The many Mexicos. Stochastic forecast 2001-2050*

Katja Kesseli y Carlos Galindo

University of Helsinki/
Netherlands Interdisciplinary Demographic Institute

Abstract

Demographic data from Mexico has serious problems of coherence. Most recent estimations are increasing the discrepancies instead of reducing them. Based on analysis of inconsistencies in demographic data, we made a stochastic forecast of Mexican population for the period 2001-2050. The stochastic forecast is composed of random simulations of fourdifferent scenarios, which are given by dissimilarities on demographic estimations of the period 1985-2000. This technique allowed us to take into account the uncertainty embedded in Mexican data. Our results imply that is very unlikely (probability 0.07) that Mexican population size in 2005 was lower than 103.2 millions as published in the recent official population count. This result adds up to requests made by other researchers about revisiting and composing consistent demographic estimations.

Key words: stochastic forecast, consistent demographic estimations, data evaluation, Mexico.

Resumen

Los muchos Méxicos: proyección estocástica 2001-2050

La información demográfica en México presenta problemas de coherencia. Las estimaciones más recientes incrementan las discrepancias en lugar de reducirlas. Basados en un análisis de las inconsistencias entre varias fuentes, realizamos una proyección estocástica de la población mexicana para el periodo 2001-2050. La proyección se compone de simulaciones aleatorias de cuatro diferentes escenarios dados por las mayores divergencias en las estimaciones del periodo 1985-2000. Esta técnica nos permitió considerar la incertidumbre embebida en los datos mexicanos. Nuestros resultados implican que es muy poco probable (0.07) que el tamaño de la población mexicana en 2005 fuera menor a 103.2 millones de personas, como fue publicado en el reciente Conteo de Población. Este resultado se suma a las peticiones de otros investigadores sobre la necesidad de revisar y realizar estimaciones demográficas consistentes.

Palabras clave: proyección estocástica, dinámica poblacional, evaluación de fuentes, estimaciones inconsistentes, México.

Introduction

Here we performed a stochastic forecast of Mexican Population for the period 2001-2050. We chose the stochastic approach because it allowed us taking into account the uncertainty embedded in present

* Acknowledgements: thanks to Juha Alho for useful comments and guidance. This research was possible thanks to the financial support of the University of Helsinki, the Netherlands Interdisciplinary Demographic Institute, and the Consejo de Ciencia y Tecnología del Estado de Aguascalientes, and also to the Max Planck Institute infrastructure and working environment.

demographic data and its future behaviour. The assumptions on levels and trends of demographic variables are expressed in terms of predictive probability distributions, i.e. we ascribed probability values to different scenarios. Based on analysis of problems in Mexican data we considered four initial scenarios. Then, using the Program for Error Propagation, we generated one thousand simulations for each scenario (simulations consist on assigning random values to the future behaviour of demographic variables according to selected probability distributions). The simulated future trends are applied to the initial population of each scenario using the cohort-component method (linear growth model with a Leslie matrix). The Leslie matrix is built with age-specific fertility and mortality rates for both sexes at each time step, net-migration is given in absolute numbers also for both sexes at each time step. The set of all the random simulations for all the scenarios is the outcome of the stochastic forecast. The results are presented in the form of prediction intervals.

This paper is divided in two major sections. First we present the dataset used in the stochastic forecast, which includes jump-off populations, and fertility, mortality and migration levels and trends (description of problems in Mexican demographic data is based only in public information, and does not refer to any confidential data or personal communications). In this section we also assign future scenarios. Then we add a small description on how the future uncertainty is quantified in probability distributions. The second part of this paper is the presentation of our results.

Dataset

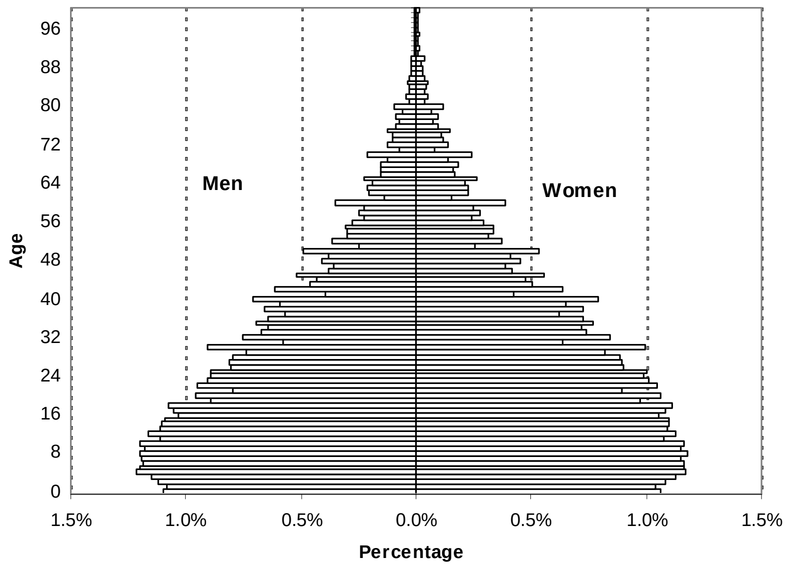
Jump-off populations

Today there is no agreement on the amount of Mexican population. Two national institutes are related to population figures and estimates: The National Institute of Statistics, Geography and Informatics (INEGI) which is in charge of Vital Statistics and National Census; and the National Population Council (CONAPO) which is in charge of population estimations and projections. INEGI (2006) has just published definitive results of their Second Population Count (a fast national census), and they claim that in October 2005 Mexican population consisted of 103.1 millions. The actual counted population was slightly less than 100.5 millions; INEGI (2006) added to this number more than 2.8 millions that were assumed to live in more than half a million houses where no information was collected. On the other hand, Conapo (2003) projected that in the middle of 2005 Mexican population should consist of 106.5 millions. The main reason of disagreement is that INEGI does not publish official estimates of undercounting of any national census, so Conapo makes its own estimations of under coverage and population size.

There are several reasons why undercounting in Mexican census is expected: huge territorial extension and geographical diversity, several communities where ethnic tongues are spoken, regions with high rates of no-response, organizational problems and etcetera. In the recent population count of 2005, besides problems mentioned above, there were major difficulties related to the period chosen for data collection (the month of October): after the raining season in Mexico (June to October) unpaved roads are flooded and some of them can not be used, the season of tropical storms and hurricanes starts in August and ends in October, so there were many flooded and displaced communities during this month (the 2005 Atlantic basin hurricane season was the most active and destructive season on record according to Gray and Koltzbach 2005); also major agricultural cycles start in October (sowing season of cotton, beans, maize and wheat, and harvesting season of maize, sugar crane and coffee), many communities moved to the fields to work and lived there during this month; and finally, two major national celebrations that last almost the entire month (El Festival Internacional Cervantino y Las Fiestas de Octubre) influenced population labour mobility in the centre of Mexico (information from several sources listed in References section, among them The Tropical Meteorology Project, Servicio Meteorológico Nacional, Centro Nacional de Prevención de Desastres CENAPRED, Secretaría de Turismo (SECTUR), Secretaría de Agricultura, Ganadería, Desarrollo Social, Pesca y Alimentación (SAGARPA), and Conapo). Because of these problems the figures reported for 2005 are highly controversial. Data validation is needed for achieving demographic coherent trends.

We decided to avoid using the controversial figures of 2005. The next most reliable estimation can be taken from the 2000 Population Census conducted by INEGI (2006). Even if INEGI does not give an estimation of undercounting, the data has been fully published and its quality can be assessed. The census considers usual resident population, referring to the people that live most of the year inside Mexican territory (is not *de facto* population because residents can be outside the country for a brief period, but is not *de jure* population in the European sense of legal residency).

FIGURE 1
MEXICAN POPULATION AS PUBLISHED IN 2000 CENSUS, INEGI



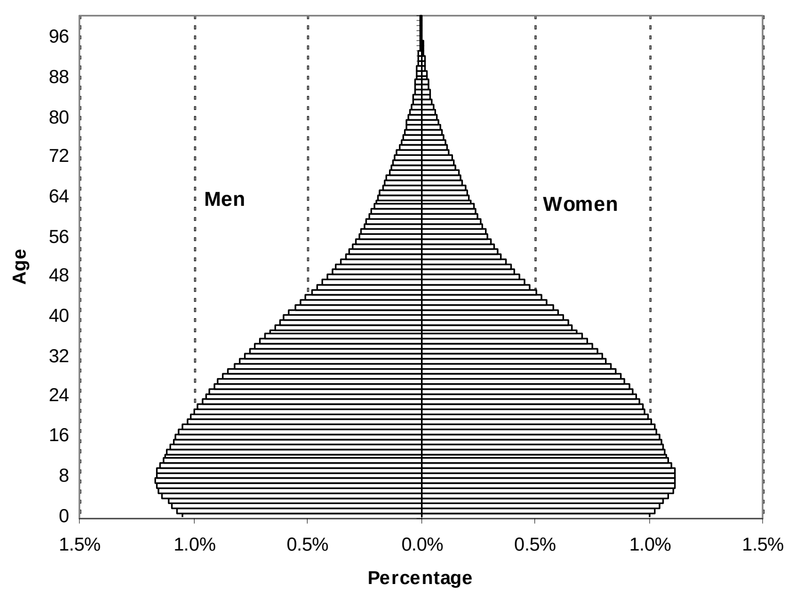
Source: INEGI (2006).

The total population actually counted in the 2000 census was 95.4 millions. INEGI estimated that more than 425.7 thousand houses were not included in the census, so they assumed that 4 persons lived in each one of these houses (5 persons for the state of Chiapas), so the total published population was 97.5 millions. Conapo (2003) corrected the 2000 census population: age heaping problems and undercounting were corrected using Luther and Rutherford method; Conapo also considered that the inclusion of 4 persons inside the uncounted houses led to bias, they assumed that only 1 person lived in each of these houses and distributed these persons only on working ages; they also corrected the excessive age declaration or rejuvenation effects using the Gray method based on moving averages. Conapo applied these corrections to the population for every State in Mexico and they estimated total population size of 100.5 millions.

If we compare the population counted by INEGI (95.4 millions) and their total published population (97.5), then the census undercoverage was 2.2 per cent. If we use counted population and corrections made by Conapo (100.5

millions) the undercounting was 5.1 per cent. If we think about INEGI's total published population and Conapo corrections the undercounting was 3 per cent. The estimations of undercounting using corrections by Conapo are reasonable numbers if we consider that the US Census Bureau (2001) estimated an undercounting in the 2000 US census of 2.3 per cent, Statistics Canada (2001) estimated a gross undercoverage in their 2000 census of 4 per cent and a net undercoverage of three per cent, and National Statistics UK (2006a) estimated that the undercount in their 2001 census was equal to 6 per cent in England and Wales.

FIGURE 2
CORRECTED MEXICAN POPULATION BY CONAPO, TAKEN TO 1 JAN 2001

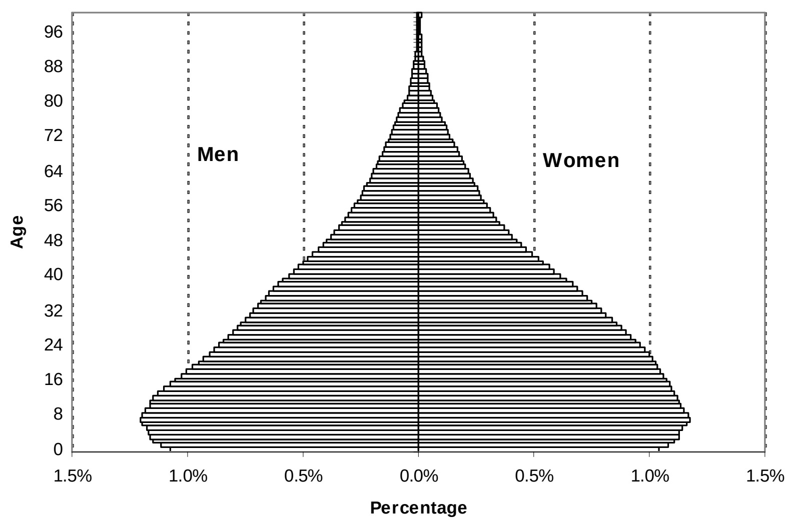


Source: Conapo (2003).

Thus we have two possible starting populations, the one published by INEGI and the one corrected by Conapo (taken to 1 January 2001). We decided to use both as jump-off populations (Figures 2 and 3); each starting population will imply a different forecasting scenario, all scenarios combined will shape a final forecast named Many Mexicos. The published population by INEGI has heaping problems and includes almost 2 million people with age not-specified.

To correct these problems we made five years age groups, then we used the correction method (1/16) for age and sex distribution proposed by UN, and finally we distributed the individuals with not-specified age by giving bigger weights to those ages which are usually undercounted (infants, very old people, and working individuals living alone). Single age groups were obtained using Karup-King multipliers. This population was also taken to 1 January 2001 (figure 3).

FIGURE 3
OWN CORRECTIONS TO MEXICAN CENSUS POPULATION, TAKEN TO 1 JAN 2001



Source: own corrections to population from INEGI (2006).

Vital rates

Conapo (2003) used its own corrected 2000 population to reconstruct population dynamic trends in the period 1990-2000. Then they used estimated trends to make a population projection until year 2050. Ordorica (2001) showed that the estimated births from Conapo differ significantly from the registered births in the Vital Statistics (collected by INEGI), then this author explained that considering the different birth rates, from Conapo and Vital Statistics, the estimated population growth and net migration were also significantly different;

his conclusion is that a revision is needed for coherence in Mexican demographic data. The estimations of migration by Conapo become more problematic when compared to estimations of the US Census Bureau. The main destination for Mexican emigration is the United States; for example, Alba (2000) estimated that 96 per cent of Mexican emigrants in 1995-2000 went to this country. Durand and Massey (1992) explain that estimations of migration between these two neighbour countries have been controversial since the 30's; they quote Corwin when he refers to the estimates of undocumented migration as a *numbers game*. These authors also give examples of several papers that participated in this migration debate arguing for 'false assumptions' and 'wishful calculations'. The emigration estimates of Conapo (2003) in 1990-2000 were based on the assumption of eight million Mexicans living in the United States, but the US Census Bureau (2006) counted 9.2 millions in the 2000 Census. Furthermore, the estimated emigration trend by Conapo, which always increases during the whole period 1990-2000, contradicts other estimations. Ordorica (2001) estimated that Mexican emigration was 40 per cent higher than Conapo figures for the period 1990-1995, and Alba (2000) estimated a smaller amount of emigrants than Conapo for the period 1995-2000, and the combined estimations of these two authors give numbers somewhat closer to the amount of Mexicans counted in the US Census.

Given this lack of coherence in Mexican demographic data, the different estimations of vital rates suggest different scenarios. For some scenarios we used vital rates estimated by Conapo (2006b) and for other scenarios we used vital rates estimated from Vital Statistics. These last rates are not published by INEGI, so we made our own estimations derived from registered births and deaths. With our estimation of vital rates we calculated net amounts of Mexican emigration. Our goal was to achieve a migration estimation coherent with calculations of Ordorica (2001) and Alba (2000), and therefore coherent with figures of US 2000 Census.

Population for own estimation of vital rates

Population data can be obtained from national census, information about births and deaths can be obtained from Vital Statistics. We assumed that both sources have the same underestimations problems, so when rates are calculated then the under coverage from both sources cancels out. We acknowledge that these sources, and our assumption, are subject to criticisms and that they present serious problems, but we also believe that they can be used as a starting point for avoiding the vicious cycle of lack of demographic coherence.

Data about births with the age of the mother (needed for age specific fertility rates) is only available for the period 1985-2000; therefore we made our estimations only for this period. The census data for years 1970, 1980, 1990 and 2000 is available from INEGI (1996, 2006). Populations are given by five years age groups with different last open groups, to make them comparable we set all published populations with highest age group of 85 and more (this also avoids problems of excessive age declaration in higher ages). We also corrected these populations using the 1/16 method of United Nations. From INEGI (1996) we know that 1980 census faced many problems. The total published population of 1980 was estimated using an undercoverage survey (Conapo 2003). This has been the only Mexican census with a post-survey to correct for undercounting; in this sense the 1980 population differs from all the others. We want to use comparable populations over time, even if we have to assume similar undercounting problems, so we can not work with 1980 corrected population. Instead we used information from 1970 and 1990 censuses to estimate a coherent population trend that includes the year 1980.

We reconstructed the mid-year population classified by sex and 5 years age groups for all the reference period 1985-2000. First we distributed people whose age was unknown into all the age groups. We gave bigger weights (a 2 per cent increase from proportional distribution) to age groups that usually present underestimation problems (small children, young adults living alone and old people). Second step was moving census populations towards mid-year estimations. The growth rates were calculated for different periods in-between 1970-2000 based on census data (we assumed geometric growth). The adjustment of 1980 was done using growth rates from the periods 1970-1980, 1980-1990, and 1970-1990. Using these rates, populations of 1970, 1990, and 2000 were taken to middle years. Population of 1980 was estimated as the average of forward projection of 1970 population and backward projection of 1990 population using the adjusted rates. The mid-year population for years 1985 and 1995 were also calculated as the average of forward and backward projections. The last step was to formulate mid-year population for every year of the first and second half of 80's and 90's decades. The first half of each decade was formulated using forward projections and the second half using backward projections (results shown in table 1 and figure 4).

TABLE 1
TOTAL POPULATION SIZE, MIDDLE YEAR ESTIMATIONS (MILLIONS)

1985	1986	1987	1988	1989	1990	1991	1992
73.227	74.858	76.526	78.182	79.923	81.703	83.217	84.758
1993	1994	1995	1996	1997	1998	1999	2000
86.330	87.929	89.558	91.217	92.907	94.629	96.382	98.168

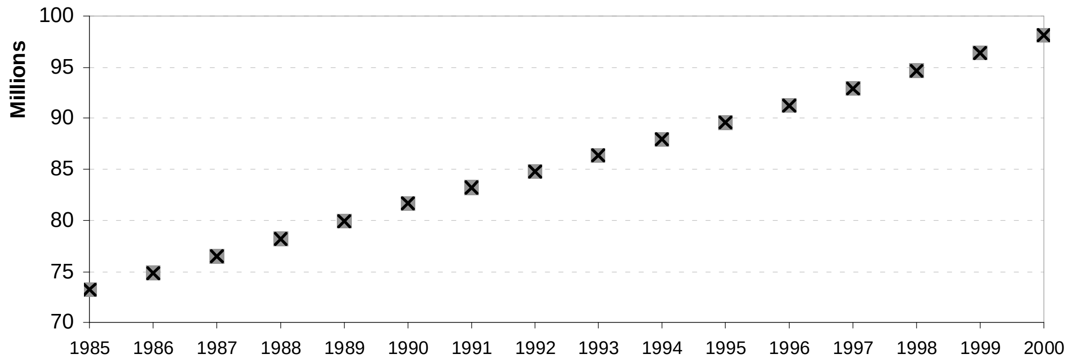
Source: own calculations.

Fertility

We used the registered births to estimate fertility rates. The main problem of birth registries in Mexico is the timing of birth registration. It is parent's duty to register the newborn child in the Register Offices but there is no maximum age specified by law at which parents have to register their child. So it is common that births are registered some years after actual birth year. Because of this problem most recent fertility rates are lower than expected and comparison over time is questionable. There is also the problem that some children are registered more than once. Galindo (2004) dealt with these problems and estimated occurred and registered births for each year during the period 1950-2000. Basically this author made projections following the behaviour of birth registries for each age of the registered child, then using Lexis diagram the projected registered children were ascribed to a given year of birth; in this way all years were made comparable.

From Vital Statistics we obtained the age of the mothers of registered births at age 0 of the newborn for every year in the period 1985-2000. Data for previous years was lost in the 1985 earthquake in Mexico City, this was the main reason why we restricted our study period only the years already mentioned. Using the age pattern of the mothers from Vital Statistics, the total estimated births by Galindo, and our population estimates, we calculated the age specific fertility rates for each year during the period 1985-2000 (results shown in figure 5 and table 3). Here we believe that is very important to note that these data, age-specific fertility rates from Vital Statistics, have not been estimated before and it will be very fruitful, for future research, to analyze it and compare it with official estimations (which are not based on Vital Statistics).

FIGURE 4
ESTIMATES OF MEXICAN TOTAL POPULATION FOR VITAL RATES



Source: own calculations based on INEGI (2006).

TABLE 2
OCCURRED BIRTHS (MILLIONS) AND TFR, ESTIMATIONS FROM VITAL STATISTICS DATA

Year	1985	1986	1987	1988	1989	1990	1991	1992
Births	2.483	2.497	2.520	2.532	2.540	2.604	2.597	2.634
TFR	4.18	4.09	4.03	3.80	3.72	3.73	3.67	3.65
Year	1993	1994	1995	1996	1997	1998	1999	2000
Births	2.617	2.598	2.543	2.510	2.477	2.516	2.531	2.555
TFR	3.46	3.37	3.25	3.15	3.05	2.97	2.93	2.90

Source: Galindo (2004) and own calculations.

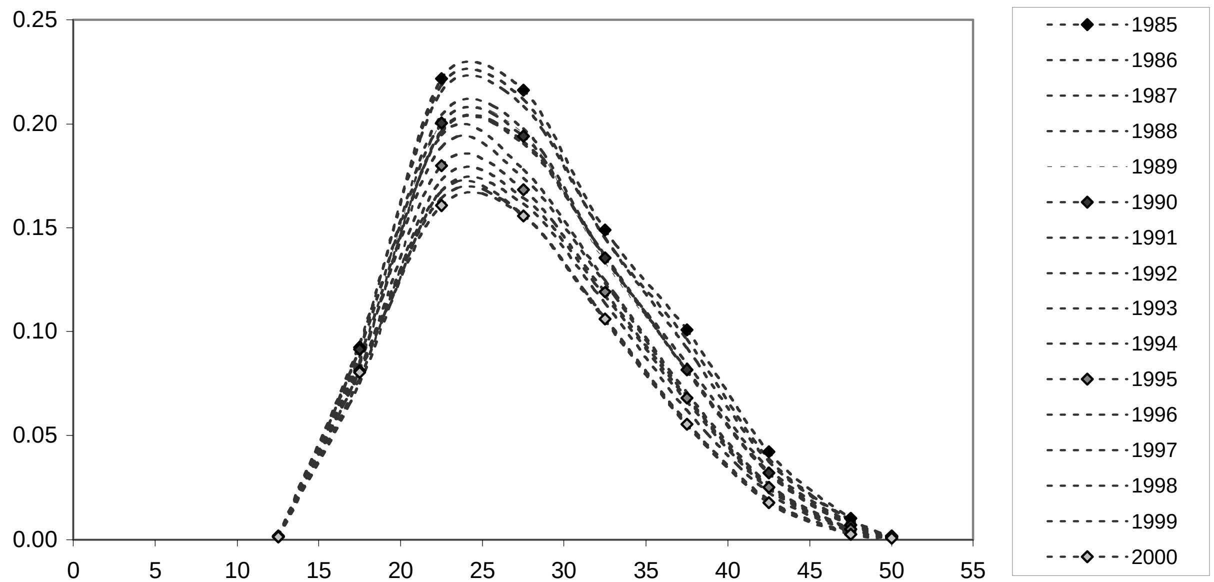
Finally, we need age-specific fertility by single year age groups, for the year 2000, in order to forecast future behaviour. We estimated single year rates using Karup-King multipliers. Thus we obtained two scenarios for fertility, the one estimated by Conapo (2003) and another one estimated from Vital Statistics data. Both scenarios are shown in figure 6.

Fertility forecasting assumptions

We also need to make assumptions about the future behaviour of the Total Fertility Rate (TFR) and the Mean Age at Childbearing. Mexican TFR presents a steady decline in several estimations as it can bee seen in Table 4 (we made our TFR estimations using the age-specific fertility rates described above).

Conapo (2006b) and UN (2006) assume a steady decline of Mexican TFR until it reaches the value of 1.8 and then they keep it constant. In order to explore also a different scenario, we argue that Mexico shares many common features with other countries that have seen their fertility reduced to lower levels. For example, Mexico as Italy and Poland has strong family traditions, the majority of people are catholic, its unemployment rates are high, there is a big proportion of precarious jobs, there is no universal health insurance, there is no unemployment governmental insurance, and the state does not guarantees schooling and health services for all children.

FIGURE 5
AGE SPECIFIC FERTILITY RATES, 5 YEAR AGE GROUPS, MÉXICO 1985-2000



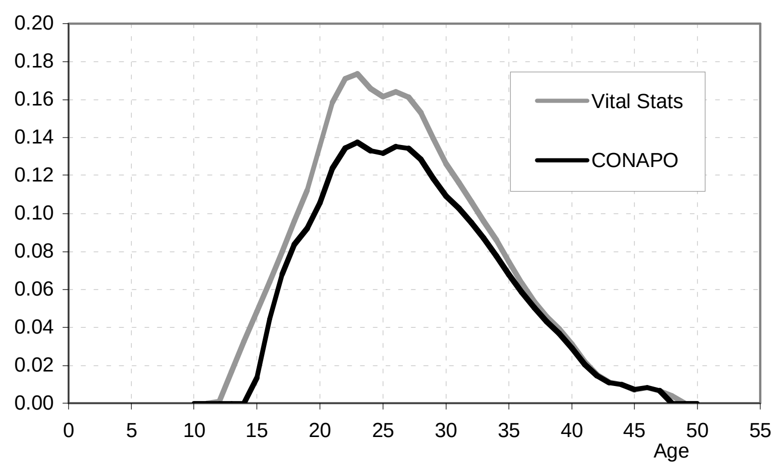
Source: own calculations.

TABLE 3
AGE SPECIFIC FERTILITY RATES

Age	1985	1990	1995	2000
10 - 14	0.0017	0.0018	0.0013	0.0013
15 - 19	0.0922	0.0913	0.0809	0.0802
20 - 24	0.2216	0.2005	0.1800	0.1609
25 - 29	0.2162	0.1941	0.1685	0.1559
30 - 34	0.1491	0.1355	0.1191	0.1060
35 - 39	0.1009	0.0815	0.0680	0.0553
40 - 44	0.0422	0.0320	0.0252	0.0179
45 -49	0.0103	0.0071	0.0050	0.0027
50	0.0013	0.0016	0.0019	0.0008

Source: own calculations.

FIGURE 6
TWO SCENARIOS OF AGE-SPECIFIC FERTILITY, CONAPO AND OWN ESTIMATION FROM VITAL STATISTICS, MÉXICO 2000



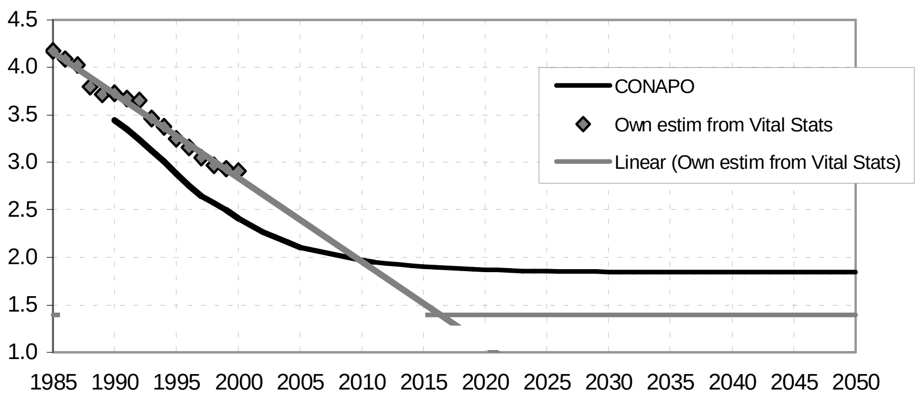
Source: own calculations and Conapo (2003).

TABLE 4
TFR ESTIMATES

Period	Conapo	UN	Own estim.
1985-1990		3.70	3.83
1990-1995	3.17	3.20	3.54
1995-2000	2.63	2.70	2.85

Source: Conapo (2006b), UN (2006) and own calculations.

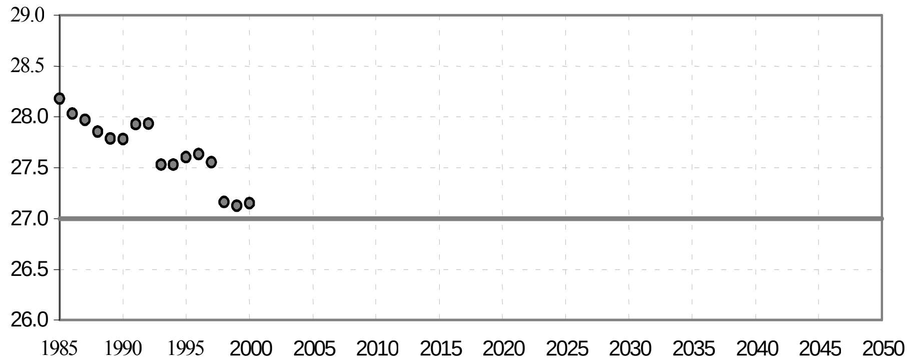
FIGURE 7
ESTIMATION OF MEXICAN TOTAL FERTILITY RATE, AND POSSIBLE FUTURE TRENDS



Source: own estimation and Conapo (2006b).

These characteristics lead us to believe that Mexican people will face the same problems on family formation and childbearing as people in Italy and Poland, and so, Mexican fertility will reduce to low levels as in these two countries (Italy 2000 TFR 1.21, Poland 2000 TFR 1.48, UN 2006).

FIGURE 8
ESTIMATION OF MEAN AGE AT CHILDBEARING AND POSSIBLE TREND



Source: own estimations.

We assumed that TFR will decline with a linear trend (as suggested by the historical data) and it will reach a low level of 1.4 children per woman. This level is assumed to be reached in 17 years and from then on, TFR will stay constant until the end of the forecast period. In figure 7 we plotted the two expected scenarios, the one assumed by Conapo (2006b) with a high fertility level of 1.8 in the long run (which is high when compared to the European fertility experience) and other low level scenario of a low fertility level of 1.4 in the long run.

We found no assumptions from Conapo for the mean age at childbearing (all parities). Our estimations show a decline, starting from more than 28 years in 1985 to almost 27 years in 2000. The decline has not been linear and sometimes the mean age has also increased as shown in figure 8. However, this behaviour might be due to problems with our data. We expect that Mexican women will keep reducing their fertility, which might imply a further reduction in the mean age at childbearing; but we also expect that Mexican women will start postponing having children, which will increase or at least will held constant this fertility indicator. Furthermore, a usual trend (seen in many countries) is that when fertility declines the mean age of childbearing increases. For simplicity, we decided keep constant this indicator during the forecasting period.

Mortality

We used the number of registered deaths from Vital Statistics and our estimations of Mexican population to perform our own estimation of mortality rates. Mortality rates were estimated for men and women separately, the rates show a slight but continuous decrease over the period 1985-2000 (shown in figure 9). Mortality rates were estimated for five years age groups, the higher open age interval was 85 and more years.

Estimations of mortality rates, from Vital Statistics data, yielded reasonable values except for infant mortality. Table 5 shows a comparison of several estimates of this indicator. It is a well-known problem that infant mortality is under registered in developing countries, so it is highly probable that our calculations underestimate this phenomenon. We adjusted to our infant mortality rates using UN (2006) estimated values as a target. This adjustment is also shown in Table 5. Here we consider that is important to note that more research is needed in order to achieve accurate estimations of Mexican infant mortality and its under-registration.

For the year 2000 we estimated age specific mortality probabilities from the mortality rates. Using Karup-King multipliers we obtained single year estimations. For the old ages, 81 and more years old, we used a linear extrapolation on the log scale. Then we compared our estimates with the ones from Conapo (2003). Estimations from both sources turned out to be very similar (shown in figure 10), the only noticeable differences occur at young ages. Because the strong similarities among estimations we decided to have only one mortality scenario given by our mortality probabilities.

Mortality forecasting assumptions

We need to assume a future mortality trend for the forecasting period. Our starting mortality conditions are given by the age specific mortality probabilities of the year 2000. So we calculated this indicator for all the years of the reference period 1985-2000. From these values we were able to obtain their rates of decline. We decided to assume a future mortality behaviour given by the average of 1985-2000 rates of decline. However, we decided not to apply these rates of decline to all the forecasting period.