of the population raises total cost of enumeration even if the cost per capita remains the same. Second, salaries rise, and since a major component of census costs are salaries, the total cost of the census will increase even if the population doesn't increase. Third, censuses will become increasingly dependent on computer technology. The technology used in one census will most likely not be used in the following census a decade or more later: it may no longer be available, or even if available, may have been superseded by technologies which provide more for the money - though they also cost more.

6. Public cooperation with the census is also likely to decline, for at least three reasons which are interconnected: resistance to government intrusion; increasing awareness of, and concern about, privacy; and growing individualism which makes people less willing to participate in communal endeavors such as a census.

7. The obvious alternative to a conventional census is one that is "register-based," and utilizes official lists of persons created for administrative purposes. There are many examples of such lists: births; driver's licenses; pupils in school; tax filers; dwellings; social security contributors; etc. The "best" official list for census purposes is a Population Registry that contains information on everyone in the population; many countries, however, lack such a Registry. Israel has a Registry and each person in the registry has a unique registry ID number. In theory, Israel's Population Registry could provide the same information as the short census form: age, sex, address, place of birth, date of immigration, race/ethnicity, marital status, religion, kinship relations.

8. Why, then, do we need a census in Israel at all? For three reasons: First, the Population Registry is not coterminous with the list of persons defined as comprising the census population. It contains persons who are not part of the population as defined for census purposes, in particular emigrants who no longer live in Israel. Moreover, it does not include persons lacking Registry ID numbers who are resident in Israel continuously for a year or longer, legally or illegally. Second, the geographical information in the Registry, and in particular the addresses, is of poor quality: approximately one fourth of the persons in the Registry are listed at addresses other than where they actually live. Third, the Registry doesn't include the socioeconomic information obtained on the census long form.

9. The goal of the integrated census is to combine the benefits of a conventional census with those of a register-based census, while reducing the shortcomings of each. The integrated census is designed to provide population counts and estimates of population characteristics by combining information from the Population Registry with that obtained from a sample survey in the field.

10. The integrated census will provide estimates for localities and, within localities, for statistical areas (equivalent to "census tracts"). We do not, at this point, envisage that we will be able to provide estimates for geographical areas which are smaller than statistical areas or whose borders are not congruent with those of statistical areas or localities.

## I. THE CONCEPTION UNDERLYING THE INTEGRATED CENSUS

11. The conception behind Israel's 2008 integrated census is simple. We take the information in the Population Registry as a first estimate of the size of the population and its geographical distribution. We undertake two parallel operations: updating the address information in the Population Registry on the basis of address information from other administrative data files; and selecting from the Population Registry a sample of households for enumeration in order to use the information obtained to evaluate the coverage of the Registry and to collect the demographic and socioeconomic data on persons and households that is obtained in the sample questionnaire of a conventional census.

12. The basic census procedure is as follows: We divide the country into Enumeration Areas (EA) - groups of 50 geographically contiguous households based on address information in the Population Registry. The Registry is a list of individuals, each of whom has a unique ID number. It is not a list of dwellings, nor a list of households. But the Registry record for each individual includes the ID number of his or her children and spouse. It is therefore possible to create "administrative households" within addresses ("administrative", because the individuals listed don't necessarily live at that address) and use them to define the EAs. We sample 20 percent of these EAs, taking into consideration various population characteristics in order to make the sampled EAs as representative as possible of the statistical area, and enumerate all the households found in the dwellings at the addresses located in each EA. In localities that are divided into statistical areas, approximately 20 percent of the EAs in each statistical area are sampled. In smaller localities that are not divided into statistical areas, we sample approximately 20 percent of the EAs in the locality. For each "actual" (as opposed to "administrative") household we find in the sampled EA, we fill out a long questionnaire. At the end of the enumeration, we have for each EA two lists of persons: those who were listed in the updated Population Registry (with address corrections) at the addresses included in the EA, and those whom we actually found in the field enumeration. We generate these two lists independently: by that I mean that the enumerator is not provided with a list of persons or households who, according to the Registry, live at the addresses he has to cover, and hence the list of persons which he or she generates as a result of the enumeration is independent of the list of persons in the Registry.

13. Each person at a given address in the sampled EA falls into one of three categories:

- (i) Listed at that address in the Population Registry, and actually enumerated at that address in the 20 percent sample survey;
- (ii) Listed at that address in the Population Registry, but not enumerated at that address in the sample survey - "Registry overcoverage" (compared to the field);
- (iii) Enumerated at that address in the sample survey, but not listed there in the Population Registry. These are "field overcoverage" (compared to the Registry). (There is, of course, a fourth category, of persons living in the EA but neither included in the Registry nor enumerated in the field survey. We have not yet decided whether we will try to estimate their number.)

14. I noted above that we update the address information in the Population Registry on the basis of other administrative data files. We do so to reduce the number of persons in the category of "Registry overcoverage." This is important because we will, in the next stage, attempt to locate them. If we can reduce their number by updating their addresses (thereby probably removing them from the Registry list for a particular EA and adding them to the Registry list for another EA), we won't have to spend time and money looking for them unnecessarily.

15. Since the basis for the integrated census is the updated Registry, we have to evaluate its errors - these errors are the Registry overcoverage. To carry out this evaluation, we must locate the people who were not enumerated at their Registry address in order to find out where they actually live. Although the sampling unit for the integrated census is the EA, the census itself is intended to provide estimates not for EAs, but for localities and statistical areas. Thus, in our search for them it is sufficient to place them in the appropriate statistical area.

16. The result of our comparison between the list of persons we enumerated in the EA, those whom we located after the enumeration and those listed in the Registry for that EA allows us to classify the persons on the Registry list into six outcome categories:

- (i) Persons listed in the Registry at the address in the EA at which they were enumerated;
- (ii) Persons who live at an address other than the one in which they are listed in the Registry for the sampled EA (their "actual address"), but in the same statistical area;
- (iii) Persons who live at an address which is not in the same statistical area as the one in which they are listed in the Registry for the sampled EA, but is in the same locality;
- (iv) Persons whose "actual address" is not in the same locality as the Registry address, but is in Israel;
- (v) Persons whose actual address is not in Israel;
- (vi) Persons whose actual address is unknown to us (these persons will be allocated among the other five categories).

17. By this point we have obtained information on the location of each of the persons appearing on the updated Registry list for the EA. We use the information about the results of our search for the persons appearing in the Registry lists for the sampled EA's in each statistical area in order to infer what the results of the search would be were it carried out for all the persons listed in the Registry for a given statistical area. We compute the probability that a person listed in the updated Registry in the sampled EA's of the statistical area falls into one of the six "outcome categories" listed in para. 16, on the basis of the information obtained from the field survey and the search for "registry overcoverage."

18. These outcome probabilities are computed separately for each of four age groups within each statistical area: 0-20; 20-30; 30-40; 40 and older. The result of this procedure is as follows: Each person listed in the updated Population Registry is assigned a probability of being in one of the six outcome categories according to his age and statistical area of residence, which serves as

a weight. The estimate for any population group is the sum of the weights assigned to its members (age group x statistical area).

## **II. GENERATING THE DATA REQUIRED BY THE INTEGRATED CENSUS**

19. As described above, the integrated census involves two main stages. In the first stage, two lists are created for each EA - one of persons listed in the updated Population Registry, and the second of persons enumerated in the field. In the second stage, we try to locate those on the Registry list ("Registry overcoverage") who were not enumerated in the field. Our efforts to update the Registry focus on two components: address, and presence in the country (which determines whether a person should be included in the census population). Addresses are corrected by comparing the content of the Registry file with the content of three other administrative data files - driver's licenses; electric company accounts; pupils in elementary and secondary schools - using the person's unique Registry ID number to link individuals across data files. Presence in the country is corrected by using information from the border control files. In addition, Registry "balances" files are obtained in order to identify persons who have died since receipt of the previous Registry file.

20. Field enumeration involves three basic stages:

- (i) A pre-enumeration canvass of the EA, during which the enumerator lists dwelling units;
- (ii) Enumeration, during which the enumerator returns to the listed dwelling units and interview the residents;
- (iii) A "clean-up" stage, in which the enumerator makes a final effort to enumerate dwellings that were closed. The field enumeration has two goals: to obtain socioeconomic information on the population, and to obtain the information needed to evaluate the address information in the updated Population Registry so it can be used as the basis for census estimates.

21. We obtain information on the location of persons in the Registry overcoverage group in a number of stages. In the first stage, the enumerators return to their EA's with lists of persons whose Registry address is in the EA but who were not enumerated, and try to locate them - either by actually finding them in the field, or by obtaining from others information on their location. In the second stage, the names of persons not located by the enumerators are transferred to a CATI system, and information on them is sought by phone. Additional potential sources of information on the location of Registry overcoverage persons are updated versions of the administrative data files used to correct the Population Registry, as well as other administrative data files that were not used for correcting the Registry.

22. At the end of this search process we are, in principle, able to assign each person listed in the Registry a probability that his Registry address reflects where he actually lives, and if it does not, a probability that his address can be characterized by one of the alternatives listed earlier. This is the basis for creating the final 100 percent census demographic file. The 20 percent sample survey of the EAs is the basis for the final census file containing demographic and socioeconomic information on persons and households.

### **III. 2008 CENSUS TECHNOLOGY**

23. The 2008 Israeli census is computer-based in almost all aspects of its operations. Computer technology is what has made the integrated census feasible. The main computer-based components of the census include mapping applications involving preparing maps for enumerators, geographical anchoring of addresses and creating the network of EA's; creating the updated Population Registry file by merging administrative files, including the development of algorithms for record-linkage among those files; computer-based field work, including enumeration by means of laptop-based CAPI using Blaise software with computerized questionnaires in three languages - Hebrew, Arabic and Russian; questionnaire data transmitted by enumerators via the internet from their home telephones; field work administration based on data transmitted by enumerators via the internet from their home telephones; field staff management, hiring, allocation and payment integrated with census geography and enumerator production; training of field staff, including computer-based training materials and procedures; dissemination of results, including the development of applications for web-based table generators, summary tabulation generators and tables available on the internet.

### **IV. ENUMERATING SPECIAL POPULATIONS**

24. The dual-list enumeration procedure described above is appropriate for approximately 70% of the population that lives in urban localities having an organized network of named streets and numbered buildings that can be mapped onto the addresses in the Population Registry. The others either live in localities without an organized system of addresses, live outside localities, live under arrangements which don't permit their enumeration by the standard dual list procedure, or are not listed in the Population Registry. We are in the process of deciding for which of these populations it is necessary to develop special enumeration procedures:

- (i) Large Arab localities having more than one statistical area and without an organized system of addresses;
- (ii) Small localities (both Arab and Jewish) having only one statistical area and no organized system of addresses;
- (iii) Persons in communal quarters (institutions);
- (iv) Residents of kibbutzim (collective settlements);
- (v) Persons living outside the borders of localities;
- (vi) Concentrations of bedouin living outside the borders of localities in southern Israel in the Beer Sheva area;
- (vii) Foreign workers;
- (viii) The homeless.

### **V. WHERE DO WE STAND NOW? 2004 FIELD TEST AND 2006 DRESS REHEARSAL**

25. Field tests are conducted as part of census planning. These tests are particularly important in the case of the integrated census since both its conception and many of its procedures are new. Testing of the 2008 census procedures involves three main components: the behaviors required of the enumerators and other field staff in order to obtain census

information; the process of creating the integrated administrative file used for the census; and the functioning of the computer programs and technology which underlie the work of the field staff.

26. The first field test, carried out in Bet Shemesh (a town of 50,000 inhabitants located 30 kilometers west of Jerusalem) in the spring of 2002, had four goals:

- (i) Creating an improved address file (IAF) on the basis of the Population Registry and additional administrative files;
- (ii) Carrying out a field survey to obtain the information needed to correct the IAF;
- (iii) On the basis of the field survey, creating a final weighted census data file;
- (iv) Computing population estimates using the final census data file. The results of the Bet Shemesh test showed the need to improve the procedures for locating the Registry overcoverage. That was set as a principal goal for the next field test in 2004.

27. The November, 2004, field test will be carried out in five localities: Giv'atayim (47,000 inhabitants, bordering Tel Aviv), Tira (an Arab town with 20,000 inhabitants near Tel Aviv), Yarhiv, Neue Yamin and Elishama (three small Jewish localities each having fewer than 1000 inhabitants). The goals of this field test are to evaluate:

- (i) Procedures developed for locating the Registry overcoverage;
- (i) Enumeration procedures developed for large Arab localities that don't have an organized system of addresses;
- (ii) Enumeration procedures in small localities that don't have an organized system of addresses;
- (iii) Procedures developed for managing field work;
- (iv) On a small scale, the functioning of one local field office;
- (v) Effectiveness of the procedure for recruiting enumerators;
- (vi) In addition, improved versions of procedures implemented in the 2002 field test in Bet Shemesh will be carried out in 2004.

28. The third field test, in the fall of 2006, will be the major field test for the 2008 census. Our current plan is to include localities with a total of some 600,000 inhabitants, approximately 10 percent of the population. The test will be conducted in clusters of localities in two separate geographical regions, north and southeast of Tel Aviv. Strictly speaking, it will not be a "dress rehearsal," because it will include components which are being field-tested for the first time. The main goals of the 2006 field test are to:

- (i) Test enumeration procedures for special populations;
- (ii) Evaluate the functioning of the procedures for identifying and locating registry overcoverage in localities varying in the likelihood of their residents being listed at their actual address;
- (iii) Test the final versions of the enumeration procedures implemented in the 2004 field test and revised on the basis of the results of that test;
- (iv) Test on a large scale the functioning of field offices, both vis-à-vis the field staff in each office and vis-à-vis census headquarters at the ICBS in Jerusalem;



- (v) Evaluate the procedures for arriving at census estimates;
- (vi) Test preliminary versions of the internet-based data-dissemination tools.

## **VI. MAKING THE RESULTS ACCESSIBLE AND COMPREHENSIBLE TO THE PUBLIC**

29. A variety of means exist for making the census results available to the public. These means have to meet two kinds of needs: for differing degrees of flexibility in generating the desired information; and for different kinds of content. To some degree, these different needs are represented by different kinds of users, with varying degrees of sophistication in using internet-based "do-it-yourself" tools to generate tables, but we also recognize that any particular user may be able to live with different degrees of flexibility in designing his desired product, according to his purpose. Therefore, we plan to allow users flexibility in obtaining the desired information by providing a product mix which includes final publications; predefined aggregate tabulations (via a data warehouse); flexible tabulations (via a table generator, an example of which can be accessed at [www.cbs.gov.il](http://www.cbs.gov.il) - click on **English** and then on the link to the **Social Survey Table Generator**); access to microdata files; and tailor-made tabulations by special order. All of these products and services are already available at the ICBS; we plan to adapt them so they are appropriate for the needs of census data users. All of them will be capable of providing tabulations by census geography, by sub-groups of the population and by subject, individually or in combination. In addition, we hope to develop the capability to prepare analytic reports.

30. One challenge posed by the integrated census is making the results comprehensible to users accustomed to products based on a conventional census methodology. Three aspects of the integrated census require particular attention: geographic detail; linkage to current population estimates; weighted estimates. Unlike in a conventional census, we will enumerate only 20 percent of the dwelling units, and plan to provide census estimates for localities and their statistical subdivisions, down to the level of the statistical area. In a conventional census that enumerates the entire population over all the country's geography, it is possible by using GIS applications to obtain data for any desired geographical area by drawing a border around it and generating the required tabulations for the resulting polygon. The integrated census, however, samples only 20 percent of the geography, and we have not yet determined whether we will be able to provide census estimates for polygons not coterminous with statistical areas.

31. The ICBS is developing a new procedure for providing current population estimates using data from the Population Registry. Under the existing system, the census provides aggregate population estimates for statistical areas according to combinations of demographic characteristics, and these aggregate estimates are updated on the basis of counts received from the Population Registry on the number of births, deaths and changes of address in each statistical area. Under the new system, the census will provide a list of persons for each statistical area, and on the basis of the information from the Population Registry on births, deaths and changes of address we will add to or subtract from statistical areas particular identified persons. The 100 percent enumeration in a conventional census provides the list of identified persons that serves as the basis for intercensal updating. The integrated census will not provide such a list, since it will not physically reach each dwelling. Instead, it will provide probability estimates that a person's listed address in the Population Registry is correct. We have not yet adapted the new procedure for population estimates to the kind of information that the integrated census will provide.

32. A conventional census based on 100 percent enumeration with a short form, and a sample enumeration with a long form, produces two final data files - a demographic file containing the entire census population, and a socioeconomic file containing the sample. Whatever error exists in the 100 percent enumeration (overcoverage; undercoverage) is not documented in the final demographic data file, nor is this file viewed as subject to sampling error. The estimates in the final sample socioeconomic data file are subject to sampling error. The 100 percent demographic file of the integrated census will also be subject to sampling error, since it is based on information from the 20 percent sample survey. Moreover, the 100 percent demographic file is comprised, essentially, of the updated Population Registry, with each person listed having attached to them a probability that their address listing is correct. These probabilities serve as weights. Aggregate estimates based on the 100 percent file (for geographical areas, population groups, etc.) are, therefore, weighted estimates subject to sampling error. In practice, we do not expect this to cause any particular difficulty for users.

## **VII. FUTURE PLANS FOR THE INTEGRATED CENSUS**

33. Israel's 2008 integrated census will serve as the basis for the censuses to follow. The integrated census combines data from administrative registers with data from field enumeration. The cost of the field enumeration is a major part of the cost of any census. Therefore, the long-term goal is to use administrative registers to substitute for the field enumeration and gradually reduce its scope. Planning for future integrated censuses will have to consider a number of issues regarding major components of the integrated census procedures:

- (i) Improving the Population Registry, so that more addresses are accurate;
- (ii) Identifying administrative sources of socioeconomic data with adequate coverage of the population as an alternative to collecting this information from households;
- (iii) Developing statistical tools to provide valid estimates on the basis of partial information;
- (iv) Reducing the sample proportion of the field enumeration in order to reduce costs;
- (v) Providing census estimates more frequently.