Consistent Projection of Mortality for Subnational Areas Using a Logistic Curve

by

Tom McDevitt
Population Studies Branch
International Programs Center
Population Division
U.S. Census Bureau

Extended abstract prepared for a paper to be presented at the 2009 annual meetings of the Population Association of America, April 30 – May 2, Detroit.

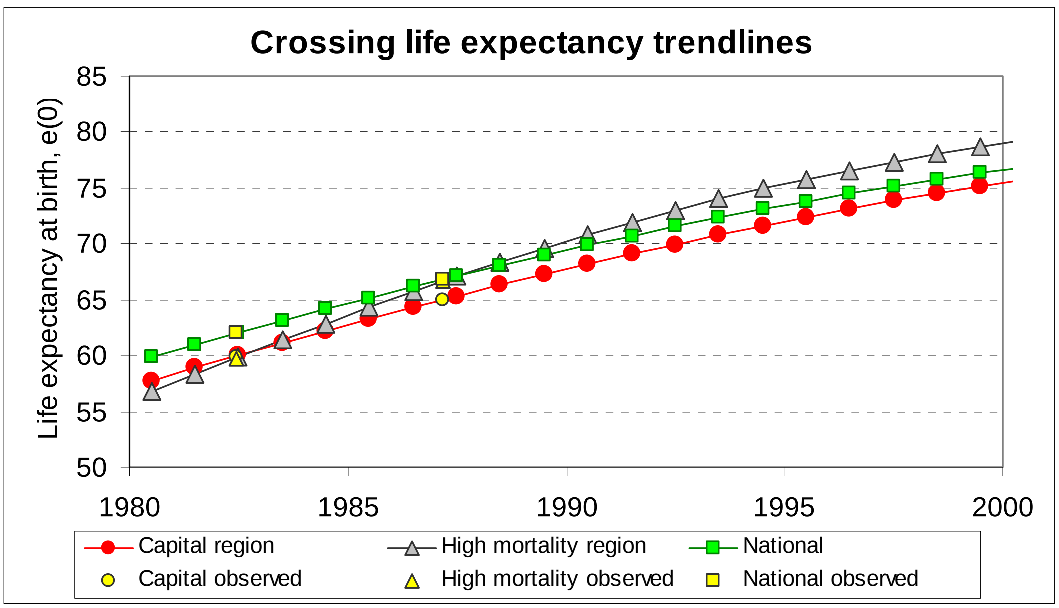
USCENSUSBUREAU

Consistent Projection of Mortality for Subnational Areas Using a Logistic Curve

Extended Abstract by Tom McDevitt, Population Studies Branch, International Programs Center, Population Division

Ensuring consistency in projected future mortality between subnational and national populations is one of the challenges involved in making subnational projections. This paper discusses the use of linear and logistic extrapolation of life expectancy at birth (e(0)), and the use of two methods based on the logistic, for ensuring consistency in projected mortality across subnational areas and between subnational and national projected mortality. Reported under-5 mortality from Demographic and Health Surveys (DHSs) is used to estimate e(0) at three points in time, and estimated regional e(0)s for the third time period are used to assess the relative goodness of fit of projected regional e(0)s based on estimates from the two prior time points. Comparisons are made using alternative projection models for regions in 12 developing countries not seriously affected by HIV/AIDS. A shared-slope logistic model performs better than alternatives considered.

The impetus for the paper stems from observations that improvements in life expectancy at birth have slowed in many countries during the second half of the Twentieth Century, as improved sanitation, the introduction of new drugs, and a variety of public health initiatives used to combat a number of infectious diseases have not been matched by comparable reductions in cardiovascular disease, diabetes and other obesity-related illness, and other diseases less susceptible to public health measures. While the debate among demographers about the nature of human longevity continues (NRC 2000, Bongaarts 2008), improvements in life expectancy in many developing countries are often slowing. From a practical standpoint, projection of life expectancy at both national and subnational levels benefits from an assumed upper limit and an asymptotic approach to that limit. The logistic curve, which accommodates an asymptotic increase in e(0) provides a much better model for fitting empirical e(0) observations than a model assuming linear increase.[1] Still, the seeming appropriateness of the logistic at the national



[1] For discussions of the use of the logistic curve in the projection of population or its components of change, see Pearl and Reed (1920), Arriaga (1984, 1994).

level provides little guidance for modeling mortality decline at the subnational level, where the pace of decline may vary widely across areas and unconstrained projection of region-specific e(0)s may lead to unlikely outcomes (Figure).

Two questions are addressed in the paper:

1. Does the logistic curve improve predictive accuracy for assumed e(0) at the subnational level for developing countries since the late 1980s?

2. Is predictive ability of the logistic model further improved by using variants of the logistic that force greater consistency between assumed national and subnational trends in e(0)?

The paper seeks to answer these questions by assessing predictive accuracy of projected e(0) using alternative mortality projection models for subnational areas in twelve countries with three or more DHSs. Predictive accuracy for regional (subnational) e(0) is operationalized as mean absolute percent error (MAPE) between the 10-year average regional e(0) from the latest DHS and the trendline-based estimate of e(0) for the same date where the trendline is fitted to estimated e(0)s from the first two DHSs.

The first part of the paper describes a series of models for projecting e(0) for subnational areas. These are:

- Linear extrapolation
- Logistic fitting and extrapolation
- Constrained logistic fitting and extrapolation
- Fixed-slope logistic fitting and extrapolation
- Shared-slope fitting and extrapolation

The procedure for deriving each e(0) is in two parts. First, gender-specific infant mortality (${}_1q_0$) and child mortality (${}_4q_1$) are derived for subnational areas (regions) for the 10-year period preceding each survey using national-level gender-specific patterns of infant and under-5 mortality and both-sexes infant and under-5 mortality reported for each region. Second, gender-specific infant and child mortality are used to estimate an abridged life table assuming West regional model mortality for each of the three surveys.[2]

A curve fitted to the e(0)s from the first two surveys is used to predict e(0) for the reference period of the third survey. MAPE is used to quantify the difference between predicted and observed e(0), and to compare performance across extrapolation methods and countries.

The second part of the paper reports the results of repeating the procedure using an age pattern of mortality that may be more appropriate to each country than West regional mortality. This second pattern is the mortality age pattern adopted by the U.S. Census Bureau for its estimates and projections work for each country. That is, the sensitivity of the results of the exercise based on West regional model mortality is assessed using age-sex-specific mortality for each country

[2] Coale and Demeny (1968).

and reference year to convert infant and child mortality into an abridged life table, and e(0), for each region of each country.

The third part of the paper summarizes the results of repeating the part 2 test using each of the DHS surveys after the first two in order to better determine sensitivity of findings to length of projection period.

Countries used in the analysis were selected because (1) at least three DHSs with infant and child mortality estimates are available to generate e(0)s for fitting and comparison, (2) the subnational regions for these countries are comparable, and (3) the countries are not among the group considered seriously affected by HIV/AIDS. The third criterion is important because e(0) trend in HIV/AIDS countries tends not to be monotonically rising and, equally important, regional mortality changes in these countries need not follow a common trend since some regions may be seriously affected and others may be little affected by the epidemic. Countries used are shown in Table 1.

Table 1. Demographic and Health Surveys Used in the Analysis

Country	Surveys used for fitting		Parts 1,2 testing	Part 3 testing
Sub-Saharan Africa				
Madagascar	1992	1997	2003/04	
Senegal	1986	1992/93	2005	1997, 1999
Middle East/North Africa				
Egypt	1988	1992	2005	1995, 2000
Morocco	1987	1992	2003/04	
Asia				
Bangladesh	1993/94	1996/97	2004	1999/2000
Indonesia	1987	1991	2002/03	1994, 1997
Nepal	1996	2001	2006	
Philippines	1993	1998	2003	
Latin America & The Caribbean				
Bolivia	1989	1994	2003	1998
Colombia*	1986	1990	2000	1995
Dominican Rep.*	1986	1991	1996	
Peru	1986	1992	2000	1996

* For some countries the latest DHS could not be used because the geography used to tabulate results differs from that defined for earlier surveys. This is the case for the 2005 DHS for Colombia and the 1999 and 2002 DHSs for Dominican Republic.

The exercise is expected to show that the combination of shared-slope fitting and extrapolation provides better forecast accuracy than the other methods for projecting mortality at the subnational level for developing countries.

References

Arriaga, Eduardo. 1984. "Variaciones sobre un tema de la function logistica," ***Notas de Poblacion***, San Jose, Costa Rica: CELADE, Year 12, No. 36: 85-97.

_____. 1994. ***Population Analysis with Microcomputers***. With Peter D. Johnson and Ellen Jamison. Washington, D.C.: U.S. Census Bureau, Vol. I: ch. VIII.

Bongaarts, John. 2008. "Trends in Senescent Life Expectancy." Draft paper prepared for an expert group meeting on adult mortality estimation, United Nations Population Division, July, photocopy.

Coale, A. and P. Demeny. 1968. ***Regional Model Life Tables and Stable Populations***. Princeton, New Jersey: Princeton University Press.

National Research Council (NRC). 2000. ***Beyond Six Billion***. J. Bongaarts and R. Bulatao (eds.). Washington, DC: National Research Council.

Pearl, R. and L. Reed. 1920. "On the rate of growth of the population of the United States since 1870 and its mathematical representation," ***Proceedings of the National Academy of Sciences***, Vol. 6: 275-288.

Appendix. Logistic Models

Each of the logistic model variants used in the paper is described here. The model descriptions that follow avoid reference to sex because, with the exception of the third variant, the same models may be fitted to male, female, or both-sexes e(0)s. All fitting and testing for the paper has been sex-specific. .

The first logistic variant expresses predicted e(0) as a function of defined lower and upper asymptotes and the estimated slope and intercept of an ordinary least squares regression line fitted to the logits of observed, dated e(0)s.

$$\hat{e}(0) = \left(\frac{e^{a+bt}}{1+e^{a+bt}} \right) * \left(UB - LB \right) + LB$$

where UB is the upper asymptote for the logistic function
LB is the lower asymptote for the logistic function
b is the slope of a line fitted to the logit transformations of observed e(0)s
a is the intercept of a line fitted to the logit transformations of observed e(0)s
t is time in years
and the logit of each observed e(0) is

$$\ln\left[\frac{(e(0)-LB)/(UB-LB)}{1-(e(0)-LB)/(UB-LB)} \right]$$

This variant is used to extrapolate e(0) for both national population and subnational (regional) population.

The second variant, referred to in the paper as the "constrained" logistic, is defined in terms of the complement of the e(0); that is, of the difference between e(0) in year t and the lower bound of the logistic. The specification ensures that proportionate changes in regional e(0) track proportionate changes in (extrapolated) national e(0) and, therefore, that regional e(0) tracks (extrapolated) national e(0). In most cases, the national e(0) trend used to guide the constrained logistic fitting for regions is the unconstrained logistic described above (first logistic variant). The constrained logistic is written:

$$\hat{e}(0)_{t,R} = UB - (UB - \hat{e}(0)_{t-1,R}) * \frac{(UB - e(0)_{t,N})}{(UB - e(0)_{t-1,N})}$$

where $e(0)_{t,N}$ is the extrapolated e(0) for country N in time t
$\hat{e}(0)_{t,R}$ is the predicted e(0) for region R in time t
$\hat{e}(0)_{t-1,R}$ is the predicted e(0) for region R in time t-1. The starting value for this term is the estimated e(0) from the second DHS.

This specification is equivalent to

$$\left(\frac{UB-\hat{e}(0)_{t,R}}{UB-\hat{e}(0)_{t-1,R}}\right)=\left(\frac{UB-e(0)_{t,N}}{UB-e(0)_{t-1,N}}\right)$$

The third variant, referred to here as the "fixed sloped" model, projects sex-specific e(0) for a population using a pair of equations fitted to a series of gender-specific life table e(0)s for a large number of countries at the Census Bureau in 2001. This specification has the same functional form as the first variant but with gender-specific slopes taken from the fitted cross-national e(0) equations. That is,

$$\hat{e}(0)_{t,R}=\left(\frac{e^{a+Ft}}{1+e^{a+Ft}}\right)*(UB-LB)+LB$$

where F is the slope of a line fitted to the logit transformations of cross-national e(0)s. F is +0.0258 for males, +0.0271 for females

a is the intercept of a line fitted to the logit transformations of regional e(0)s

This specification ensures that regional e(0) trend for a specific country is consistent with cross-national e(0) trend (actually, change in the logit transformation of e(0)) in every period. This specification does not allow for the possibility that national e(0) will not follow a logistic trend.

The fourth variant, referred to here as the "shared-slope" variant, projects e(0) for region R using the same functional form as the first variant but with the slope taken from the fitted equation from the national e(0) trendline. That is,

$$\hat{e}(0)_{t,R}=\left(\frac{e^{a+Bt}}{1+e^{a+Bt}}\right)*(UB-LB)+LB$$

where B is the slope of a line fitted to the logit transformations of national e(0)s

a is the intercept of a line fitted to the logit transformations of *regional* e(0)s

This specification ensures that regional e(0) rise is not only consistent with, but is parallel to, national e(0) improvement (again, defined in terms of the logit of e(0)). This specification also does not allow for the possibility that national e(0) will not follow a logistic trend.

For each logistic variant, and for the linear extrapolation, the upper asymptote for each series of extrapolations has been set equal to 82.56 years for males and 88.40 years for females. The lower bound for each logistic has been set equal to 25 years. These are the same upper and lower bounds used by the U.S. Census Bureau in fitting logistic curves to empirical e(0)s for developing countries.