XXVI IUSSP INTERNATIONAL POPULATION CONFERENCE, Marrakech (Marruecos), 2009.

Adjusting Small Population Projections by Symptomatic Variables. Colon Department (Argentina) 2008-16.

González, Leandro M.

INTERNATIONAL UNION FOR SCIENTIFIC STUDY OF POPULATION

XXVI IUSSP INTERNATIONAL POPULATION CONFERENCE

1905 Session - Obtaining data on special populations: survey techniques and sampling methodologies for developing countries

"Adjusting Small Population Projections by Symptomatic Variables. Colón Department (Argentina) 2007-16."

Autor: Leandro M. González (CONICET – Universidad Nacional de Córdoba)[1]

Correo electrónico: leandrogonzalez@yahoo.com.ar

ABSTRACT:

An adjustment procedure of small population projections is proposed in this paper, combining the techniques of components projections and symptomatic variables. Based on projections with two or more different migration scenario, symptomatic variables allow selecting the closest projection to poscensal estimation or defining a new projection. The poscensal migration derives from the proportional difference between the estimated population by symptomatic variables and the closer projections.

This methodology is particularly useful in populations with high migratory movements. It is applied to Colón department (Cordoba, Argentina), with an annual migration rate of 16 per thousand in 1991-2001 period. Estimates are made for 2007 with records of births, deaths, pupils and primary voters. Symptomatic variables confirm the continuing population growth, which could lead to an increase of 52 percent between 2001 and 2016 if assumptions of fertility, mortality and migration are maintained.

KEY WORDS: POPULATION PROJECTION – SYMPTOMATIC VARIABLES – SMALL POPULATION

---

[1] Doctor en Demografía, investigador del Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET) y del Centro de Estudios Avanzados (Universidad Nacional de Córdoba, Argentina).

1. INTRODUCTION

This article presents part of a larger research, which developed a methodology designed to adjust projections of small populations by symptomatic variables, after the last available census (Gonzalez, 2007). We analyze here the usefulness of the procedure as a tool for indirect estimation of migration level, as it represents one of the three key variables of population dynamics. Unlike the national or provincial populations, migration may represents the main factor of change of lower stocks.

Demographic projection techniques allow the updating of population data with significant degrees of approximation to reality. Based on assumptions scientifically supported, it is possible to have estimates of size and composition of populations, with disaggregations responding to the interest of researchers and users of information.

There are basically three types of population projection methods: the mathematical methods are used to estimate the total population of an area and are based on mathematical functions; the method of components (or cohorts) to simulate the growth and composition of the population from the analysis of components of population dynamics (fertility, mortality and migration); and symptomatic variables estimates, which involve the use of information associated with the population dynamics in absence of full demographic information or their quality is deficient.

Third mentioned process -symptomatic variables- uses available statistical records on small areas to detect trends in population growth. This allows complementing traditional projection techniques when there is no specific information. Examples of these variables are car registration records, new housing, taxes, public services, voters and pupils, among others.

Population estimates for small areas allow a multiplicity of uses, as input into governance and private economic activity. Projections can be used to plan public works and services, as well as their appropriate budgets. It also offers an adequate guide for the promotion of joint activities between the municipalities and private entities to promote local development.

In this paper we consider the possibilities of applying the techniques of population estimates and projections to estimate future population of small areas. In these cases migration is the most unpredictable factor of population dynamics. While fertility and mortality can be extrapolated by mathematical functions, migration depends on unpredictable factors. For this reason it is proposed a methodology that allows deriving indirectly the level of migration in a small population, after the last census.

First, the article describes the theoretical implications of the procedures used with symptomatic variables, analyzing their assumptions and methodological scope. Secondly, we propose the use of these techniques as a process of adjustment of population projections made by method of components, based on indirect estimation of migration level.

To illustrate the proposed procedure, it’s applied to Colón department (Córdoba, Argentina). This will attempt to show the effectiveness and limitations that can offer this procedure when it is applied to a test case. Colón department population was chosen for is its high growth, based on a significant immigration flux.

Colón department is on the northern edge of Córdoba city, capital of the homonymous province. According to the National Institute of Statistics and Censuses (INDEC), the population of Colón department was 125,402 inhabitants in 1991 and 171,067 in 2001, with a average annual intercensal growth of 29.5 per thousand (INDEC, 2003). In 2008 a provincial census was conducted resulting in a population of 205,030 inhabitants; annual growth in the period 2001-08 was 26.7 per thousand, and notes the persistence of a high population growth in the area. Unfortunately there are no quality studies in the coverage of this census, so it's taken as an indicative figure.

Córdoba city has a population of around 1.3 million inhabitants but a growth rate much lower. Córdoba in the period 1991-2001 showed an annual growth rate of 9.9 per thousand and 2.7 per thousand in 2001-08. This allows underlying the fast increase in the population of outlying areas, particularly driven by migration from the Córdoba city to neighbouring municipalities.

As we reduce the scale population, greater may be the effect of migration on the evolution of a population. Since components projections require assumptions of births, deaths and migrants, future evolution of migration in a small area may compromise seriously the effectiveness of the projections. For that symptomatic variables are proposed as a method of estimating a small population migration.

2 - THEORETICAL CONCEPTS

A population projection is a calculation procedure of the future evolution of a population, based on assumptions of fertility, mortality and migration levels. Projections are formal calculations that show the effects of assumptions on a known population (IUSPP, 1985:115).

The most frequently used procedure is the cohort component method, which performs a simulation of a changing population according to the components of growth: fertility, mortality and migration. Starting with the base population by age and sex, the population at each specific age is exposed to the possibility of death, reproduction and migration. Entire procedure is repeated for each year of the projection period, resulting in the projected population by age and sex for each year (ARRIAGA, 2001:309-10).

Other procedures for calculating prospective population are based on mathematical functions. Future population of an area is based on the results of recent population censuses and some mathematical function (usually exponential or logistic). Arriaga notes that these methods do not produce projections themselves, but population estimates. The main difference is that estimates only provide the age the total population stock, although the age composition can be estimated by other procedures (ARRIAGA, 2001:412-13).

A variant of mathematical methods for demographic estimates are symptomatic variables. Andrew Howe defines symptomatic variables as "*any available set of data which in some way relates to changes in population size*" (Howe, 2004:3). These are statistical records that are associated with the volume and change in a population.

CELADE specialists say:

> *Symptomatic variables methods attempt to face the challenge of update finding associated variables with population change, when information is permanent and good quality. This search involves finding variables that satisfy two requirements: i) present a high correlation with the size and evolution of the population, ii) to have permanent records. In truth, there are regular lists of many variables, such as records of consumers of basic services (electricity and water, for example), school enrollment statistics, electoral registers, records of vital statistics, construction housing, etc. (CELADE, 1998:78-79).*

A not insignificant number of these variables is related to demographic change but this link is not accurate. Because of this it is necessary to find a formula or algorithm that relates population size with symptomatic variables (CELADE, 1998:78-79). Main procedures used to estimate population of small areas by symptomatic variables are listed below (BAY, 1998; TEIXEIRA JARDIM, 2001; CHAVES ESQUIVEL, 2001):

- Apportionment distribution: Assumes that the population is divided between areas under the same ratio observed for the symptomatic variables. It is easy to calculate and requires information on a temporary time, but is very sensitive to the quality of records.
- Proportional distribution: Only difference with the previous method is that estimates is adjusted to changes in both symptomatic variables and total population, in two moments of time. It has the advantage of being consistent with the total population and is less sensitive to changes in coverage of the symptomatic variables; the disadvantage is it requires information for two moments of time.
- Censal ratio: It assumes that a local population changes in proportion to occurrence rates of a symptomatic indicator of the wider area. It requires information for two different dates.
- Difference rate: A variant of the previous method, which uses differences in the calculation of occurrence rates rather than reasons, and has the same scope.
- Composite: It uses different symptomatic records to estimate population of a local area by age group. Main advantage is the use of each symptomatic variable to estimate the age group that is more related with; the disadvantage is the need for a set of records simultaneously.
- Correlation of ratio: Population change is correlated with the variation of one or more symptomatic variables, through a linear regression model. Its main strength is the estimates are derived from probabilistic models. The disadvantages are the need for a reasonable number of observations in two moments of time to build the regression models, and the requirement of independence between symptomatic variables.
- Correlation of rate: It is a logarithmic fit of previous method, which attempts to adapt the estimation procedure for short periods of time.
- Correlation of difference: It is also a linear regression model, which measure demographic change in variation of differences rather than ratios. It has the same advantages and limitations of the correlation methods.

3 - METHODOLOGY

To use symptomatic variables to adjust small population projections, the necessary steps are:

1 - Collecting of symptomatic information (primary pupils, voters, births and deaths) for recent years (2002-2007).

2 - Estimate of total population of Colon department for the last year which symptomatic information is available (2007 in this case), using more accurate procedures in relation to last census (González, 2007).

3 - Projection of Colon department population by component method, for 2001-07 period. Projections are made from the same population base, vital records and assumptions of future fertility and mortality, and two different migration scenarios at least. This would produce two projections at least, that differ only in migration level.

4 - Comparison between projections and estimated population with symptomatic variables. Total population estimated in step 2 and projections set calculated in the previous step are collated. Comparison can confirm any of projections or recommend adjusting them. In the second case, a new hypothesis of migration can be defined by proportional difference between the estimated population with symptomatic variables and the closer projections: migration rate is derived applying the same proportional difference to the migration hypothesis of the closer projections. To calculate the annual migration rate from projections and symptomatic variables estimates, for a postcensal year, the following formula is applied:

$$m = \left[\frac{SN - CN_{inf}}{CN_{sup} - CN_{inf}} * \left(m_{sup} - m_{inf} \right) \right] + m_{inf} \quad (1)$$

Where:
m = annual migration rate.
SN = estimated population by symptomatic variables.
CN_{inf} = lower projected population by components than SN.
CN_{sup} = upper projected population by components than PS.
m_{sup} = migration rate of CN_{sup} projection by components.
m_{inf} = migration rate of CN_{inf} projection by components.

5 - Final projection 2007-16. A new projection by component method is built, based on 2007-estimated population by symptomatic variables and derived migration by formula (1). A period of 15 years after last national census (2001) is taken as an intermediate term, especially for fertility and mortality assumptions.

Applying the formula (1) requires at least two component projections, whose values could cover estimates from symptomatic variables. A migration rate can be derived from each population estimate, using the relative location of estimates from symptomatic variables respect to component projections. Under the assumptions of defined fertility and mortality hypothesis for all projections and equal quality of symptomatic information, the difference between estimates from symptomatic variables and component projections responds exclusively to migration.

Another possible interpretation could postulate that the relative difference between the estimates by symptomatic variables and component projections comes from combined effect of: 1) differences between assumptions and actual levels of fertility and mortality; 2) present migration level; 3) changes in quality symptomatic records. In this way, net migration rate represents an estimation of migration level plus a residual component (errors of assumptioms and variations in quality of symptomatic records).

4 - IMPLEMENTATION

The proposed methodology is applied to Colón department for 2001-16 period, to show its effectiveness and limitations in a test case. First component projections are describeb.

The base population (171,625 inhabitants) is obtained from 2001 census population (171.067 inhabitants), adjusted for census omission (2.75%) and correction of the population under 5 years from the population defined by INDEC for projection of Córdoba province (Gerencia de Estadísticas y Censos, 2003). Age structure of population over 5 years remains in proportion to the census results, as the test of quality for a declaration of age is acceptable.

Fertility is defined taking into account births by age of the mother registered during 2001-07, and total fertility rates calculated in 1991 (3.12 children per woman) and 2001 (2.33). A hypothesis that total fertility rates will decline gradually to 2.2 children per woman by 2010 and 2.1 by 2016 is defined. Mathematical functions are not used for rates extrapolation because they produce a very low fertility for the period 2010-16.

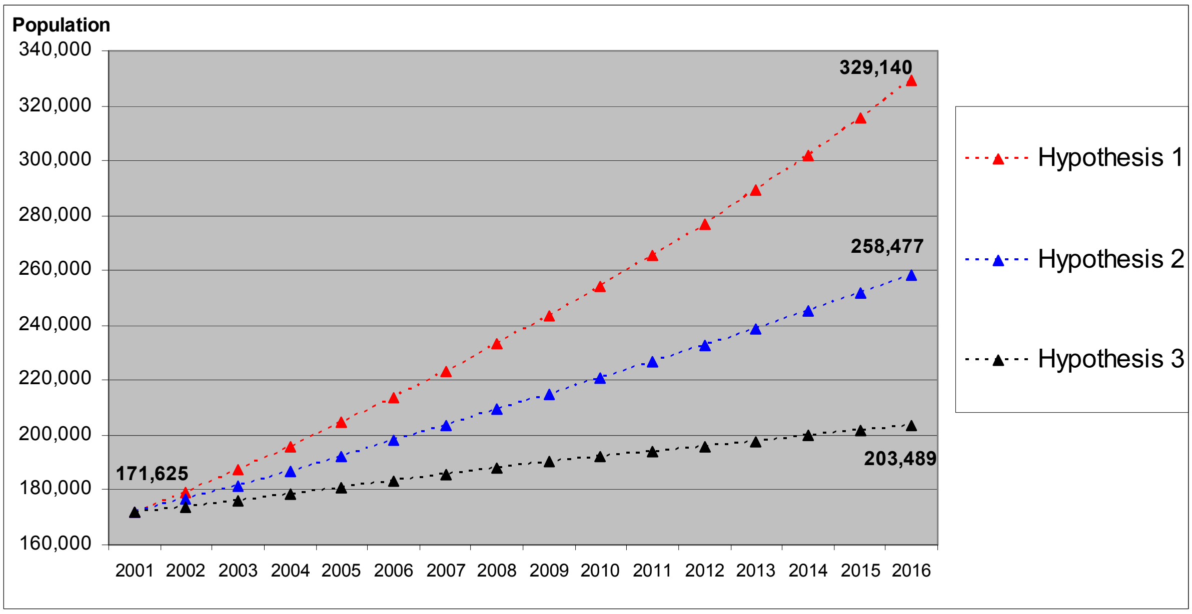
Mortality is estimated throug life expectancy at birth of the area formed by Colón and Capital departments, since the study population is close neighborg of Córdoba city. This proximity allows Colón´s residents to benefit from the medical supply of Cordoba city. Based on the life expectancies at birth for both departments in 1980, 1991 and 2001, an logistic extrapolation to 2016 is calculated. Projections also incorporate deaths by sex and age registrered in 2001-07 period.

In the case of migration, a set of three different scenarios are proposed on intercensal net migration rate (15.9 ‰):

- Hypothesis 1: assumes that annual migration rate is about twice the rate observed in the period 1991-2001, 31.8 per thousand inhabitants;
- Hypothesis 2: assumes that migration is equal to intercensal period, 15.9 per thousand inhabitants per year;
- Hypothesis 3: assumes no migration.

Due to the definition of three different scenarios of migration, three different population projections are obtained. As projections remain the same assumptions of fertility and mortality, differences between them respond only to different levels of migration. In Graphic 1 illustrates the resulting projections.

Graphic 1: Total population projected by migration assumption. Colón department 2001-16.



After population projections are developed, it is necessary to estimate postcensal population with symptomatic variables for years in which records are available. This will employ symptomatic variables and procedures that showed an error of less than 10 percentage points on the 2001 census for Colón department (Table 1).

Table 1: Percentage differences of estimated population with symptomatic variables respect to 2001 census. Colón department.

METHOD	VARIABLES SINTOMÁTICAS					
	Births	Deaths and pupils	Vital records	Primary pupils	Voters	Mean*
Apportionment distribution	-1.3		-8.2		-0.7	1.6
Proportional distribution				**6.1**	**0.1**	**-5.7**
Censal ratio				**5.8**	**0.2**	**-5.6**
Difference rate				**3.9**	**0.2**	**-6.8**
Composite	**-2.7**	**-2.7**		**-2.7**	**-2.7**	
Correlation of ratio		**-3.5**		**-7.1**		
Correlation of rate		-5.5		**-7.5**		
Correlation of difference		**2.8**				

(*) Average of the estimates made with births, vital events, students and voters.

Source: González (2007:103).

Table 1 shows that the most accurate estimates are derived from voters, throug the first four procedures; differences are smaller than 1 per cent. In descending order of accuracy, following estimates are obtained from records of births by apportionment distribution method for (-1.3%), and average of births, vital events, students and voters with the same procedure (1.6%).

In third place are estimates obtained by composite method, which employs a births, primary students, voters and deaths records together (-2.7%). Below is the correlation of differences method, from records of deaths and students (2.8%). Followed in order of accuracy is correlation of ratio method, based on records of deaths and students (-3.5%). In fifth place is estimation of difference rate method, based on primary school enrollment (3.9%).

Finally there are a number of models whith differences between 5 and 10 percentage points: correlation of rates on deaths and students (-5.5%), the average of estimates by censal ratio and proportional distribution (-5.6 and -5.7%), and censal ratio applied to students (5.8%).

The collected postcensal information for Colón department are births, deaths and primary school enrollment for the years 2001-07; registered voters in elections in 2003, 2005 and 2007. Table 2 presents the figures.

Table 2: Symptomatic variables selected of Colón department, 2001-07.

Variable	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Primary pupils	23,739	24,346	24,303	24,716	24,941	24,330	25,758
Voters	120,058		127,926		136,265		144,430
Births	2,835	3,196	3,402	3,818	3,753	3,780	3,838
Deaths	1,050	1,075	1,209	1,245	1,227	1,261	1,499
Vital events	3,885	4,271	4,611	5,063	4,980	5,041	5,337

Source: National Ministry of Health, Provincial Ministry of Education, and Federal Courts of Córdoba city.

Above information is used to calculate the population of Colón department between 2002 and 2007. Table 3 details the obtained estimates by applying the procedures selected.

Table 3: Estimated population with selected symptomatic variables. Colón department, 2007.

METHOD	**Population estimates by symptomatic variables**						
	Births	**Vital events**	**Primary pupils**	**Deaths and pupils**	**Voters**	**Mean***	**Every variables**
Apportionment distribution	235,453	216,358			**201,943**	221,698	
Proportional distribution			**201,682**		**204,385**	220,470	
Censal ratio			**201,426**		**204,486**	220,815	
Difference rate			**199,666**		**204,407**	220,241	
Composite							**214,956**
Correlation of difference				227,201			
Average estimate	213,012						
Limited average estimate	**204,119**						

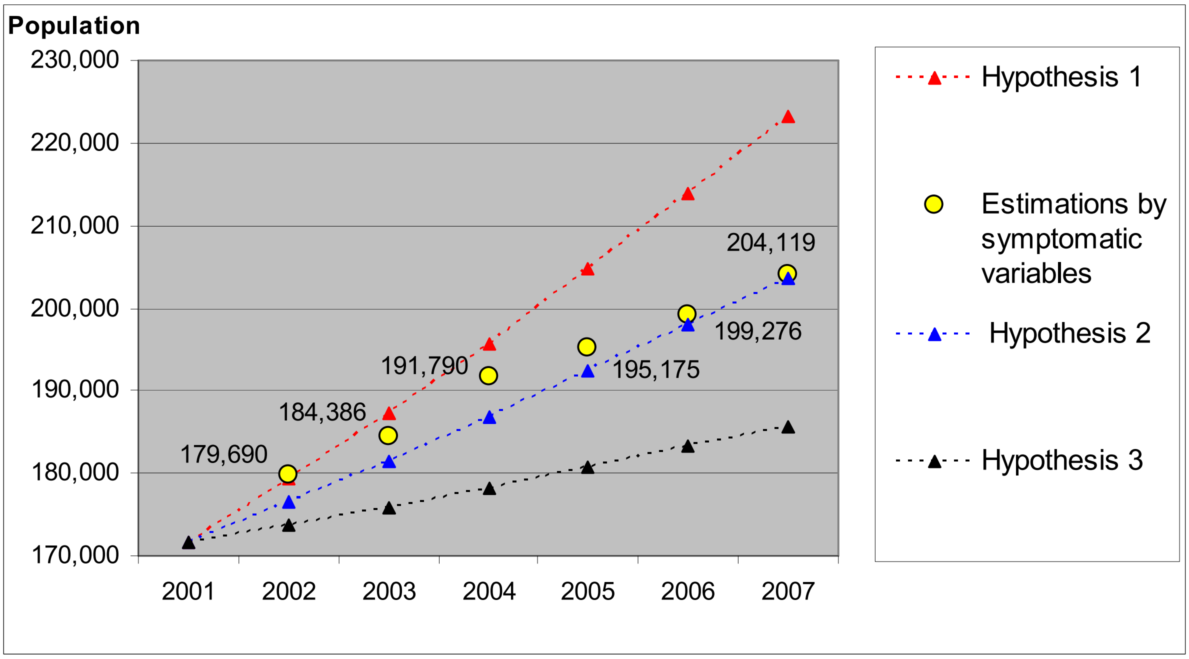
(*) Average of estimates made with births, vital events, primary pupils and voters.

Source: Calculations based upon data from National Ministry of Health, Provincial Ministry of Education, and Federal Courts of Córdoba city.

As there are many estimates for each year is recommended to calculate an average, excluding outlying values. This average also diminishes the effect of implicit assumptions in each procedure and the limitations of each source. Estimated population for 2007 year is taken as a “limted” average -i.e. excluding the extreme values- equivalent to 204,119 inhabitants. If all estimates are considered, the average reaches 213,012 inhabitants, a higher value according to estimates from previous years and the figures of provincial census 2008 (205,030 inhabitants).

Graphic 2 shows the contrast betwen the estimates by symptomatic variables and component projections. For 2002 and 2003 year symptomatic variables indicate that population would have followed a path very close to projection 1 (migration 100% higher than registered between 1991 and 2001). In 2002, the projected population overcomes the estimated population (179,690 inhabitants).

Graphic 2: Estimated population by symptomatic variables and projections. Colón department, 2001-07.



Adopted estimate for 2004 shows that population would be located between projections 1 and 2. Estimates for 2005 to 2007 years are progressively closer to projection 2. This indicates a slowing population growth since 2002, near to the middle migration rate.

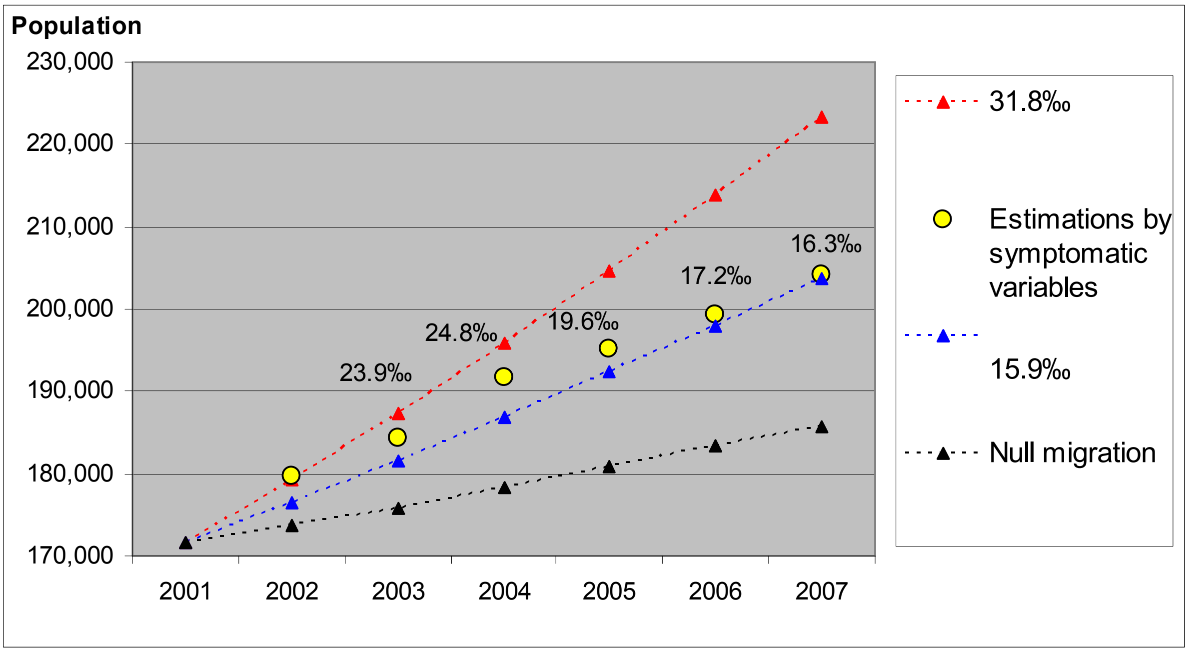
As estimate for 2007 year is slightly higher than projection 2, there are two options: adopting the nearest projection (hypothesis No. 2) or to run new projection. If the second option is chosen, estimate of migration level is requiered. Table 4 and Graphic 3 present the calculation of net migration rates for each year after the census, according to the proposed formula (1).

Table 4: Calculation of the annual migration of the population of the Columbus department with symptomatic variables, years 2002-07.

	2002	2003	2004	2005	2006	2007
Estimated population by symptomatic variables	179,690	184,386	191,790	195,175	199,276	204,119
Proyected population hypothesis 1	179,262	187,256	195,745	204,685	213,893	223,310
Proyected population hypothesis 2	176,446	181,448	186,758	192,324	197,952	203,577
Migration rate hypothesis 1 (‰)	31.8	31.8	31.8	31.8	31.8	31.8
Migration rate hypothesis 2 (‰)	15.9	15.9	15.9	15.9	15.9	15.9
Estimated migration rate (‰)	*	**23.9**	**24.8**	**19.6**	**17.2**	**16.3**

(*) It is not calculated because the estimated population by symptomatic variables exceeds the projected population by components and migration rate may have a distorted value.

Graphic 3: Annual migration rates derived from estimates by symptomatic variables and component projections. Colón department, 2002-07.

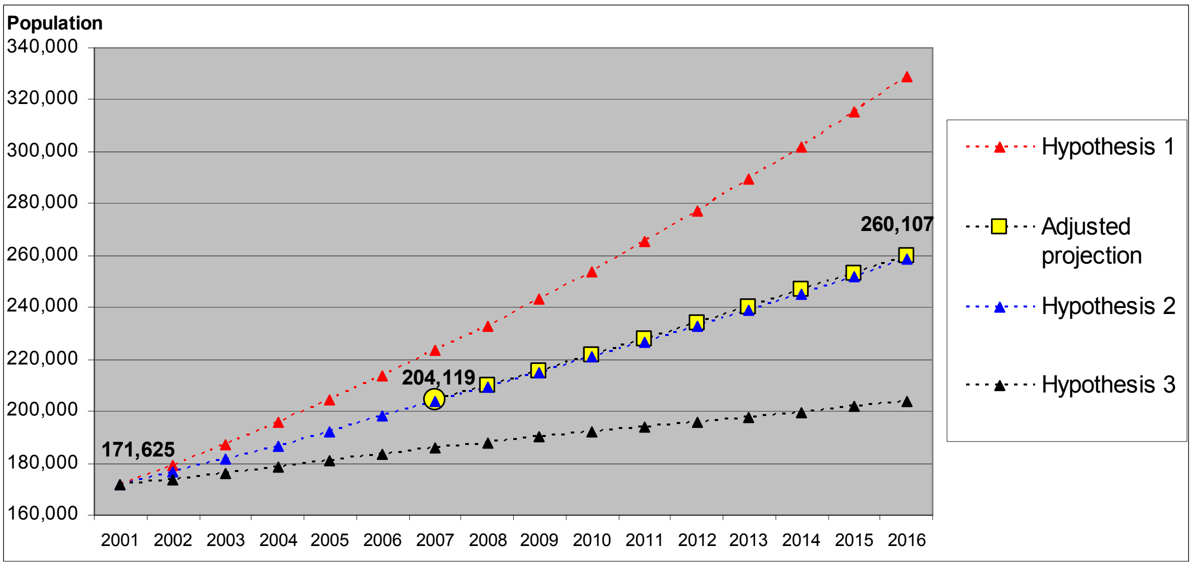


During the considered period only the 2002 year estimate exceeds the highest migration hypothesis, so its migration rate should be recalculated from a new projection by components having a higher rate of migration 1 (>31.8 ‰).

For 2007 year, symptomatic variables suggest a migration rate of 16.3 per thousand. That means that in 2007 the population of Colón department was evolving under the fertility and mortality assumes of all projections, and an annual net migration rate of 19.6 per thousand inhabitants.

Once the new migration rate was calculated, it is able to run a new projection. New base population is defined by the symptomatic variables (204,119 inhabitants). Structure by sex and age can be applied proportionately to the neares projected population (hypothesis No. 2) for the same year. After incorporating the estimated level of migration, 19.6 per thousand, divided by sex and age. If the same assumptions of fertility and mortality are kept, Colón department can reach 260,107 inhabitants in 2016, which represents a population increase of 52 percent in just 15 years (Graphic 4).

Graphic 4: Total projected population according to adjustment by symptomatic variables. Colón department, 2007-16.



This new projection assumes that the estimated migration rate for 2007 year remains constant all the period. This assumption can be reviewed regularly, as new symptomatic records are available. This allows making a new projection each time that reliable information be availabe, or select the projection that user considers most convenient.

5 – FINAL REMARKS

The proposed procedure for adjusting a projected population by symptomatic variables, do not theoretical and methodology inconveniences that invalidate its use. This is due to the separate application of each method, since the adjustment is made to the results of each procedure and is based on the opinion of the researcher. Its originality lies in the definition of the population base of a projection by symptomatic variables and a net migration rate, for postcensal years.

The proposed adjustment profits from advantages of each procedure and smooths their limitations. Method of components offers robust projections, a detailed annual population structure and a wide range of demographic indicators. It also be able to test demographic projections under different scenarios, and incorporate available vital statistics. In the other hand, its accuracy can not be assessed until a new census, usually every ten years in Argentina.

The symptomatic variables offer the possibility of estimating population for each year that records are available, and may know the accuracy of each procedure to the last census. They allow having estimates for the years after the last census, with a delay of approximately one to two years with respect to the present time. Their main drawback is that only the total population is calculated, without age and sex structure.

It must also consider that records used as symptomatic variables may have quality problems. Administrative records often have varying degrees of coverage, omissions or bias (e.g. the actual residence of voters or registration of vital statistics). These shortcomings are difficult to solve if there are not coverage surveys of the sources. Even so the data can be used if biases remain relatively constant over the period considered and tested procedures demonstrate an acceptable approach to the census results.

With regard to implementation of the proposed methodology to Colón department, results show the persistence of a sustained population growth. Estimate of annual migration rate implies that migration in 2007 was slightly higher than in 1991-2001 period (15.9 ‰). If these parameters were kept constant, estimated population can reach 260.107 inhabitants in 2016, implying a growth of 52 percent in just 15 years.

In short, the splicing of both methods can redo a set of projections for each year after the last census. Although censuses are the appropriate instruments to assess the demographic projections, symptomatic variables allow adjusting projections annually until a new census. This would result in a significant saving of time for users of demographic information, which can use adjusted projections until the disposition of the next census results.

Finally, this proposal can be particularly useful for small populations that register significant changes. As it has shown in the case of Colón department, migration can become the most important dynamic factor in a smaller population. Symptomatic variables can update the last census data, register indirectly the migration level annually, and provide a medium or long term view when are combined with component projections. In this way, provided information may be a valuable tool for planning of public and private management.

BIBLIOGRAPHY

ARRIAGA, Eduardo (2001): *El análisis de la población con microcomputadoras*. Córdoba, Doctorado en Demografía - Universidad Nacional de Córdoba.

BAY, Guiomar (1998): “El uso de variables sintomáticas en la estimación de la población de áreas menores”, in Notas de Población magazine; Santiago, CELADE, Nº 67/68, p. 181-208.

CELADE (1998): *Demografía II*. México, Programa Latinoamericano de Actividades en Población, Carlos Welti Editor.

CHAVES ESQUIVEL, Edwin (2001): “Variables sintomáticas en las estimaciones poblacionales a nivel cantonal en Costa Rica”, in Notas de Población magazine; Santiago, CELADE, Nº 71, p. 51-72.

GONZÁLEZ, Leandro (2007): Ajuste de proyecciones de poblaciones menores con variables sintomáticas. Departamento Colón 2001-16. Córdoba, Ph.D. Thesis in Demography, Universidad Nacional de Córdoba, unpublished.

INDEC (2003): Censo Nacional de Población, Hogares y Viviendas 2001. Buenos Aires, INDEC.

IUSPP (1985): *Diccionario Demográfico Multilingüe*. Liege: CELADE.

HOWE, Andrew (2004): "Assessing the accuracy of Australia's small área population estimates, 2001". Canberra, Australian Population Association, http://www.apa.org.au/upload/2004-5C_Howe.pdf.

TEIXEIRA JARDIM, María L. (2001): "Uso de variables sintomáticas para estimar la distribución espacial de población. Aplicación a los municipios de Rio Grande do Sul, Brasil", in Notas de Población magazine; Santiago, CELADE, Nº 71, p. 21-50.