Economic Commission for Europe

Conference of European Statisticians

Joint Eurostat/UNECE Work Session on Demographic Projections
Geneva, 18-20 April 2016
Item 2 of the provisional agenda
Methodology

Preparing population projections in developing countries

Instituto Nacional de Estadistica y Geografia. Mexico[1]

Summary

Mexico has a population of 119 million and 1.96 million of kilometres of land. This big numbers produce some problems in basic information for population, then projections have different limitations which require a careful analysis called demographic conciliation. This stage is crucial and it determines the base population, and levels, trend and structures of main demographic variables, which allow the elaboration of scenarios for future developments. This document show a summary of the procedure followed by the preparation of the demographic conciliation for the latest Mexican projections, highlighting the main problems of methodological character what it had, as well as those who are considered, in the author's opinion, the main challenges to resolve in order to improve the population projections. Conciliation and projections were made using the components method. The exercise was done by each State (Mexico has 32 States) and was made bottom to top, i.e. was done by Federal State, then the national projection was obtained by sum. The conciliation use data for the population census between 1990 and 2010. Data sources were the five census of the period (1990, 1995, 2000, 2005 and 2010), the statistics of births and deaths registered in the system of vital statistics, four national surveys of demographic dynamics (1992, 1997, 2006 and 2009) as well as international migration estimates it is derived from the United States most adequate census samples and surveys.

---

[1] [Prepared by Garcia Víctor, Planning Director.

I.   Introduction

1.   In Latin American countries, demographic data from census, administrative registers and surveys, have different limitations which require a careful analysis in order to be useful in studies.

2.   In the field of population projections, while in developed countries the population base and levels, trends and structures of demographic variables are information that can be taken almost directly from the available data sources, the scope and limitations of this information in the Latin American countries require a carefully analysis and assessment

3.   One of the most useful tool used for this analysis and assessment, at least in Latin America, is the so called demographic conciliation[2], which in addition to the analysis and evaluation, allows to set the levels, trends and structures of main demographic variables in the recent past, and a population estimates. This stage is crucial in preparing the projections raw data, throughout it, the experts determines the bulk of demographic needs, like base population, levels, trends and structures of main demographic variables which allow to elaborate scenarios for future developments.

4.   This conciliation is made for a period in the recent past, limited by two population censuses and includes a key step that is known as selection of initial values, process that is completely arbitrary in its conventional form.

5.   It is important to mention that this process takes specific characteristics in each case, which depends on the information available in the country and on the decisions that takes the group responsible for its preparation.

6.   Among the first, availability of at least two censuses of population, administrative records or demographic surveys that allow to estimate fertility, mortality and migration indicators is essential.

7.   Also is required that a responsible group take some decisions between highlights the period which will be analysed and the geographical and age disaggregation level.

8.   This document is a summary of the procedure followed in the preparation of the demographic conciliation which served as the basis for the latest projections of Mexico, highlighting the main problems of methodological character which has, as well as those who are considered, in the author's opinion, as the main challenges to resolve in order to improve conciliation and population projections.

II.   Demographic conciliation

9.   The overall objective of the demographic conciliation is to build throughout components method a consistent set of estimates of population demographic dynamics, for a period in the past, based on all basic information available.

---

[2] Rincón, Manuel. Conciliación Censal y Determinación de la Población base. Métodos para Proyecciones Demográficas. Centro Latinoamericano de Demografía. Serie E, No. 1003, San José, Costa Rica, 1984.

10.   The demographic conciliation can be defined as an indirect method to establish the volume and structure of the initial or base population in order to accomplish population projections, reconstructing the demographic dynamics of the recent past. It is essential to update trends and levels estimates of each demographic phenomenon.

11.   Traditional census conciliation exercises consist on simultaneous location of the population data by age and sex of two or more censuses at one time point, using forward and backward demographic projections. Once located the various census populations at a given time, the expert chooses, for each age and sex, the most appropriate values.

12.   During this process, in its classic form, intervenes, in a definitive way, the judgement of the person responsible for its production when he or she selects the named "initial values". This selection, although considers some basic principles, is subjective and is, in the opinion of the author, his biggest weakness.

13.   Demographic conciliation starts by choosing a period in the recent past, delimited by two population. This period can include more intermediate censuses.

14.   Once defined the period, all the demographic information available for it, is collected and analysed. This analysis concludes with estimates of trends, levels and structures of each one of demographic variables.

15.   A specific date in the year is selected and each population census is moved to that date. For the selection of this date is important to consider the specific reference date of each census, a common practice is to select the middle year point.

16.   Using components method and estimates of demographic variables they are projected to the same year. Although the recommendation is the year of the census which is considered more reliable, usually is preferred the most recent. With this, different estimates of the population for each sex and age are obtained, as many as censuses are included in the selected period.

17.   At this point the selection of the initial values is made, according to a series of criteria that requires extensive knowledge of the characteristics of the demographic data and population dynamics, but that does not remove its subjective character. Their results are crucial because these selected values usually are taken as the base population for the new projections.

III   Population projections in Mexico

18.   In Mexico, the General Secretariat of The National Population Council (CONAPO) is the institution responsible for elaborating population projections. This institution is in the Ministry of the interior of the Mexican Government.

19.   From 1970 until today, whenever rises a new census in Mexico, CONAPO elaborates new projections. These projections are always prepared at national and state level.

20.   The last population census was raised in 2010[3] and as a consequence, CONAPO proceeded to develop new population projections whose first phase was the preparation of the demographic conciliation.

21.   The version published in 2013, included the following projections:

   a. National population: yearly, by simple age and sex, from 2010 to 2050
   b. Population by State: yearly, by simple age and sex, from 2010 to 2030
   c. Population by municipality, yearly, by five large groups of age and sex, from 2010 to 2030
   d. Population for localities over 2500 inhabitants: Total, from 2010 to 2030

22.   For the period 2010-2030 it was used a bottom-up process, i.e. first, State projections by components method were made, then national projection was built by sum. From 2030 to 2050, national projection was built by components method directly.

23.   Projections by municipalities were prepared by a cohort's relation method and from localities were used growth rates.

IV   Demographic conciliation in Mexico

24.   From 1990 to 2005 based versions, demographic conciliation Mexico was made for a five years period and top to bottom.

25.   Recent 2010 demographic conciliation 2010 was prepared bottom to top, i.e., was made for each one of 32 States and the country data was calculated by addition. The selected period was 1990-2010, twenty years, which includes 5 census projects, one every five years. As it is necessary to consider the Interstate migration, conciliation of the 32 States must be simultaneously elaborated.

26.   The general procedure consisted in projecting with estimated demographic variables, the population from previous census (1990, 1995, 2000 and 2005) to a single date (July 1, 2010), this leads to a set of five population estimates for each sex and age on that date and consequently requires selecting a single estimate. Finally, estimates of demographic variables, including the population were integrated into a backward projection that covers the period 1990-2010

27.   First, with all available sources of information, were estimated annual fertility rates and mortality tables. International migration were estimated in volume, i.e. the annual number of international immigrants and emigrants by sex and age. For internal migration by State, were used annual emigration rates by age and sex for each State. Finally, the distribution of emigrants of each State of destiny State was estimated.

28.   Age data has some problems. The first is age heaping, this was adjusted using the smoothing procedure proposed by Gray4. The second is a significant number of households in which census cannot be taken (448,195 in 2010). Using

---

[3] In 2015 the INEGI rose a new national survey. Using its results, CONAPO will prepare new projections

[4] Gray, Alan. "The Missing Ages: Adjusting for Digit Preference", in Asian and Pacific Population Forum. Vol. 1 No. 2, Feb1987, East West Population Institute. East West Center.

the information corresponding from homes really surveyed, was estimated the inhabitants number and age distribution in these homes without information.

29. This "new" and "total" census population was moved to July 1 of the respective year with growth rates by age and sex.

30. These populations were projected by the components method to July 1, 2010, with the equations presented below.

$$N_{x+1}^{t+1} = N_x^t \cdot S_x^t + IE_{x+1}^{t+1} - EE_{x+1}^{t+1} + II_{x+1}^{t+1} - EI_{x+1}^{t+1}$$

N_x^t Population of age x in year t

S_x^t Survival relationship for population of age x in year t

IE_x^t International immigrants of age x in year t

EE_x^t International emigrants of age x in year t

II_x^t Internal immigrants of age x in year t

EI_x^t Internal emigrants of age x in year t

International migration was estimated in number, then it was applied directly. For internal migration, were estimated:

r_x^t Emigration rates for age and sex

$s_{i,j}^t$ Proportion of emigrants from the State i having as destination the State j

Then, with emigration rates and population, out-migrants by age and sex for each state are calculated

$$EI_{x+1}^{t+1} = \left\{ [N_x^t \cdot S_x^t] - EE_{x+1}^{t+1} \right\} \cdot r_{x+1}^{t+1}$$

Then out-migrants of each state are distributed among all of them

$$II_{x+1}^{t+1} = \sum_{i=1}^{32} EI_{x+1}^{t+1} \cdot s_{i,j}^{t+1}$$

31. By this way, there are five estimates of the population by sex for every age in July 1st, 2010, one by each one of the censuses.

32.   These five estimates should be reduced to one for each age and sex. Usually, this is done by making decisions considering the characteristics of this type of information. The team responsible makes this selection analyzing case-by-case.

33.   In this point the problem is the subjective nature of the procedure. It can be said that if several experts apply this process independently, they would hardly get the same results. In last exercises made in Mexico, instead of analyzing case-by-case basis, has decided to establish general rules for the selection of these initial values.

34.   Until 2005 census based exercises, the decision was in general, take the maximum value, under the following considerations[5]:

    a.   The estimates of demographic variables are correct.

    b.   Census present lack of coverage but not over coverage

35.   For the current version, it was decided to make a demographic conciliation for the period 1990-2010. This results in a longer series of estimates of demographic variables and is considered to have better elements for the elaboration of its future evolution hypothesis as well as for the determination of the base population.

36.   It was detected that in more than half of the cases, the maximum value corresponds to the data from the 2010 census. This leads to the conclusion that this is the census with best coverage.

37.   In terms of the selection of initial values, it was determined to make it with specific criteria, in order to eliminate the subjectivity of the process. to do this it was considered:

    a.   Estimates derived from the previous censuses (1990, 1995, 2000 and 2005) there are affected by Census coverage and estimates of mortality and migration, both internal and international, as well as estimates of fertility for the early ages.

    b.   Estimates derived from the Census of 2010, depend only on their own census coverage.

38.   Then, criteria used was:

    a.   When the maximum corresponds to data from 2010 census, chose it.

    b.   If maximum corresponds to another census, take an average between this one and the 2010 data.

39.   The assumption is that the differences are due, on the one hand to the census coverage differences, and on the other, to errors in the estimates of demographic variables, primarily migration. Take the average it is equivalent to assume that half of the error is due to each of these two factors.

40.   In terms of population, a census under coverage of 1.58% was estimated. This figure is slightly higher than the estimated by post enumeration survey,

---

[5] it should be noted that in these cases, we worked with a five-year period and therefore we had only available two estimates for each age

carried off by INEGI (1.30%). By State, this coverage varies between 0.69% and 3.19%.

41.   Similarly to other exercises, including the previous made for Mexico, this last close the exercise and the selected values are taken as the population base. In this exercise it was decide to make a backward projection until 1990. This allowed to assess the impact of the selected values in the coverage level of previous censuses.

42.   In terms of the evaluation of the estimates of demographic variables, this was only done for fertility. With backward projection annual births were estimated, then, initial estimates of fertility rates were reset. Changes in Total Fertility Rates (TFR) were relatively small (the differences with initial values vary from -0.26 to 0.29), but consistency in the demographic dynamics of the States was improved

V.   Discussion

43.   Demographic conciliation exercises performed in Mexico in the last twenty years have not included the arbitrary selection of initial values, so the subjectivity of the specialist does not intervene, Although persists some arbitrariness in the selection of initial values.

44.   In Mexico, in previous exercises, under the assumption that the estimates of demographic variables are correct and that censuses have errors of under coverage and not over coverage, it was selected the maximum value.

45.   In 2010, initial values were calculated averaging the maximum value with the corresponding to the last census[6]. In this case, assumptions are:

> a.   Last census is the best in coverage. This assumption is based on the fact that in more than half of the cases, this census value corresponds to the maximum among the five estimates.
>
> b.   The latest census estimates depend only of census coverage, while the other four are affected by limitations in estimates of fertility (for early ages), mortality and migration
>
> c.   Differences between the value of the last census and the maximum is caused, with an equivalent weight, by both the under coverage of the last census and errors in estimation of demographic variables.

46.   A major limitation in the selection of the initial values in the exercise of 2010, is that for every age, there are five available estimates and for the selection of the initial values are only considered two, i.e., only a part of the information is used.

47.   It is important to consider that the five estimates have different characteristics. Figures that come from the 2010 Census nearly correspond to the census data. Only minor adjustments were made, as the smoothing to eliminate digital preference (procedure that does not alter the size of the population), and the

---

[6] Obviously, if the maximum value corresponds to the last census, the selected value corresponds to this census.

transfer to July 1, that was just 20 days (the census reference date is June 12, 2010).

48.   Figures from other censuses, which in addition to the same treatments that it did to the 2010 census, were projected for until 20 years and incorporate estimates of mortality, internal and international migration.

49.   However, for each age and sex, are available 5 estimates and only two are used. Even though one of the two (of the last census), has the elements to consider that it is the best quality, the second used is only selected for being the largest.

50.   See, for example, graph 1. In these ages. There are three estimates that are not considered, these correspond to values lower than 2010, only the 1995 census data are higher than 2010 and the only one which suggest under coverage in 2010. 1990, 2000 and 2005 suggest over coverage.

Graph 1
Tabasco. Females. 30-34 years
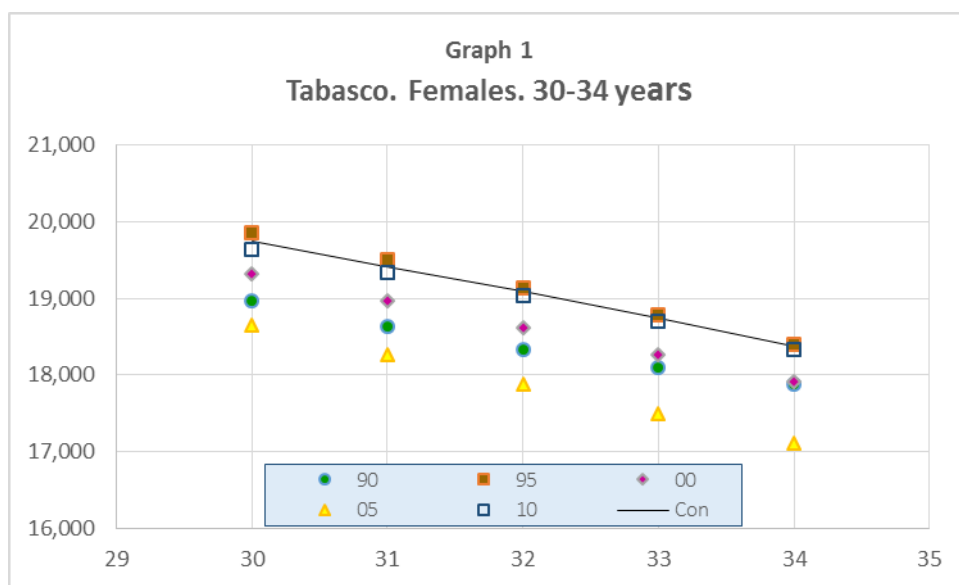
51.   This kind of considerations are used to select initial values in the traditional procedure, but as already mentioned, the decisions are subjective.

52.   The question is, is it possible to develop an algorithm to select the initial values, using all available information? We are referring not only to data (in this case five), also their distribution.

53.   On the other hand, one of the objectives of the demographic conciliation is t**o make consistent and compatible population by sex and age, estimates of mortality, fertility and migration,** within a specific time in the past, based on all available demographic information.

54.   However, in demographic conciliation, usually, estimates of mortality, fertility and migration, are only used in order to calculate the population data by sex and age, those are used as initial values for projections. The demographic variables rarely are revised.

55.   In given the limitations of data sources and its impact on the estimation of the demographic variables of the recent past, in the 2010 exercise, fertility estimates were adjusted, but estimations of mortality and migration were not evaluated.

56.   The Mexico's exercise in 2010, TFR estimation was adjusted using demographic conciliation, but migration and mortality were not revised.

VI.   An alternative algorithm

57.   Looking for alternatives in order to determine initial values without subjective criteria and using the five data for each age-sex group, we combine two simple statistical tools, which were applied to selected states:

58.   The four selected states were

    a.   Chihuahua. This state has the major under coverage estimated.

    b.   Queretaro. Whit the lowest under coverage estimated.

    c.   Tabasco. The difference between post enumeration survey and conciliation was the lowest.

    d.   Sinaloa. The difference between post enumeration survey and conciliation was the largest.

59.   First, using Box Plot, "outliers" were identified. These points were eliminated.

60.   With the remaining data, using Bootstrap technique, and assuming a normal distribution, confidence interval limits for the estimation of the population of each sex and age group were estimated.

61.   Finally, maximum value between the upper limit of the confidence interval and 2010 census data, was chosen.

62.   In table 1, total under coverage estimates from demographic conciliation, Bootstrap and pos-enumeration survey are presented.

63.   In general, selected values are similar, but Bootstrap consider the five estimates, conciliation only one or two (when data from 2010 census is the maximum, it is the selected value).

64.   For a more detailed comparison between both exercises, in each State was selected the ten-year age group in which are presented the biggest differences between both stimations These ten-year groups and associated differences are presented in table 2.

65.   In graphs 2-9, initial data are compared with original conciliation and Bootstrap selection. Differences are explained by the behavior of points not considered in conciliation, for example, in Chihuahua males, practically all the values estimated with the previous censuses are superior to 2010. In conciliation was considered only one of them (the maximum), while with B-S all are included

66.   It can be considered that B-S estimates improve the corresponding to conciliation, the argument is which incorporate more information, however, further analyses are required

VII.   Pending tasks

67.   It is considered that the proposed exercise complies with the objective, propose an alternate algorithm that considers all available information. However, it is necessary to assess its results more broadly, for example, analyzing their impact on estimates of census coverage in previous events.

68.   Although the last conciliation prepared for Mexico included an adjustment of fertility rates, no attempt was done to assess, and if necessary modify, the initial hypothesis of internal and international migration.

69.   One possibility that can be explored to assess this variable would be, for example, calculating the difference between a census population and the projected from the previous census, contrast age structure of these differences with the structure by age of immigration or emigration, and evaluate the possibility of making adjustments in the initial estimates.

70.   If conciliation is made, as in the case of Mexico, by regions, this would lead to the comparison of the results of a State with the major point of origin or destination for the migration. Of course that does not seem easy, but given current developments for processing, it is deemed feasible to explore this and other alternatives.

VIII.   Conclusion

71.   The demographic conciliation is an exercise that has proved very useful in the determination of the base population for the preparation of population projections in Latin America. However, it suffers from at least two limitations that should be solved

72.   On the one hand, the selection of initial values is an arbitrary and subjective process, which largely depends on the person responsible for its production. It is considered, that initial values selection must be done in such a way that does not depend on subjective judgements of the person in charge of its preparation.

73.   In last exercises developed in Mexico, has been eliminated this subjectivity using specific rules for the selection of these initial values, However, these alternatives do not consider all available information.

74.   It is considered that these solutions can be improved through the development of algorithms with statistical bases and which include all available information. In this paper, it is presented an alternative.

75.   Another limitation is that, although conciliation is defined as "the reconstruction of the demographic dynamics of the recent past", in general is limited to a selection of initial values, leaving virtually identical estimates of other variables, even when they are not consistent among themselves and with the selected population. Usually what is doing, once determined the base population, it is a backward projection for the period.

TABLE 1

ESTIMATED UNDER COVERAGE. SELECTED STATES

State	Sex	Survey	Conciliation	Bootstrap
Chihuahua	Both	1.78	4.37	5.46
	Male		4.05	4.82
	Female		4.69	6.08
Querétaro	Both	1.01	0.34	0.17
	Male		0.47	0.25
	Female		0.22	0.10
Tabasco	Both	0.55	0.61	0.47
	Male		1.02	0.90
	Female		0.22	0.06
Sinaloa	Both	0.47	2.88	3.27
	Male		0.47	0.25
	Female		4.35	5.10

TABLE 2

AGE GROUP SELECTED AND DIFFERENCES

State	Males		Females	
	Age group	Difference (%)	Age group	Difference (%)
Chihuahua	5-14	2.20	40-49	2.69
Queretaro	0-9	-0.78	0-9	-0.60
Tabasco	20-29	0.43	45-54	-0.43
Sinaloa	0-9	-0.78	25-34	2.47

Graph 2
Chihuahua. Males. 5-14 years

Graph 3
Chihuahua. Females. 40-49 years

Graph 4
Queretaro. Males. 0-9 years

Graph 5
Queretaro. Females. 0-9 years

Graph 6
Tabasco. Males. 20-29 years

Graph 7
Tabasco. Females. 45-54 years

Graph 8
Sinaloa. Males. 0-9 years

Graph 9
Sinaloa. Females. 25-34 years

12