_____

United Nations Expert Group Meeting on
The Scope and Content of Social Statistics
9-12 September 2008
United Nations, New York

Population: Size, structure and distribution

Prepared by

Pnina Zadka
Central Bureau of Statistics, Israel

_____
* This document is being reproduced without formal editing.

Population : Size, structure and distribution

This paper addresses issues that relate to population static in contrast to population dynamics. It deals with the description of a population at a fixed time. Most of the considerations would refer generally to national data, with a few exceptions. Nevertheless, most of the concepts apply to sub-regions within a country, as well.

1. Population Size

Population size is the basic demographic fact and the most important figure for a community. As such, estimates of population size go back to the biblical period. The population size of a country/nation is the primary fact any government tries to obtain.

Population size refers to the number of people living within defined geographic boundaries at a specific time. The size of the territory within these boundaries may vary; from the entire world to a small village, a neighborhood, a building or even a single dwelling. The concept has three main dimensions (1) spatial- contiguous geographic territory, (2) time and (3) number of people. Clear definitions and feasible applications of geographic boundaries are the first requirement for an estimate, of the number of people residing in a pre-defined area at a pre-defined date. Population size is always an estimate and never an actual full count, regardless of the definitions used. There are always miscounts, omitting people from the count due to various reasons or counting people who should not, that will have no impact on the estimated population size.   Precise definitions and stable standards are essential for the reliability of population estimates.

1.1 Definitions, Terms and Standards

i.   <u>Time</u> – time can be defined as a "point" (specific date) or an interval of time such as a decade, a year or a month.
ii. <u>Geographic territory</u> - The space defined closed borders. The borders can be presented by a line on a map (a paper or a digital map) preferable accompanied by an exclusive textual description of the boundaries (physical markings such as waterways, walls etc.). Geographic boundaries should form a polygon (as opposed to a line). The geographic boundaries always relate to a specific point in time. Within countries geographic boundaries may result from jurisdiction or administrative issues.
iii. <u>Population size (count)</u> – Number of people residing in a specific area at a specified time (point in time or interval of time). For an interval, usually an average of the population size at several points within the interval or a point representing the middle of the interval. Events that have an impact on the population size are deaths, births, immigration and emigration. Data on changes are usually obtained from official registration by authorities. In some cases, specific types of residents are excluded from population counts, such as diplomatic corps, foreign military corps, nomadic people (wanderers between countries).
iv. *De jure* <u>population</u> – All people who belong to a given area at a given time by virtue of legal residence, usual residence or similar criteria.
v.   *De facto* <u>population</u> – All the people actually present in a given area at a given time.

vi. <u>Place of usual residence</u> – people whose usual residence is within the set boundaries, who might reside in other areas part of the time, or whose affiliation is to that area (family ties)
vii. <u>Permanent residents</u> – People who have a formal residency status and are registered as living in the specified area.
viii. End of year population / beginning of year population – refers to the number of people counted at a specific point in time.
ix. <u>Average population</u> – A geometric average of population counts at specific points over the time interval.
x. <u>Mid-year population</u> – population at the middle of the time interval.

The standards for calculating population estimates were set by demographers and are based on census counts. In many countries administrative registration of vital events (deaths and births) was used to update the annual population size obtained by census. Who is an immigrant and should be included in the count and who is an emigrant and should be deleted from the count, were in the past and still are, political issues, and there are no international standards in effect on the inclusion and exclusion of certain people from the count. There are sets of recommendations by various international organizations and most countries use them as guidelines. The current recommended standard for the 2010 census round is related to the term "usual residents" which is based on the usual place of residence in the census questionnaire. This is mainly to assure that a person is not counted in more than one place at the same time.

1.2 Data Sources

Population estimates are based on two main data sets; census data and current flows of changes in the population. Both data sets can be obtained either through surveys or from administrative sources. In many countries census data are still collected in the traditional way, full enumeration of all dwellings within the set geographic boundaries. Data on current changes are mostly based on administrative registration of vital events and migration events. Census data are subject to over counts and under counts. In most developed countries census data are collected over 3-4 weeks. The census counts relate to the census date, a date chosen for logistic convenience. Administrative sources are subject to lag in registration and reporting, due to various reasons; intentional miss-registration and erroneous registrations. Data on international migration is usually based on border control reports and depends heavily on the quality of these reports. Legal options to cross the border without being reported may affect the quality of the data. Data on internal migration are mostly based on self-reporting, and are subject to special interest and conformity with regulation.

1.3 Israel's Population Size Estimates

Israel's population size estimates are calculated by the traditional method. Census results are the base line for each census date and current population estimates between censuses, are based on updates in the Population Register (PR).

The PR in Israel was formed shortly after the formal foundation of the country. The register was established during the first census in November 1948. The PR contains demographic information, names (surname, given name and parents' given names) and a unique ID number for each person. The ID number was granted during the first census, (which was held under curfew) to all the residents in the country at that time. Newborns are granted an ID number within 24 hours of birth and immigrants arriving with an Immigration Visa are granted the ID number after crossing the border by the

Immigration Authorities. People, who change their residency status during their stay, are granted the number when acquiring permanent residenc y status. All permanent residents in the country have a unique and only one ID number.

Census results are used as a base line and the changes in the PR are used to calculate the current estimates between censuses. The ID number follows the person from birth (immigration) to death. It is widely used for all identification purposes and is therefore reliable. The PR was computerized in the late 60's of the previous century. Israel has closed border, with a limited number of formal crossings. All border passages are computerized. This is a great advantage to assure reliable registration of formal entries and departures to and from the country.

Israel's population size is defined as a "permanent population", which is a modification of the term "*de-facto* population" standard. In addition to all the people that are present in the country on the census date, people who are absent on the census date for less than 12 months and didn't notify the authorities of emigration, are included in the census population. The same applies for the current estimates based on the PR updates. People who return after a 12 month absence and stay in the country for less than 90 days are not included in the population.

Censuses, despite the efforts invested in the enumeration process, are still subject to under enumeration of "hard to enumerate groups" such as homeless and nomadic people. On the other hand, there are also duplications in the enumeration. Census results are corrected and imputed to adjust for the enumeration omissions . In Israel the census results were imputed and corrected only for the aggregated results. Therefore the annual size estimates were calculated on an aggregated, multi-dimensional, detailed matrix. The matrix values are calculated from the final corrected census results. Updates in the PR are not necessarily synchronized with census dates; information obtained during the enumeration process may not have been updated in the PR. A major deficiency of this system is that cells with few cases may become negative in time due to mistakes in enumeration information or duplication due to late updates in the PR. The current matrix in use for Israel contains about 7 billion cells, for 7 million people. By definition, most of the cells will be empty and many of them contain only one case per cell. Due to the difficulties encountered in previous years, of cells becoming negative, an alternative system is tested since the last census (1995) simultaneously. The alternative system is a "statistic PR" in which the records are updated individually based on the individual updates in the PR. The estimates are then calculated from the individual information and the final results are adjusted to conform to total size of the population using statistical models. The upcoming census (at the end of 2008) will be used to evaluate the quality of the two methods.

Two sub -population groups traditionally under reported in the Israeli censuses are: "Bedouins" - nomadic tribes who live in the South and a "religious"-ultra-Orthodox" group. The latter live in closed, heavily populated neighborhoods in urban areas. Both groups, who have very high fertility rates, are registered with high accuracy in the PR, since they are eligible for child allowances paid by the government. In the censuses they are traditionally misrepresented. Their census data was corrected either by imputation or corrections on the final aggregated results. The use of a "Statistic Register" should correct these flaws.

In the last 25 years the Bedouins situation has been changing. Economic incentives introduced by the government, to encourage the Bedouins to settle in designated localities and build permanent housing in an urban type of settlement, have significantly reduced the nomadic group. Over the last two decades the proportion of nomadic

Bedouins has decreased, and currently only less than half of them do not have permanent residences.

Another group that is difficult to estimate is people who stay in the country illegally. The size of this group is always a rough estimate, which is based on sources with differing reliability and whose quality changes over time. Currently three different models are used to estimate this group. Therefore the estimates have a wide range of variations. The final estimated figure is an arithmetic average of the three models.

2. Population Distribution

The concept of population distribution refers primarily to geographic distribution. Geographic distribution has several aspects that should be considered, such as size of area, type of area, political affiliation, administrative civil divisions and type of settlement. Geographic distribution is based on a variety of spatial definitions. Geographic sub-divisions, differ in their size, density, and number of inhibitors within countries. Sub-divisions within a country may also differ for different purposes, and by the authority responsible within the country. This is even more difficult in comparing geographic sub-divisions between countries. The spread of population over the land is seldom even within a country, and even over the land of a single locality. Population dispersion over land is a result of economic forces and historic and political changes. Statistical offices have very little, or no, influence over such decisions, but have to respond with population estimates for the sub-groups that were imposed by the country's civil authorities. Nesting the sub-divisions in a hierarchical mode is a useful way for a geographic sub-division within the country. This is not always possible for all purposes and different sub-divisions do exist. These subdivisions may overlap or not fully cover the entire land. Another issue that refers to geographic sub grouping is the type of locality; urban or rural and the subdivision within these two major categories. The definition of "rural" and "urban" differs among countries. The definition and change over time within countries, which are not always applied according to set standards.

2.1 Terms and classification related to population distribution:

i. Districts – (division, county) administrative or political sub-division within a country.
ii. Region- a geographic subdivision of a country
iii. Municipalities (localities) – a single locality with an official, legal status
iv. Quarter – A sub-area within a municipality, with some independent municipal autonomy
v. Statistical areas – a non-official subdivision of municipalities, with relative homogeneity regarding geographic, demographic, economic, social, historical and cultural characteristics.
vi. Block- the smallest contingent area surrounded by streets.
vii. Type of locality; urban – rural – type of residence setting of the locality.
viii. Metropolitan – agglomeration of localities with economic affiliation to a central city or town.
ix. Population density – number of people per sq. kilometer
x. Center of population over area – a geometric function representing the "center" point of the polygon determining area boundaries.

Many of the concepts, definitions and classifications related to geographic distribution of population are still under dispute. Very few have been agreed on or accepted, and they are set by countries and adjusted to national requirements. Very few standards were imposed by international organizations and subsequently applied by countries' statistical

offices. Some classifications and terms have changed over the years within countries to comply with administrative needs. For example, there are five different criteria that countries use to distinguish between a rural locality and an urban locality: (1) administrative area (2) population size (3) local governance (4) urban characteristics (5) predominant economic activity. In recent years more criteria have been suggested, such as density – concentration, average height of buildings and   water use per capita.

2.2. Data Sources – data on the distribution of population over a country are obtained through censuses as well as updates based on vital events and internal migration flows. Some countries base their internal migration and population distributions on local registers in municipalities. Local estimates are aggregated and verified against the total size. Some countries use central registers to calculate the population distribution over the land.

2.3 Population Distribution – Israel

Israel is among the countries with a very high population density. The spread of the population over the area is not even. Half of the country's land is a desert, which is mostly uninhibited. Over 72% of the population is concentrated in 14% of the land, resulting in very wide disparities in population distribution over the country.

The definition of urban/rural has been modified a few times over the 60 years of the country's existence, usually after censuses.

The definitions which were used up to the 1983 census were based on the main type of activity and main income source of the people living in the area. Localities with main income and activity from agriculture were defined as rural, and all the others were defined as urban. With agriculture becoming more industrialized, and more people living in agricultural localities changing their activity to non-agricultural occupation, the definitions for "rural" and "urban" were modified after the 1983 census. The new definition was based on size; localities with less than 2000 residents were defined rural, all the others were defined urban. A locality with 20,000 residents or more is defined a town. There is no formal definition for a city in Israel. Only two metropolitan areas were defined in Israel, one over 20 years ago – the Tel Aviv Metropolitan area and the Haifa Metropolitan area, in the last decade. The Tel Aviv Metropolitan area consists of more than 45% of Israel's population, portraying the density in the central "shore" and "off-shore" regions of Israel. Over 90% of the population of Israel lives in urban localities. In recent years, as agriculture has contributed less than 2% to the GDP, rural land has been attracting real estate developers and some rural localities have changed the land from agricultural to constructed land -residential and commercial construction. These trends have resulted in rural localities with agricultural activity to grow to over 2000 residents. The definition was examined by a special committee, which dealt with criteria to modify the definition, taking into account additional factors besides population size, such as: height of buildings, concentration of constructed land and use of water per capita. The final decision of the committee was to add a new category for localities with 2000-4999 inhabitants, which would be an "Urban-Rural" locality and would be grouped with the urban localities only when a dichotomy is presented. The main incentive for the decision was to maintain comparability of data for time series. Another decision taken was to preserve a secondary classification that would enable to disaggregate rural localities that have modified their status from rural to urban.

3. Population Structure

The structure of the population can be described through demographic, social, cultural, ethnic, and economic characteristics. The composition of the population is the subgroups into which the population is decomposed in the statistical sense of the word. Two types of decomposition are commonly used; one is the decomposition by individual characteristics (biological and acquired), and the other is decomposition by family typology.

3.1 Terms, definitions and standards of population structure

The most traditional and widely used Individual characteristics for decomposition are sex and age.

i. Sex -The personal characteristic of sex (currently referred to as gender, in the social sense rather than the biological sense) is the prime separator in demographic considerations. For any demographic analysis separation into male and female groups is essential. Any planning of services has different approaches to males and females. Demographic events such as fertility, mortality, migration and marriage are linked to sex. Sex is also closely linked to various social and economic indicators, such as education and labor force participation. Hence, most statistics present separate figures for each sex for most demographic, social and economic statistics. Sex definition presents no statistical problem and is, in most instances, easily obtained. Statistics, especially national statistics disregards ambiguity statuses of sex and treats the variable purely as a dichotomous situation. Abnormal sex ratio is a basic indicator for sex discrimination in a society.

ii. Age -Much of the rational for the need to present data by sex also applies to age. All core demographic and social variables, like natality, mortality, naptuality and migration are linked to age, and hence affect the statistics. Communities with different age compositions will differ in their totals, despite a similarity in behavior.

The age of an individual at a certain point in time is commonly defined as their age at their last birthday. In some cases age is defined as "age at the nearest birthday", or even "age at next birthday". The two latter definitions are rarely used by statistical offices, nor are they used in censuses. Nevertheless, in some cultures other definitions of age might exist and other calendars, different from the 365 days per year, might be used and affect the age reported by an individual. It is always recommended that the date of birth rater than age be reported. By using conversion of the reported date to the common calendar, age at each date can be calculated. Only if date of birth is unavailable, should age be obtained as a reliable proxy to the date of birth. Age is rarely presented in less than one year groups, except for infants aged less than one year, especially in the case of infant mortality statistics. Age in statistical tabulation is often aggregated into five year age groups or even broader groups, in some cases. For specific statistics which change sharply at a specific age (marital status, school enrolment, labor force participation) single year groups at certain ages might be required.

Reporting of age may be biased by the tendency of digit preferences, such as zero and five, resulting in heaping in the age distributions at the preferred ages. In some instances the exact age of an individual is unknown and proximity of the age is used.

iii. Race and ethnicity - compositions of race and ethnicity, differ in their classification and definition by the country practices, culture, history and tradition. The international recommendations and definitions are vague on the collection and presentation of these

data. In practice, each country subdivides its population according to its traditions and practical experience and planning needs.

iv. <u>Marital status</u> - an acquired feature, unlike the previous characteristics, which were biological features. This feature requires a formal relationship with another individual and a formal institution. The definition is therefore more complex. The institutions and legalities differ between nations and within subgroups in a country with cultural autonomy. Information on marital status is also reflected in the family formation and typology. Recent changes, such as postponing marriage, and several new cohabitation statuses, introduce new definitions of family typology.

Since marriage in most countries is performed by an authorized institution, the legalities associated with definitions of marital status are the result of the legislative procedures in each country. Eligibility to form or break the ties is embedded in the legal system. <u>Same sex marriages</u> are a contemporary example of different approaches in different countries to the legality and registration of such marriages and subsequently changing their formal recognized marital status.

v. <u>Families and households</u> -A household is the basic socio-economic unit. For most people the two concepts are identical. Some people live in settings which are not a single family household. Such are multi-generational families who share a dwelling, and "collective households" (military barracks, prisons, college dormitories). There are formal recommended definitions which are proposed for censuses. The UN recommended definitions for censuses are:
<u>Household</u> – One or several persons living together, who make common provision for their own food or other essentials of living, regardless of family relations between the persons comprising the household.
<u>Family-</u> The family within a household is defined as those members of the household who are related to a specified degree through blood, adoption or marriage.

3.2 Data sources for population structure

The two major data sources for all types of decomposition of the population into subgroups by demographic characteristics are censuses and updates in the PR. Sex, age and marital status are registered in the PR. In some countries ethnic and racial characteristics are not registered.

Data on households and families is collected mainly through censuses and household current surveys between censuses. Information obtained from administrative sources usually relates to "core families" who share living quarters. Household definitions may have some minor variation but are generally common.

Household typology and family typology are based on national experience and specific societal needs within countries. The aim of statistical offices is to collect data that is as detailed as possible, to enable variations in typology according to specific demands by users.

3.3 Israel's population structure

Sex is collected in censuses and is also fully registered in the PR. Meaning, there is no missing data on sex in the PR. Nevertheless, erroneous registration of sex in the PR may occur, as so does misreporting and mistakes in data entry in censuses. But as they happen randomly they do not affect the structure and will usually be detected when

other events do not comply with the registered sex such as giving birth by a person registered as a male. As these are very rare events, we assume that this type of mistake is scarce and has no effect on the primary distribution by sex. This is not applicable to age distribution.

A substantial group of immigrants, who arrived in Israel shortly after its independence, arrived with no formal documents such as passports or birth certificates. Their age was, declared to the immigration authorities. This, declaration of their age without the exact date of birth, resulted in a proximal calculation of year of birth and after some years a default exact date was assigned to each person with missing information on date of birth. As ages were declared, there was a preference for certain digits such as 0 and 5 and accordingly the year of birth was heaped. This resulted in a distorted age pyramid with heaping, which moved along ages over the years. As these mainly occurred, during the first few years after independence, the heaping is currently mainly observed in the upper age groups. The same applies to the population that lived in the country before its independence the ages were obtained during the first census. People who did not have formal birth certificates declared their date of birth or if not known their year of birth was calculated based on the reported age, causing similar heaping. Age heaping is mainly affecting the mortality statistics, by distorting the mortality function. Israel has a relatively young population (Median age 30 years), therefore the irregularity is mainly displayed in the fluctuations of the mortality rates at the older ages. To overcome this fluctuation a special mathematical smoothing function was developed, which is applied to the mortality rates.

Ethnicity and race are not collected in Israel, neither on census questionnaires nor in any other survey. A proxy for ethnicity is the combination of country of origin (country of birth or for Israeli born, father's country of birth) and religion. Religions are grouped into the two main national groups, Arabs and Jews. The main obstacle encountered is for country of birth; the term refers to country by current borders as opposed to borders at the time of birth. Merger of countries is handled automatically while split of countries is almost impossible to adjust, since the exact place of birth (name of locality) is not registered and often also localities names change after separation of countries.

Families and households in Israel –
Data is obtained from censuses and the labor force survey. An attempt has been made recently for the upcoming integrated census, which will be held at the end of 2008, to formulate families from the PR and some supporting sources such as school enrolment files. The attempt has been only partially successful so far. The pilot census held in 2006 revealed that only in about 82% of the cases did the algorithm formulate the correct family. The algorithm was improved and retested and the results were somewhat improved and reached approximately 85%. The anticipation is that with the census results, the algorithm can be further improved; however, it will probably not exceed 90% success.

4. Summary

Population size, geographic distribution and the population structure of are the basic data needed for any planning or analysis of the changes that occurred or are expected for the society. Demographic projections, used for planning services and determining the allocation of resources is based on the society's composition and size.

All society performance indicators are adjusted to the population size and composition. Demographic characteristics of the society are the fundamental building blocks for any

evidence based decisions. The main source for the data on the size, distribution and structure of a country's population are obtained from censuses and updated mainly from administrative sources, in developed countries, and rarely from data obtained in sample surveys. As countries move toward non-traditional censuses, most of the census data is originates from administrative sources as well. In future with censuses becoming an operational and budgetary heavy burden on one hand and availability of rich administrative computerized data sources there will be a shift to rely more on administrative sources for population size, distribution and structure.

5. References

Principles and Recommendation for the Population and Housing Censuses, Rev. 2, United Nations, Economic and Social Affairs, New York 2008.

Recommendation for the 2010 Population Censuses, Conference of European Statisticians, Economic Commission for Europe, United Nations, New York and Geneva 2006.

The Methods and Material of Demography, H.S. Shryock, J.S. Siegel and associates, US department of Commerce, Bureau of Census, 1971.

Statistical Abstract, Central Bureau of Statistics, Israel 2007.