Methodology for Population Estimation

Types of Population Estimates

Population estimates can be divided into several types on the basis of their time reference and how they are derived. The most common types of estimates are as follows:

Intercensal estimates relate to a date intermediate to two censuses and take the results of these censuses into account. *Postcensal estimates* relate to a past or current date following a census and take the most recent and possibly earlier censuses into account, but not later censuses. *Projections* relate to dates following the last census, usually future dates, for which no current reports are available.

Both postcensal estimates and projections can be regarded as extrapolations and intercensal estimates as interpolations.

General Methodology for Producing Estimates

Several methods for making estimates of the total population of a country are available, each applicable under different circumstances. They include sample surveys, population registers, complete or partial registrations, the component method, mathematical procedures and various combinations of these methods.

The most basic methods of estimating population belong to a group called component methods. This method takes account of the components of population change. In such procedures each component is estimated separately and is added to or subtracted from a base population to arrive at the population as of some other date.

The basic formula for the component method, also referred to as the balancing equation, is:

$P_t - P_o = B - D + I - E$

Where

P_t= the population at the end of the period

P_o= the population at the beginning of the period

B= births

D= deaths

I= immigration

E= emigration

This is the most preferred method as it utilizes the census count as the base and direct data on postcensal changes from registration systems and population registers.

The most important factor determining the choice of method to be used in preparing a given population estimate is the type and quality of data available for the purpose. The component method may be used in a more limited way, in combination with mathematical or other procedures not using components. This variation becomes applicable where reliable death statistics by age are missing and the volume of net immigration is not significant. Allowance also needs to be made for under registration of births.

Revision of Intercensal Estimates

Intercensal Estimates and the Error of Closure

Intercensal estimates are produced following each census in order to reconcile existing postcensal estimates with the census counts, thus assuring the internal consistency of the estimation system.

The national census conducted at regular intervals provides a range of demographic data on the population. One of the fundamental uses of the census is to provide the base population for the estimation of annual population size, distribution and composition in the postcensal period. The extent to which the census at the later date differs from the estimates based on the earlier date is dependent on the level of completeness of the vital registration system as well as the census enumeration. The difference between the expected count from the census and the actual count is called the *error of closure*.

This error would be reflected in the balancing equation as:

$$P_t - P_o = B - D + I - E + e$$

The error of closure comes basically from two sources; differences in the amount of undercoverage/ overcoverage in successive censuses and errors in the components of population change over the intercensal period. On the basis of additional knowledge about the accuracy of the various terms, it is possible to decide whether *e* can be attributed as a measurement error to a particular term in the equation.

In developing intercensal estimates the assumption can be made that the total *error of closure is a linear function of the time elapsed since the first census.* Hence the correction for each year is derived by interpolating linearly between zero at the earlier census date and the error of closure assigned to the later census date. These interpolated corrections may then be combined with the original postcensal population estimates.

The error of closure may also be distributed over the intercensal period in proportion to the postcensal population, total population change or one or more of the components of change.

Where mathematical interpolation is employed in the calculation of intercensal estimates, the error of closure is taken care of in the principal calculation and a separate adjustment for this error is unnecessary.

Methodology

The methodology utilized for the preparation of the estimates for 2001–2011 was a combination of a limited component method and mathematical procedures. The Population Analysis Spreadsheet (PAS) system developed by the International Programs Center Population Division of the US Bureau of the Census was used. The PAS system is a set of Lotus 1-2-3 and Excel spreadsheets which were created to help in the analysis of demographic data for use primarily by developing countries with limited information. The choice of method was dictated by deficiencies in the vital statistics and the system for estimating migration.

The spreadsheet used was AGEINT which forms part of the set developed for analysing the age and sex composition of populations.

AGEINT- interpolates between two age/sex distributions. The linear or exponential interpolation performed by this spreadsheet is made using the population in the same age groups at the two dates. *The procedure assumes that the average annual change in each population age group has been constant during the intercensal period.*

The linear interpolation is performed as follows:

$$P_s = kP_i + (1-K)P_j$$

Where:

P_i , P_s, and P_j represent the population of each age group at dates i, s and j;

s is the date for which the interpolation is desired (it must be chronologically between dates i and j); and

k is a constant for all age groups, calculated as: k = (j-s)/ (j-i)

Steps in the Calculation

Step 1 – The error of closure was distributed among the adjusted components of growth (Table 4).

Step 2– The total population by sex was derived for each year from 2001, on the basis of the components.

Step 3 – Using the 2001 and 2011 census data by sex and 5 year age groups (under 1, 1-4, 5-9….80+) as the base population, AGE INT was used to derive (interpolate) end of year population by age and sex for each year for the period 2001-2011 (census).

Postcensal Estimation

The methodology outlined below has been used to produce estimates of the population and components for the period since the census of 2011. *Postcensal age estimates are provisional and subject to change when a detailed evaluation of the 2011 census age distribution is done. There is some evidence of age heaping at specific ages. Corrective measures would include 'smoothing' to correct the irregularities for single years of age only. The five-year age groups will not be affected.*

From Census to the end of 2011

Step 1- Estimating the Total Population

The first step is to derive the total population at the end of 2011 by sex. This involves adding the births, subtracting the deaths and migration for each sex for the remaining portion of 2011 to the estimates at the census. (The period between April 4 2011(day after census day) and December 31, 2011 was 271/366 days or .7404 of the vital events for 2011.)

Step 2- Estimating the Population by Age

(i) The five-year age distribution of deaths and migration are broken down into single year estimates using Beers Ordinary Interpolation Coefficients. For migration, the multipliers are used for all age groups. For deaths the multipliers are used for ages five years and over only. For deaths under five years, the numbers as recorded by the RGD for single years 0, 1, 2, 3, 4 were used.

(ii) Estimates for single years of age are then derived as follows. The estimates for the census, beginning with age 0 are used as the base. It is assumed that .7404 of each age (including age 0) at the census will attain a birthday and move to the next age, by the end of 2011. Age 0 at the end of 2011 will comprise .7404 of all the births for the year and .2596 of the of the census population at age 0.

(iii) The final step is to subtract the decrements by single years from each age.

Step 3- Adjusting the totals to ensure consistency

The total of the estimates by age and sex derived after the births, deaths and migration are applied to the population at the end of the previous year and are then adjusted to match the total derived at Step 1.

For 2012

Step 1- Estimating the Total Population

For end year 2012 the first step is to derive the total population by sex. This involves adding the births, subtracting the deaths and migration for the full year 2012 for each sex to the estimates at the end of the previous year;

Step 2- Estimating the Population by Age

(i) The estimates for the end of year 2011, beginning with age 0 are used as the base for the beginning of year 2012. Each age is aged one full year. All persons at age 0 at the end of 2011 become age 1 and all the births for 2012 are at age 0.

(ii) The five-year distribution of decrements (deaths and migration) for 2012 are broken down into single years using the Beers Coefficients as described above. For deaths, the multipliers are used for ages five years and over only. For deaths under five years, the numbers as recorded by the RGD for single years 0, 1, 2, 3, 4 were used.

(iii) The decrements for each year of age have to be separated into those which occurred before a birthday and those which occurred after a birthday. For example, some of the persons who started the year at age 20 years will die or migrate at age 20 and some will die or migrate after attaining age 21 years.

The separation factors used for deaths are as follows:

Age	f	(1-f)
0	.8025	.1975
1 and over	0.50	0.50

The separation factor for age 0 (f = .8025) indicates that, in any given year, about 80 per cent of infant deaths occur to infants who were born in the estimation year. The remaining 20 per cent of deaths (1-f) occur to infants who were born in the previous year.

For migration the separation factor of .50 is used for all ages.

Step 3 - Adjusting the totals to ensure consistency

The total of the estimates by age and sex derived after the births, deaths and migration are applied to the population at the end of the previous year are then adjusted to match the total derived at Step 1.

Data Limitations

One essential aspect of a good system of vital statistics is the provision for critical evaluations. The United Nations (2001) recommends that evaluation or performance monitoring should be part of the operation of the civil registration and vital statistics systems. The elements of the evaluation programmes will vary in detail according to the degree of development of the system, but should include at some stage, intensive studies of the degree of completeness of registration and of statistical reporting of events, with a view to evaluating the quantitative accuracy of vital statistics. Measuring the degree by which registration is deficient, is a difficult problem, but one which is of great importance to the improvement of vital statistics. No test or measure can be considered definitive of course because of the potential "incompleteness" of any source of information used as a check.

In Jamaica, a number of national agencies routinely collect information on births and deaths and the data derived from these sources provide the basis for the assessment and adjustment of vital statistics produced by the Civil Registration System, managed by the RGD. The civil registration system is the repository for information emanating from the Health system: public and private institutions, from homes, medical practitioners, public and private, and from the various arms of the justice system.

Coverage of Births and Deaths

No regular schedule for the conduct of tests for coverage of birth and death registration is in place and over the years they have been done on an 'ad hoc' basis.

There are indications that birth registrations have seen considerable improvement, especially since the inception, in January 2007, of a programme for bedside registration by the RGD. Through this programme the agency places a registration officer or a Local District Registrar in hospitals across the island to facilitate immediate and more effective registration. Coverage is estimated at over 95 per cent. Statistics on delays in registration for births occurring in each year between 2001 and 2010 show that on average about 98% of the events were registered within twelve months of the birth.

The coverage of deaths has lagged and continues to be problematic. The areas of particular concern over the years have been infant deaths, sudden and violent cases and maternal deaths. An Evaluation of the Quality and Completeness of Registration for deaths occurring in 2008 (McCaw-Binns et al 2012), found that only 76% of the deaths had been registered. Of the unregistered, nine per cent were in a RGD database of unregistered events, based on police reports of deaths due to violence, suicide and land transport accidents. The remaining 15 per cent which were unaccounted for were mainly sudden deaths due to natural causes and accidents such as fire and drowning. The majority of unregistered deaths remain Coroners' cases.

International Migration

The intercensal estimates for migration are derived as the residual difference between the total annual population increase and the natural increase (the difference between births and deaths). For the period following the census, the estimate of migration is derived from the statistics on legal migration to the receiving countries for which data are available, the United States of America, Canada and the United Kingdom. The age-sex distribution of migrants is assumed to conform to the distribution of persons migrating to the United States of America, the only destination for the majority of migrants and the destination for which the details required are available.

References

McCaw-Binns A and Holder Y. February 2012. *Quality and Completeness of Vital Registration in Jamaica, 2008: Comprehensive Review of the Completeness, Coverage and Quality of Death and Foetal Death Registration, Coding and Classification of all Deaths including Foetal Deaths and Maternal Deaths.* Report to the Planning Institute of Jamaica. Kingston.

Calculation of Mortality Rates and Life Tables

2The life tables presented in Tables 48-50 have been constructed using the Population Analysis Spreadsheet (PAS) system developed by the International Programs Center Population Division of the US Bureau of the Census. The PAS system is a set of Lotus 1-2-3 and Excel spreadsheets which were created to help in the analysis of demographic data for use primarily by developing countries with limited information. The spreadsheet used was LTPOPDTH which forms part of the set developed for constructing life tables The LTPOPDTH spreadsheet constructs and smoothes a life table.

3The data required for constructing the life table using this spreadsheet are as follows:

1. populations for ages under 1 year, 1-4 years, and by 5-year age groups thereafter; deaths for the same age groups as the population.
2. populations for ages under 1 year, 1-4 years, and by 5-year age groups thereafter; deaths for the same age groups as the population.

Optional data are as follows:

1. Infant mortality rate;
2. Separation factors for ages under 1 year and 1-4 years.

If no infant mortality rate is provided, the program calculates one (the q_o value) based on the population and number of deaths under 1 year.

If no separation factors are available, they are automatically calculated using the Coale-Demeny formulas. For other ages the separation factors are assumed to be 2.5.

A logarithmic smoothing process is applied to the ${}_nm_x$ function *of the unsmoothed life table*. Once the ${}_nm_x$ values are smoothed, they are proportionally adjusted to reproduce the total number of deaths in the smoothed ages (starting with the age group 15-19 years).

The closing of the life table is calculated based on the open-ended central death rate.

Constructing Tables for Both Sexes

The LTPOPDTH spreadsheet allows for the calculation of life tables for male female only. Tables for both sexes combined were by weighting the q_x'$_s$ from the male and female tables, in accordance with the distribution of the population by sex at each age. A new table is then constructed from the combined q_x'$_s$ and an assumed radix of 100,000.

The formula used for this computation was as follows:

$$\{((Bm/Fm)^* d^{m}{}_{x})_{+} d^{f}{}_{x}\} / \{((Bm/Fm)^* l^{m}{}_{x})_{+} l^{f}{}_{x}\}$$

Where Bm = male births

Fm = female births

$d^{m}{}_{x}$ = male deaths (from ${}_nd_x$)

$d^{f}{}_{x}$ = female deaths (from ${}_nd_x$)

$l^{m}{}_{x}$ = male survivors to age x (from l_x)

$l^{f}{}_{x}$ = male survivors to age x (from l_x)

Explanation of Separation Factors

A separation factor is defined as the average time lived during an age interval by persons who die between the beginning and the end of the interval. For example, persons dying between ages 45 and 46 years live on average one half year between their 45th birthday and their time of death before their 46th birthday. Separation factors are usually calculated for five-year age groups. In this case, persons dying between ages 45 and 50 years live on the average about 2.5 years: some of them die immediately after celebrating their 45th birthday, while others die just before reaching their 50th birthday.

Although it is accepted that separation factors for ages five years and over are half of the age interval (when the age interval is not greater than five years) this assumption is not affected for ages under five years. Separation factors for ages under one year and from one to four years are smaller than half the age interval. This is due to the fact that mortality is very high during the first day of life and declines rapidly during the first year and up to the fifth year. This means that the number of deaths is greater at the beginning of the age interval than at the end, and the time lived by those dying during an age interval is less than half of the age interval.

Coale Demeny West Separation Factors

If Infant Mortality Rate < .100

Age	Males	Females	Both Sexes
0	.0425	.05	.0426
1	1.653	1.524	1.6518

K_o separation factors for age 0

K_1 separation factors for ages 1–4

${}_1L_o = k_o * l_o + (1-k_o)\ l_1$

${}_4L_1 = k_1 * l_1 + (4-k_1)\ l_5$

${}_5\ L\ {}_x = 5/2 * (l_x + l_x + 5)$

$L_{80} = 3.725\ (l_{80})\ .0000625\ (l_{80})2$