**A proposed method for fitting the mortality pattern in Mexico 2000, 2005 and 2010**

Claudio Alberto Dávila Cervantes[1]

**Abstract**
We present a methodology based on Cubic Splines to fit the general mortality curve in Mexico by sex at all ages for the years of 2000, 2005 and 2010. This method consists in fitting a polynomial of degree three in the first four ages (fitting them perfectly); between ages 4 and 80 we fit a cubic spline using five nodes; we fit another third-degree polynomial beginning at age 80. We employ 24 parameters to estimate the mortality curve in the full age range. This model presents an excellent goodness of fit in men, which increases over the years; for women, the goodness of fit is, in general, better than men's. This methodology is considered to be a relatively parsimonious estimation of mortality; it maintains the mortality curve's changes in concavity; and it allows the realization of mortality projections in different contexts and causes of death.

**Key words**: Mortality, Cubic Splines, Mathematical modeling, Mexico, Full age range

[1] Research Professor. Facultad Latinoamericana de Ciencias Sociales (FLACSO)-Mexico – claudio.davila@flacso.edu.mx

## Introduction

During the 20th century significant increases in life expectancy were present among most countries. In Mexico, during that same period, a profound transformation of mortality also happened which was reflected in an increase of life expectancy from a level of 30 years in 1920 to 74 years in 2000; this huge increase occurred rapidly between 1940 and 1960, when the indicator increased almost 19 years in that 20-year period. However, during the next 50 years the rate of increase of life expectancy decreased given that it *only* increased 18 years. The issue regarding the slower growth of life expectancy is to analyze how it can continue to grow; part of the answer lies in understanding the behavior of the mortality pattern and to fit it so we can make a mathematical manipulation of the phenomenon, which also could enable us to make projections and develop research about effects, modifications and possible changes in specific ages. This provokes a growing interest in the description and mathematical fit of mortality patterns.

There are several attempts to model the mortality curve over the entire age range. Among them the Gompertz model (1825) stands out and its variants Gompertz-Makeham and Lazarus. However, today it is known that these mortality laws don't fit correctly the mortality rates in the first years of the life table; and at extreme ages given that the observed rates usually are lower than those predicted by the model and with that the number of survivors at those ages is higher than the estimated with the Gompertz function. Under this framework, the objective of this research is to present a proposal of the method of cubic splines as an alternative to fit the general mortality curve by sex over the entire age range for the years of 2000, 2005 and 2010. We calculated the goodness of fit between the life table's observed and estimated number of people left alive at age $x$ ($l_x$) and the probability of dying between ages $x$ and $x+1$ ($q_x$). We also compare the estimated and the observed life expectancy in order to analyze de quality of the proposed fit.

## Material and methods

*Data*

The data for this investigation comes from the mortality vital statistics from Mexico's National Institute of Statistics and Geography (INEGI) and from the population projections from the National Population Council (CONAPO), for the years of 2000, 2005 and 2010, by sex and individual ages. The mid-year population used to calculate the life tables was obtained from the population projections of CONAPO. We only considered deaths occurred within the country, being excluded: non-specified registries by age and sex; and deaths abroad. Based on this, we discarded 1.21% of the total registries in 2000, 0.88% for 2005 and 0.91% for 2010.

*Methodology*

Is important to explain that a spline is a function $\boldsymbol{f}$ conformed by sections of cubic polynomials with the form $P(x) = ax^3 + bx^2 + cx + d$, which are joined as smoothly as possible (without $\boldsymbol{f}$ necessarily being a unique polynomial) (Barrera et al., 1996). In this study we propose a methodology fitting a polynomial of degree three in the first four ages for the $l_x$ fitting them perfectly due to the change in concavity in the mortality curve caused by the still relatively high infant mortality in the country, as the child adapts to the new environment and gains immunity from the diseases of the outside world (Helligman and Pollard, 1980). Between ages 4 and 80 we fit a cubic spline using five nodes which can vary depending the year fitted. We fit another third-degree polynomial beginning at age 80 to fit the other change of mortality existent in older ages which is related with the deceleration and leveling of mortality in the extreme ages of the life table (Gavrilov and Gavrilova, 2006). To analyze the goodness of fit we used the $R^2$ statistic.

**Results**

We adjusted the number of people left alive at age $x$ and the probability of dying between ages $x$ and $x+1$ for the total number of deaths both for men and women in 2000, 2005 and 2010 using the proposed method. The fit is considered to be good in every year and both sexes given that there is a high similarity between the observed and estimated curve (graph 1). This is numerically corroborated by analyzing the values of $R^2$ (table 1). Results obtained show a goodness of fit close to the maximum possible value of one in every case analyzed which implies a satisfactory fit of mortality in both sexes and every year of study.

**Graph 1 - Comparison of the observed and estimated mortality curves for men (left) and women (right), Mexico 2000, 2005 and 2010.**

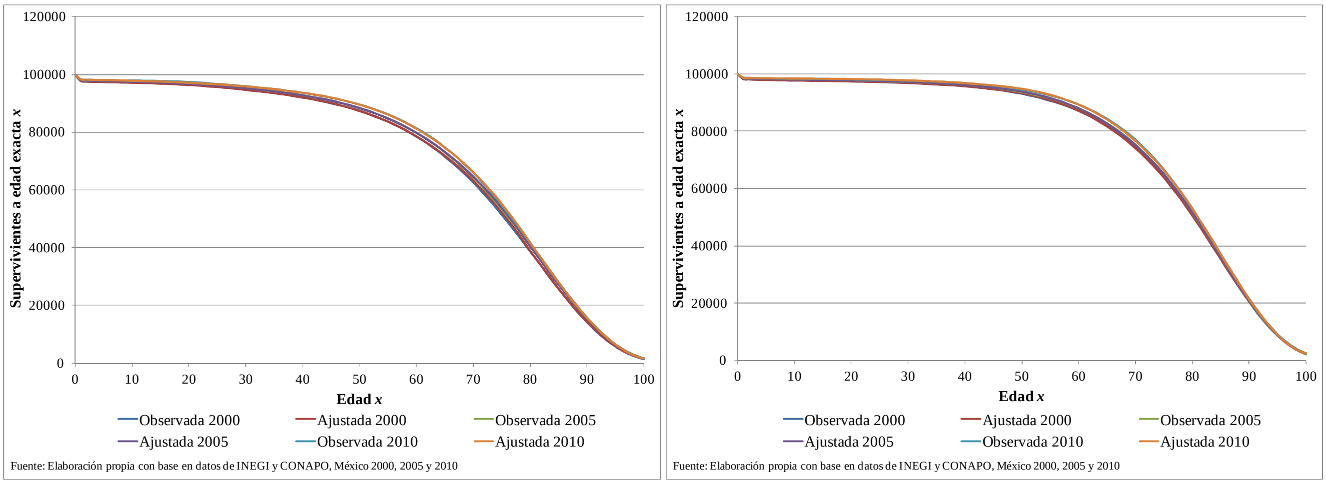


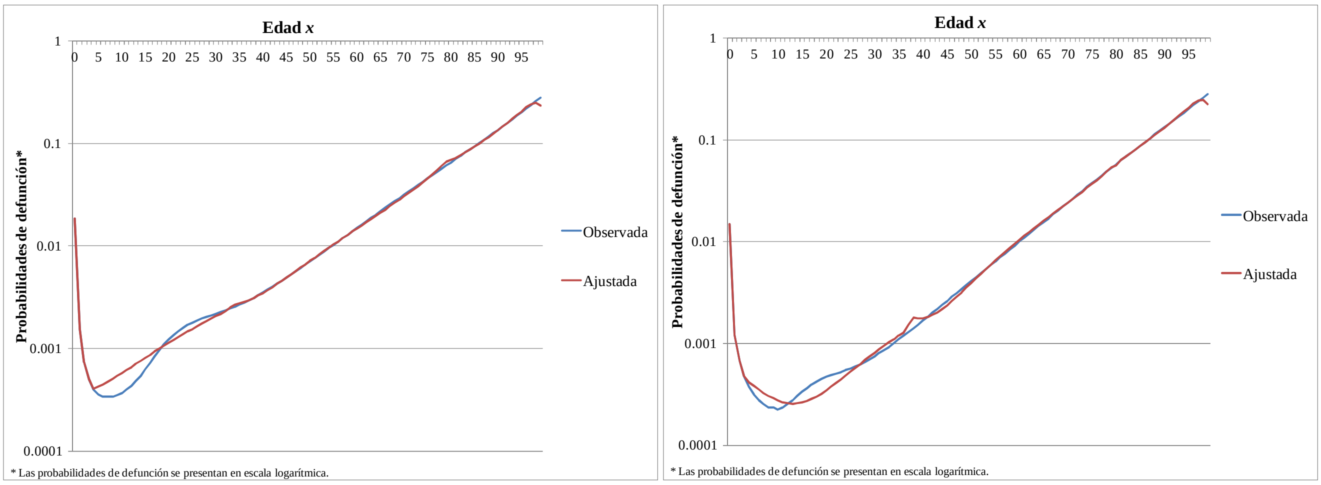

**Table 1 – $R^2$ values for $l_x$ and $q_x$, and estimated life expectancy vs observed, Men and Women. Mexico, 2000, 2005 y 2010**

| | | $l_x$ | $q_x$ | $e_0$ official | $e_0$ estimated |
|---|---|---|---|---|---|
| **Men** | **2000** | 0.9999548 | 0.9916822 | 71.3040 | 71.3685 |
| | **2005** | 0.9999695 | 0.9946900 | 72.2126 | 72.2500 |
| | **2010** | 0.9999816 | 0.9947044 | 73.0829 | 73.0776 |
| **Women** | **2000** | 0.9999936 | 0.9936219 | 76.4716 | 76.4785 |
| | **2005** | 0.9999912 | 0.9921477 | 77.0482 | 77.0570 |
| | **2010** | 0.9999807 | 0.9921435 | 77.7962 | 77.8083 |

Source: Own elaboration based on the mortality vital statistics in Mexico. INEGI

If we analyze these results in more detail, the value of the $R^2$ for men in every year is above a value of 0.9999 and lightly increases each year giving a better fit in 2005 and an even better in 2010. For women, the fit is better than men's in 2000 and 2005, but not in 2010. Another characteristic of the fit for women is that the best fit occurs in 2000 and the value of $R^2$ decreases each passing year (though the fit is still good). We also analyzed the goodness of fit of the proposed estimation using the $q_x$ in order to corroborate if the graduation formula maintains the concavity of the observed data over the entire age range and to study if the fit is adequate for another life table series and to see if it represents correctly the changes in concavity of the curve. The fit of the $q_x$ was satisfactory especially starting from the second node in every case. At young ages the change of concavity present during childhood and adolescence is well fitted (as an example of this we present in graph 2 the comparison between the estimated and observed $q_x$ in 2005). We note that we have a lower goodness of fit than with $l_x$, but the estimations can be still considered a good one (table 1). For men we obtained values of $R^2$ superior to 0.99 and consistently gets better every year; women also have a high $R^2$ being 2000 the best fit. Something worth noting is that in both $l_x$ and $q_x$ the fit for women is better in 2000, but in 2005 and 2010 the estimation for men had a higher $R^2$.

**Graph 2 - Observed and estimated $q_x$ for men (left) and women (right). Mexico 2005.**




Source: Own elaboration based on the mortality vital statistics in Mexico. INEGI

Regarding life expectancy estimated by this methodology we calculated values of 71.37 and 76.47 in 2000 for men and women respectively, which are very similar to the official values published by CONAPO (table 1); for 2005 we have a life expectancy of 72.25 and 77.05 for men and women respectively, also similar to the official values; and in 2010 these values are 73.08 for men and 77.81for women. These results indicate that the proposed methodology can be used reliably to perform estimations of life expectancy in this and possibly in other contexts. Another important result is the location of the nodal ages for men and women which can be used as a tool to analyze mortality changes over time. An increase in the nodal age can be related to a decrease in the impact of mortality. In this study however, we used the same nodes for every year because the time of analysis was not enough to indicate a significant shift in mortality levels. The first node is located at age 4 in all cases; the second is situated in similar adult ages for men (33) and women (37); the third is also located in adult ages (38 men and 43 women); the fourth is situated in different ages for both sexes 50 for men and 61 for women, this is due to the lower mortality for the latter in these group ages; the final node is located at age 80 in all cases.

## Discussion

Mortality in Mexico has decreased constantly since 1940 until today and life expectancy has increased continuously throughout that period. This has presented at different rates and it reflects that in recent years a decrease in the rate of life expectancy gains has occurred. Mortality studies and specifically the estimation of the parameterized pattern of mortality that mathematically describes the age-sex structure of mortality and its changes through time could all help to have a deeper knowledge about the composition and levels of life expectancy. Therefore these kind of researches are important because they allow us to mathematically manipulate the phenomenon (Halli and Rao, 1992). Finding the underlying patterns that rule demographic phenomenon in order to describe them mathematically allows us to propose plausible, probable or desirable future scenarios if a set of specific conditions occurred and based on them we could recommend adequate actions to try to decrease the impact of mortality in certain ages. Models like the one proposed here are also useful to develop research about specific effects in precise ages, about changes in demographic determinants and the impact those would have on the trend and impact of mortality within a certain population (Halli and Rao, 1992), which in turn could provide input to establish public policies or public health campaigns aimed to mitigate or delay mortality, as well as in mortality projections.

We proposed here a methodology that involves a third degree polynomial in the first four ages and in the last ages of the life table and a 5 node spline fitting ages 4 to 80. The main reaches this function has are a more than satisfactory goodness of fit and therefore an excellent mathematical description of the mortality pattern. We believe then that the proposed function is adequate given that takes into account the intrinsic properties of mortality curves such as: the change of concavity presented during early childhood mainly due to the still relatively high infant mortality by fitting this characteristic perfectly with a third degree polynomial; the descending mortality rate after early childhood ages where the minimum probability of dying is achieved, fitted with the first polynomial of the cubic spline; the significant rise in probabilities of dying during adolescence and young adulthood caused by accidents and violent deaths and the exponential growth of mortality rates during adult ages, which are well described achieved by fitting the series of cubic polynomials from the spline (Congdon, 1993); and the deceleration of mortality present at extreme ages of the life table through the use of another third degree polynomial that respects the change in concavity also found at those ages (Gavrilov and Gavrilova, 2002).

We conclude then that the methodology using splines fits correctly the age pattern of mortality, describing correctly these aforementioned characteristics of the mortality curve. Other main advantage of this model is that not only it gives an excellent goodness of fit of mortality data but also it only requires the knowledge of matrix algebra and the calculation of life tables in order to implement it. Also, the methodology is believed to be successful because every polynomial is joined smoothly, just as Helligman and Pollard (1980) state “a mortality graduation can only be considered successful if the graduated rates progress smoothly from age to age and at the same time they reflect accurately the underlying mortality pattern”. Finally, we believe that the model’s major advantage lies in the fact that it is a flexible model given that we used polynomials which allows us to fit mortality in other contexts, in mortality projections, in different years, and also can be used to fit the patterns of mortality by different causes (Dávila, 2012).

*Limitations*

This proposed methodology has some disadvantages. Even though the curve is continuous and applicable over the entire age range it has a relatively high number parameters due mainly to the fact that for each polynomial (within the spline or not) we had to estimate four parameters, which gives us a total of 24 parameters, making the function a relatively little parsimonious model. Another disadvantage lies in the selection of the nodes which is done manually and does not give the opportunity to generalize for different years, by sex or in the case of fitting a pattern of mortality by cause. Another limitation is related with the quality of data used here originating from the vital statistics in Mexico, especially in older ages. These data has certain deficiencies such as: coverage problems particularly in rural areas due to an inadequate coverage of the Civil Registry in poorly accessible regions; underreporting of infant and maternal deaths (Camposortega, 1992) and a long bureaucratic process which provokes a delay in the publication of mortality data (Hernández and Narro, 2010). Because of this there could be some uncertainty about the real level of mortality in extreme ages, which could also lead us not to fit the *real* level of mortality.