



STATE OF ISRAEL



The 2008 Israel Integrated Census of Population and Housing

Basic conception and procedure

Charles S. Kamen

February, 2005

This document describes the principal components of the 2008 Israel Integrated Census of Population and Housing, to be conducted by the Israel Central Bureau of Statistics. Dr. Hagit Glickman, of the Census Methodology Unit, prepared the sections on the statistical methodology of the census procedure.

Background

1. In addition to its basic function of providing counts of population and households, the Israeli census has four principal functions which are similar to those characterizing censuses everywhere: (1) the demographic information on the population is used by the Israel Central Bureau of Statistics (ICBS) 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; (2) enumeration of the population in localities provides the Ministry of Interior and other ministries with the information they require for budget allocations; (3) data from the 20% sample is the sole source of detailed socio-economic information for the entire population by geographical division and sub-groups; (4) making data available to users for a wide range of purposes: government, commercial, academic, educational, research.

2. Israel's current population is 6.8 million. 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 (PR). 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 government decided to postpone it for two years.

3. 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.

4. Conventional censuses have three major shortcomings: They are expensive, they are very vulnerable to lack of cooperation by the public and the information they collect rapidly becomes outdated. 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.

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

6. The high cost of censuses means that they can be conducted only at relatively long intervals. As a result, the data they provide becomes outdated. The lack of timeliness is particularly serious at the local level, where changes are more rapid. The advantage of a census is its ability to provide detailed data at the local level, but its cost results in such data being collected only infrequently. As time passes, the data become less useful.

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 Population Registry has a unique Population 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 at all? Israel conducts a de jure census. It counts people at their usual place of residence. At the national level this is defined, for persons listed in the Population Registry, as those present in Israel or absent for less than one year; and for those not listed in the Population Registry, as persons who have lived in Israel for one year or longer. Israel's Population Registry cannot now substitute for a census, for five reasons: First, the Population Registry is not coterminous with the list of persons defined as comprising the de jure census population. It contains persons who are not part of the de jure population, in particular emigrants who no longer live in Israel. Moreover, it does not include persons

lacking Population Registry ID numbers who are resident in Israel continuously for a year or longer, legally or illegally. Second, the geographical information in the Population Registry, and in particular the addresses, is of poor quality: approximately one fourth of the persons in the Population Registry are listed at addresses other than where they actually live. Third, one of the goals of the census is to collect data on households, which can't be identified on the basis of information in the Population Registry. Fourth, the Population Registry doesn't include the socio-economic information obtained on the census long form. Fifth, the Population Registry does not include information on housing.

9. 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 of demographic and socioeconomic characteristics from the sample enumeration for geographical areas which are smaller than statistical areas or whose borders are not congruent with those of statistical areas or localities.

The conception underlying the integrated census

10. The conception behind Israel's 2008 integrated census is simple. Instead of carrying out a complete field enumeration to obtain population counts, we use the Population Registry as the basis for population estimates. Because we know that the Population Registry is inaccurate, we have to correct the estimates that are based on it. We conduct a sample survey of approximately 20 percent of the population. The sample survey has two functions: to obtain the detailed demographic, social and economic information that is provided by a census, and to obtain the information we need to correct the estimates that are based on the Population Registry. We use the demographic, social and economic information we collect in the sample questionnaire to prepare estimates of these characteristics for the entire population. We also learn from the sample survey whether the addresses of the persons in the sample are listed accurately in the Population Registry. Just as we estimate the demographic, social and economic characteristics of the entire population on the basis of the information obtained from the sample, we will use the information on the addresses that we obtained from the persons in the sample to estimate the likelihood that the persons who were not sampled are listed correctly in the Population Registry. Thus, unlike in a conventional census, where the sample has the single function of providing information on population characteristics, in the integrated census the 20 percent sample serves two functions: it provides information on the demographic, social and economic characteristics of the population (as in a conventional census); and it provides information about whether the Population Registry addresses are correct. The information on addresses is used to compute the undercoverage and overcoverage parameters that are used in computing the census estimates.

11. The basic procedure of the integrated census is straightforward. 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. (1) We update the address information in the Population Registry on the basis of address information from other administrative data files. We call this updated Population Registry the Improved Administrative File (IAF). (2) We select a cluster sample of contiguous addresses containing either single or multiple dwellings and conduct two operations: enumeration of the households living in them, and evaluating the validity of the information in the IAF regarding these addresses. The information obtained from this sample enables us to evaluate the coverage of the IAF and to collect the demographic and socio-economic data on persons and households that is obtained in the sample questionnaire of a conventional census.

12. Since the basis for the integrated census is the IAF, we have to evaluate its errors – the IAF undercoverage and overcoverage. Undercoverage and overcoverage are concepts applying to the population as a whole, as well as to subgroups in the population; they can apply to the entire country or to geographical subdivisions; they can apply to combinations of population groups and geographical area. An important population group that the Population Registry misses completely is those people who do not have an Israeli ID number, whether they are in the country legally or illegally. Since the existing administrative information on such people is inadequate, no attempt is made to include them in the IAF (nor do we intend to estimate their number), though they will be enumerated if they are found in the field and meet criteria of duration of residence.

13. In the following discussion, we consider only those persons with an Israeli ID number. On a national level, all Israelis with an ID number are included in the Population Registry. National overcoverage comprises emigrants who no longer live in Israel but are still in the Population Registry (approximately 500,000 persons, some 7 percent of the persons listed in the Registry). However, because of the poor quality of Population Registry addresses, most coverage errors occur at the local level. Approximately one-fourth of the population appear in the Registry at an address other than that at which they actually reside. The estimation procedure of the integrated census is therefore focused on estimating local coverage errors.

14. We define the coverage errors of the IAF with respect to a statistical area (or with respect to a locality, if it is not divided into statistical areas). IAF undercoverage of a statistical area refers to persons living in that area but listed in a different statistical area in the IAF. Similarly, IAF overcoverage of a statistical area refers to persons listed in that area in the IAF, but living in a different statistical area or abroad. The estimation procedure of the integrated census extends the dual system model for estimating undercoverage, in order to provide an

estimate of overcoverage as well. It aims to provide estimates of undercoverage and overcoverage errors for each statistical area, and correct the IAF counts accordingly.

The procedure of the integrated census

15. The basic census procedure is as follows. Using the existing administrative-statistical division of the country into localities, and statistical areas within localities having 10,000 inhabitants or more, we divide them into Enumeration Areas (EA), groups of 50 geographically contiguous households whose boundaries do not cross boundaries of localities or statistical areas. The division into EA's is based on address information in the Population Registry. The average enumerator's work load will comprise five EA's with about 50 households in each. The Population 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 Population Registry record for each individual often 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, or in the same apartment) and use them to define the EA's, and assign EA's to enumerators in accordance with their planned work load. We sample approximately twenty percent of the EAs within each statistical area (or within the locality if it has fewer than 10,000 inhabitants), taking into consideration various population characteristics in order to make the sampled EAs as representative as possible of the statistical area (or of the locality). We enumerate all the households found in the dwellings at the addresses located in each EA. 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 IAF 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 IAF, live at the addresses he or she has to cover, and hence the list of persons which the enumerator generates as a result of the enumeration is independent of the list of persons in the IAF.

16. Because the sampled EA's are not necessarily contiguous, it is important to insure that the enumerator does not erroneously enumerate dwellings located outside of his EA. A number of safeguards are in place to prevent this from occurring: the enumerator's computer will not accept addresses which do not appear on its pre-loaded address list for the EA, and crew chiefs must approve adding buildings and addresses to the EA map. If the "trespass" is limited to dwellings within the same statistical area in which the EA is located the effect on the population estimate will be less damaging than if the error involves an incursion into another statistical area. Such errors will be most likely to occur in localities characterized by a

substantial amount of new construction, lack of a clearly defined street grid and an absence of an organized system of addresses – a combination of circumstances more likely to result in a misreading of the EA map.

17. The list generated by the field enumeration is matched with the IAF list, using an automated record linkage process. Matching is done nationally, on the basis of ID number and other key variables such as sex and age.¹ As a result of the matching process, we are able to determine the number of people who were both enumerated in a given statistical area and listed in that area in the IAF, and estimate the local undercoverage of the IAF. In order to estimate overcoverage, either local or national, we use additional information, as described in the following paragraphs.

18. I will distinguish in what follows between "IAF overcoverage," which refers to persons listed in the IAF who do not actually live in their IAF statistical area, and "IAF excess," which refers to persons listed in the IAF in the sampled EA's who were not found in the field. IAF excess is observed; IAF overcoverage is inferred. Each person at a given address in the sampled EA falls into one of three categories: (1) listed at that address in the IAF, and actually enumerated at that address in the 20 percent sample survey, or determined to live at another address in the same statistical area; (2) listed at that address in the IAF, but not enumerated at that address in the sample survey – "IAF excess", compared to the field; (3) enumerated at that address in the sample survey, but not listed there in the IAF. These are "field excess," compared to the IAF.

19. In order to estimate the coverage errors of the IAF, we must locate the people who were not enumerated at their IAF address in order to determine 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 EA's, but for localities and statistical areas. Thus, in our search for the persons categorized as "excess", it is sufficient to place them in the appropriate statistical area.

20. The result of our comparisons between the list of persons we enumerated in the sampled EA's and the IAF excess whom we located after the field enumeration, on the one hand, and those listed in the IAF, on the other, allows us to classify the persons in the IAF list into five outcome categories: (1) persons listed in the IAF at the address in the statistical area at which they lived; (2) persons who live at an address which is not in the same statistical area as the

¹ For a description of the matching procedure, cf. T. Yitskov and H. Azaria, "Record linkage in an integrated census." *Proceedings of the 2003 Conference of the U.S. Federal Committee on Statistical Methodology* (2003).

one in which they are listed in the IAF, but is in the same locality; (3) persons whose "actual address" is not in the same locality as the IAF address, but is in Israel; (4) persons whose actual address is not in Israel; (5) persons whose actual address is unknown to us. Standard imputation techniques will be used to allocate persons in category (5) among the other four categories.

21. By this point we have determined the actual statistical area of residence for each of the persons appearing in the IAF list for the sampled EA's. This information enables us to estimate overcoverage of the IAF for various geographical divisions. We use the information about the results of our search for the persons appearing in the IAF lists for the sampled EA's in order to infer what the results of the search would be were it carried out for all the persons listed in the IAF. The national overcoverage of the IAF is estimated based on the number of persons in category (4) – those not in Israel – together with administrative data obtained from the Border Control authorities. The local overcoverage of the IAF, defined with respect to a statistical area or a locality having fewer than 10,000 inhabitants, is estimated on the basis of the number of persons listed in the IAF in the given statistical area who are found in categories (2), (3) and (4). We will also estimate IAF overcoverage with respect to other geographical areas – for example, districts or localities. These estimates will be used for evaluating the plausibility of the integrated census final estimates.

22. Estimates of local undercoverage and overcoverage for each statistical area are computed within demographic groups that differ in their IAF address error patterns. In a field test conducted in Bet Shemesh (a town near Jerusalem) in 2002, estimates were computed separately for each of four age groups within each statistical area: 0-20; 20-30; 30-40; 40 and older. We combined the estimates of overcoverage and undercoverage with the number of persons listed in the IAF to obtain an estimate of the population size for the statistical area. We generated a census weight for each of the IAF records – a coefficient by which the record is multiplied in order to reflect the number of persons it represents in the population. The census estimate for any population group is the sum of the IAF census weights assigned to its members.

Statistical Methodology of the integrated census

23. The statistical methodology of the integrated census has two main components: estimating the IAF coverage errors, and estimating population characteristics. Both the coverage estimates and the estimates of population characteristics use information from the IAF and from the field enumeration survey. The coverage estimates are based on the methodology of Dual System Estimation, and represent an extension of the classical model for estimating undercoverage in census data; the extension accommodates overcoverage or

“false captures” in the administrative data. The lists generated in the classic “capture – recapture” stages are represented in our model by the two independent lists we create – one is based on the IAF, and the other on the field enumeration. The procedure for estimating population characteristics uses regression calibration techniques to produce sample weights. Regression estimators are used because they allow consistency with the IAF based population estimates for a number of auxiliary variables simultaneously while reducing the variance.

24. In the following paragraphs, we sketch the main idea of our coverage estimation procedure in a non-formal, intuitive manner.²

25. Our procedure produces population estimates for each statistical area. Recall that we define the IAF coverage errors with respect to a statistical area. The IAF undercoverage for a given statistical area is composed of all persons living in that area but listed elsewhere in the IAF. The IAF overcoverage for a given statistical area is composed of all persons listed in the IAF in that area but not living there. We consider two parameters, one for undercoverage and one for overcoverage. The undercoverage parameter is

$$p_{1+}(i) = \frac{\text{The number of persons listed in the IAF in area } i \text{ who actually live there }^{(*)}}{N(i)}$$

where $N(i)$ is the size of area i , that is, the number of persons living in statistical area i . Similarly, the overcoverage parameter is

$$\lambda(i) = \frac{\text{The number of persons listed in the IAF in area } i \text{ who don't live there }^{(*)}}{N(i)}$$

Note that

$$N(i) = \frac{\text{The number of persons listed in the IAF in area } i \text{ }^{(*)}}{p_{1+}(i) + \lambda(i)}$$

(*) Formally, the numerator is the expected number of persons satisfying the specified conditions.

Since the parameters, $p_{1+}(i)$ and $\lambda(i)$, are unknown we estimate them on the basis of the sample data. We use the estimates $\hat{p}_{1+}(i)$ and $\hat{\lambda}(i)$ to compute the census weight,

² For more detailed presentation of the model, and of its mathematical basis, cf. H. Glickman, R. Nirel and D. Ben Hur, "False Captures in Capture-Recapture Experiments with Application to Census Adjustment." *Bulletin of the International Statistical Institute*, 54th Session, Contributed Papers, Vol. LX (2003), pp. 413-414. R. Nirel, H. Glickman and D. Ben Hur, "A Strategy for a System of Coverage Samples for an Integrated Census." *Proceedings of Statistics Canada Symposium 2003: Challenges in Survey Taking for the Next Decade* (2004).

$$\hat{w}(i) = \frac{1}{\hat{p}_{1+}(i) + \hat{\lambda}(i)}$$

An estimate for the size of a given statistical area is now obtained as the sum of the corresponding census weights.

26. Estimating the undercoverage parameter: The field enumeration and the procedure for locating IAF excess described in Par. 30-32, below, enable us to generate for each statistical area entries for three of the four cells in the following fourfold table:

	Enumerated in the SA	Not enumerated in the SA
In the IAF for the SA	IAF identical to field	IAF net excess
Not in the IAF for the SA	Field excess	

All four entries of the above table correspond to the population of the sampled EA's of a given statistical area. The IAF net excess is obtained by excluding the IAF erroneous enumerations, i.e., persons listed in the IAF for a given statistical area but living in another statistical area or abroad. It is assumed that the field enumeration has no erroneous enumerations. Using the entries in the fourfold table, and assuming independence between the IAF and the field, we estimate

$$\hat{p}_{1+} = \frac{\text{IAF identical to Field}}{\text{IAF identical to Field} + \text{Field excess}}$$

Estimating the overcoverage parameter: Using the number of IAF erroneous enumerations in the sampled EA's of a given statistical area we generate an estimate for the overcoverage parameter:

$$\hat{\lambda}(i) = \frac{\text{IAF erroneous enumerations}}{(\text{IAF identical to Field} + \text{IAF net excess}) / \hat{p}_{1+}(i)}$$

The denominator in the above formula is an estimate of the number of people living in the sampled EA's of a given statistical area.

27. The accuracy of the population estimates of a statistical area depends on the following conditions: (1) independence between the IAF list and the field list; (2) all persons in the population have the same probability to be listed in the IAF in the correct statistical area, as well as the same probability of being enumerated in the field; (3) the distribution of the IAF overcoverage of any statistical area across its EA's is proportional to the distribution of the "true" population. The assumption of independence is met because the field enumerators have

no knowledge of the contents of the IAF. The two other assumptions are approximated by dividing the population into homogeneous groups with respect to the likelihood that they are subject to errors of overcoverage and undercoverage, basing this division on IAF variables related to coverage error. Separate estimates of the undercoverage and the overcoverage parameters are calculated for each of the estimation groups.

28. After calculating the census weights, we plan a series of comparisons between the estimates generated from the census weights and alternative population estimates. The comparisons will be made for various demographic subgroups and aggregated geographic levels such as localities, districts and the country as a whole. The census weights will be adjusted, if necessary, if the comparisons with the population estimates justifies doing so. The alternative population estimates include a direct estimate of the total population of Israel that is based solely on the IAF and its national overcoverage estimate; and the annual mid-year population estimates produced by the ICBS Demographic Division.

29. The estimation procedure proceeds by the following steps: (1) establishing estimation groups homogeneous with respect to coverage errors; (2) computation of the overcoverage and undercoverage parameters for each group; (3) computation of a "census weight" for each person in the IAF according to the estimation group to which he or she belongs; and (4) conducting a series of checks to assure the quality of the census estimates. At the end of this procedure we are able to compute population estimates for subgroups with various combinations of characteristics, and to create a final census file – an ICF – Integrated Census File – with each record assigned a census weight.

Generating the data required by the integrated census

30. 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 IAF, and the second of persons enumerated in the field. In the second stage, we try to locate those on the IAF list ("IAF excess") who were not enumerated in the field. Our efforts to update the Population 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 Population Registry file with the content of three other administrative data files – motor vehicle registrations; electric company accounts; pupils in elementary and secondary schools – using the person's unique Population 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, Population Registry update files are obtained in order to identify births, deaths and marital status changes that have occurred since we received the previous Population Registry file; they also include new immigrants.

31. Field enumeration involves three basic stages: (1) a pre-enumeration canvass of the EA, during which the enumerator lists dwelling units; (2) enumeration, during which the enumerator returns to the listed dwelling units and interviews the residents; (3) a "clean-up" stage, in which the enumerator makes a final effort to enumerate dwellings that were closed and to convert non-response. The field enumeration has two goals: to obtain socio-economic information on the population, and to obtain the information needed to evaluate the address information in the IAF so it can be used as the basis for census estimates.

32. We obtain information on the location of persons in the IAF excess group in a number of stages. In the first stage, the enumerators return to their EA's with lists of persons whose IAF 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. In order that the CATI procedure not be open-ended, we will use the census field tests to evaluate the cost-effectiveness of continuing the telephone search in order to set a cut-off point on the basis of diminishing returns. Additional potential sources of information on the location of IAF excess 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 Population Registry. Persons not located as of the conclusion of the search for IAF excess will be allocated among the groups specified in Par. 20.

33. At the end of this search process we are, in principle, able to assign each person listed in the IAF a probability that his IAF address reflects where he actually lives, and if it does not, a probability that his address can be characterized by one of the four other alternatives listed above (Par. 20). 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 socio-economic information on persons and households.

2008 Census technology

34. 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 used to prepare maps for enumerators, geographical anchoring of addresses and for creating the network of EA's; harmonizing administrative files obtained from outside the ICBS; creating the IAF 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 Hebrew, Russian and

Arabic; devising an innovative solution that enables us to display fonts for these three languages within Blaise; 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.

Enumerating special populations

35. 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 we will provide population estimates, and for which will it be necessary to develop special enumeration procedures: (1) large Arab localities having more than one statistical area and without an organized system of addresses; (2) small localities (both Arab and Jewish) having only one statistical area and no organized system of addresses; (3) persons in communal quarters (institutions); (4) residents of kibbutzim (collective settlements); (5) persons living outside the borders of localities; (6) concentrations of Bedouin living outside the borders of localities in southern Israel in the Beer Sheva area; (7) foreign workers; (8) the homeless.

Where do we stand now? 2004 Field Test and 2006 Dress Rehearsal

36. 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.

37. 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: (1) creating the IAF on the basis of the Population Registry and additional administrative files; (2) carrying out a field

survey to obtain the information needed to correct the IAF; (3) on the basis of the field survey, creating a final weighted census data file; (4) 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 IAF excess. That was set as a principal goal for the next field test in 2004.

38. 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, Neve Yamin and Elishama (three small Jewish localities each having fewer than 1000 inhabitants). The goals of this field test are to evaluate: (1) procedures developed for locating the IAF excess; (2) enumeration procedures developed for large Arab localities that don't have an organized system of addresses; (3) enumeration procedures in small localities that don't have an organized system of addresses; (4) on a small scale, procedures developed for managing field work; (5) on a small scale, the functioning of one local field office; (6) on a small scale, the effectiveness of the procedure for recruiting enumerators.

39. 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 that are being field-tested for the first time. The main goals of the 2006 field test are: (1) to test enumeration procedures for special populations; (2) to evaluate the functioning of the procedures for identifying and locating the IAF excess in localities varying in the likelihood of their residents being listed at their actual address; (3) to 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; (4) to 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; (5) to evaluate the procedures for arriving at census estimates; (6) to test preliminary versions of the internet-based data-dissemination tools.

Making the results accessible and comprehensible to the public

40. 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;³ 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. All of the data made available to the public will meet the strict confidentiality requirements specified by the Statistics Ordinance, to insure that it will not be possible to identify individual respondents to the census.

41. 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. While we expect to be able to provide estimates from the 100 percent enumeration for polygons not coterminous with statistical areas, we have not yet determined whether we will be able to do so for the sample enumeration.

42. 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. The new system is a compromise between aggregate (component) methods and individual (PR) based methods. The 1995 base population was established primarily by aggregate component estimation of the census error. Not everyone in the 1995 base population (the adjusted census population) can be identified with an individual Population Registry record. As a result, the new population base contains "fictitious" records, so that the

³ Via a table generator, an example of which can be accessed at www.cbs.gov.il – click on **English** and then on the link to the **Social Survey Table Generator**.

aggregate base can be treated as though it were a corrected 100 percent count of the population in November, 1995. The 100 percent enumeration in a conventional census provides a 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 weights for persons listed in the Population Registry. We have not yet determined how to adapt the new procedure for population estimates to the kind of information provided by the integrated census.

43. 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 socio-economic 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 socio-economic 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 IAF, with each person listed having attached to them a census weight. Aggregate estimates based on the 100 percent file (for geographical areas, population groups, etc.) are, therefore, weighted estimates subject to sampling error. Since users are accustomed to receive population estimates based on a complete enumeration, it will be necessary for us to explain the implications of using estimates based on a sample and assist in their interpretation more than would be required in a conventional census.

Future plans for the integrated census

44. 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: (1) improving the Population Registry, so that more addresses are accurate; (2) identifying administrative sources of socio-economic data with adequate coverage of the population as an alternative to collecting this information from households; (3) developing statistical tools to provide valid estimates on the basis of partial information; (4) reducing the sample proportion of the field enumeration in order to reduce costs; (5) providing census estimates more frequently.