



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Multistate life tables using Stata

Jerônimo Oliveira Muniz
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil
jeronimo@ufmg.br

Abstract. The `mslt` command calculates the functions of a multistate life table and plots a graph of conditional and unconditional life expectancies by time. The command provides linear and exponential solutions to estimate the number of individuals, transitions, probabilities, person-years, and years of life in a given cohort and state of occupancy. The input data are time-specific transition rates (or survivorship proportions) between nonabsorbing and at most one absorbing state. In addition to the mean age at transfer between states, `mslt` calculates the following summary measures: the mean age, the probability of dying, the average duration, and the proportion of life spent in a specific state.

Keywords: st0615, mslt, age, demography, increment–decrement, life expectancy, life table, model, multigroup, multistate, population, probability, proportion, rate, transition

1 Introduction

Multistate, or increment–decrement, life tables are a demographic tool to calculate the amount of time that individuals of a cohort will spend in a given *state of occupancy*, broadly defined by the presence of a categorical attribute. In comparison with single-state life tables, a unique feature of multistate tables is to decompose the overall life expectancy into segments of life span by different states, which make the analytical approach of multistate life tables applicable to many areas (Ledent and Zeng 2010). In the sociology of the family, examples of valid states for investigation are single, married, divorced, or widowed. In social stratification, one could examine the odds and the time spent in the working, middle, and upper classes. In epidemiology, the focus could be on noninfected versus infected individuals or experiencing no disability versus being disabled. In labor economics: employed, unemployed, or retired. In geography: residing in rural or urban areas. In political science: being a Democrat, a Liberal, or a Republican or having no political affiliation. Multistate life tables allow the calculation of how many years one could expect to spend in each one of these categories and provide other informative functions to understand social processes of change and composition of groups. In all these examples, an additional and ultimate absorbing state (which once entered cannot be left) could be included: death, but also unrecoverable illnesses, or physical conditions. Differently from single life tables, where death is typically the only source of exit, in multistate life tables, increments and decrements from more than one state are allowed, making it a rich analytical framework to model flows under the influence of concurrent risks.

Multistate models answer the following questions: How long could one expect to live in a given condition? What is the probability of moving from one condition to another? What is the probability of dying in a given state? At what time of exposure (that is, ages) are transitions between states of interest most likely to occur? How long can one born in a given condition expect to live in alternative states?

Once transition rates—or survivorship proportions—by time¹ and between states are entered in Stata, `mslt` calculates unconditional and conditional life expectancies and optionally provides the proportion of life, the mean age of persons, the average duration, the probabilities of transition, and the mean lifetime at transfer between states. The goal of this article is to introduce and illustrate the use of `mslt`, a Mata-based command to calculate functions and key summary measures of multistate life tables.

2 Background

Increment–decrement life tables have a long history in demography. The oldest application dates from the early 20th century, when DuPasquier (1912) investigated transitions between two conditions: healthy and disabled. Since then, much work has been done to advance and implement multistate models (Rogers 1973, 1975; Ledent 1980; Schoen 1988a,b; Palloni 2001), with applications to marital status (Schoen and Land 1979), labor force participation (Smith 1982; Willekens 1980), interregional migration (Rogers 1975), population projections by educational attainment (Lutz and Goujon 2001), active and disabled (Land, Guralnik, and Blazer 1994), and happy life expectancies (Yang and Walij 2010). In addition, a plethora of computer programs have been written to estimate multistate life tables (Willekens and Putter 2014), but most of them are not “flexible enough to handle a very broad class of applications or to implement alternative solutions” (Palloni 2001, 272). The `mslt` program was developed having this aim in mind.

1. Classical demographic theory uses categorical age groups to investigate the distribution of the population in different states. In many cases, however, the size of organisms could be more important than age to predict the response of an individual to its environment. For many species, chronological age tells little about its demographic properties, and the age distribution tells little or nothing about the behavior of the population (Lefkovitch 1965; Usher 1966; Caswell 2001). In ecology, for instance, size is more relevant to the species’s demographic fate than age. Trees (Zon 1915), crabs (Somerton and MacIntosh 1983), and fish (Alm 1959) must reach a threshold size before beginning to reproduce. Trees (Jiménez, Lugo, and Cintrón 1985), corals (Hughes and Jackson 1985), and marine invertebrates (Jackson 1985) have size-dependent mortality. Forthcoming notation and examples in this article assume age-specific dependency because this is the most relevant predictor of reproduction in human populations. Note, however, that factors or states other than age (sex, marital status, location, political and religious affiliation) also affect vital rates and the dynamics of populations.

2.1 Required data for existing methods

The input to calculate the functions of multistate life tables is a set of transition rates² or survivorship proportions for age-specific groups. Formally, transition rates from state i to j between exact ages x and $x + n$ are expressed as

$${}_nM_x^{ij} = \frac{{}_nD_x^{ij}}{{}_nP_x^i} \quad (1)$$

where x , the lower right subscript, refers to the exact age at the beginning of each age interval and n , the lower left subscript, is the length of each age group, usually equal to one or five years in the demographic literature. To present and operationalize the data, one divides the continuous variable age—including decimal or fractional years—into a discrete set of categorical groups starting at exact times or ages x and ending after n years, the elapsed time in each interval. ${}_nD_x^{ij}$ is the observed number of transfers between states i and j , from exact age x to the instant before attaining exact age $x + n$.³ ${}_nP_x^i$ comprises the midperiod population in state i between these same ages. So ${}_1M_{20}^{ij}$, the transition rate from state i to j between ages 20 and 21, refers to events occurring to persons who are aged 20.0000 to 20.9999 in exact years since birth; and ${}_5M_{20}^{ij}$ would include the events and the population of individuals who have completed 20, 21, 22, 23, and 24 years of age since birth. Although the length of n depends on the type of data available to the analyst, `mslt` provides most precise results with single-year data ($n = 1$). If such data are not available, the recommendation is to smooth the multiyear data using splines, interpolation, or any other indirect method before entering the data into Stata.

Survivorship proportions, on the other hand, are defined as

$${}_nS_x^{ij} = \frac{{}_nP_{x+n}^{ij}(t+n)}{{}_nP_x^i(t)} \quad (2)$$

where ${}_nP_x^i(t)$ is the number of persons in state i of the observed population between the ages of x and $x + n$ at time t , and ${}_nP_{x+n}^{ij}(t+n)$ is the number of persons in state j of the observed population between the ages of $x + n$ and $x + 2n$ at time $t + n$ who were in state i exactly n years earlier (Schoen 1988a). The key difference between (2) and (1) is that survivorship proportions rely on the same group of individuals, while rates may not.

2. Some authors prefer the term “exposure” or “occurrence rates”.

3. In the last open-ended (or terminal) age interval, n is equal to infinity, so in such cases notation should reflect this specificity [that is, (11), in which $n = \infty$].

Four methods—linear, mean duration at transfer, exponential, and cubic⁴—solve for the quantities of multistate life tables when transition rates are available. Alternatively, when data are entered as survivorship proportions and must be converted to rates, only two strategies—linear or exponential—are available (Schoen 1988a). `mslt` handles data of both types (rates and proportions) to provide linear and exponential solutions for multistate life tables. The other two alternatives (mean duration at transfer and cubic solutions) have not yet been incorporated into the program. Research has shown, however, that these methods “produce very similar values for single-year age intervals” (Schoen 1975, 1988a).

3 The mslt algorithms

This section presents formulas to estimate the functions and key summary measures of multistate life tables. Because of its technical content, readers not familiar with demographic notations may find it challenging to follow and may skip to the next section for an intuitive illustration of the method and straightforward interpretations of results. However, those who want to have a keen grasp of what `mslt` is assuming and doing are encouraged to read this section and to look further into its references.

The essential quantities of multistate life tables are the number of individuals in state i at exact ages x and $x + n$, l_x^i , and l_{x+n}^i . Under the assumption that l_x^i is linear (increasing risk of the event within each age interval), the number of individuals at age $x + n$, l_{x+n}^i , is calculated by solving the matrix equation originally suggested by Schoen (1975) and Rogers and Ledent (1976) and later referenced by Schoen (1988a, 70) and Palloni (2001, 269),

$$\mathbf{l}(x + n) = \mathbf{l}(x) \left\{ \mathbf{I} - \frac{n}{2} \mathbf{M}(x, n) \right\} \left\{ \mathbf{I} + \frac{n}{2} \mathbf{M}(x, n) \right\}^{-1} \quad (3)$$

where $\mathbf{l}(x + n)$ and $\mathbf{l}(x)$ are row vectors containing the numbers of individuals in states of interest, \mathbf{I} is the identity matrix, n represents, as in (1) and (2), the length of age intervals, $\mathbf{M}(x, n)$ are square matrices containing transition rates from state i and j and between exact ages x and $x + n$, and the superscript -1 indicates the matrix inverse. Each age group—defined by and between ages x and $x + n$ —has a set of age-specific transition rates. Each age group between the first and the last age x has its own matrix $\mathbf{M}(x, n)$. The integer values assumed by x and n (usually 1 or 5) may vary according to the analyst’s dataset and are thus left unconstrained.

4. The difference between these methods lies in the assumption about the force of events within age groups. These assumptions are required because age, a naturally continuous variable, is grouped into discrete groups of age. Therefore, the behavior of transition rates within discrete age groups must be assumed to obtain the functions of multistate life tables. The linear method of calculation assumes linearity on the gross flows of persons from states i to j . This implies that the underlying risk of the event is increasing with age. The mean duration of transfer method assumes quadratic gross flows. The exponential method assumes approximately constant risks of the event within age intervals. This implies that the number of individuals at each period of time are nonlinear (exponential) functions of age. And the cubic method assumes a cubic polynomial to calculate the number of person-years in each state i . Further comparisons and details about these methods are in Schoen (1988a, 70–76).

In particular, in the presence of an absorbing state,⁵ the structure of the $(k + 1)$ by $(k + 1)$ matrix of observed transition rates from state i to state j between ages x and $x + n$ is

$$\mathbf{M}(x, n) = \begin{pmatrix} \sum_n M_x^{1,j} & -_n M_x^{1,2} & \cdots & -_n M_x^{1,k} & -_n M_x^{1,k+1} \\ -_n M_x^{2,1} & \sum_n M_x^{2,j} & \cdots & -_n M_x^{2,k} & -_n M_x^{2,k+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -_n M_x^{k,1} & -_n M_x^{k,2} & \cdots & \sum_n M_x^{k,j} & -_n M_x^{k,k+1} \\ 0 & 0 & & 0 & 0 \end{pmatrix} \quad (4)$$

where the summations over j run from 1 to $k + 1$ possible destinations for a person in one of the k living states, excluding the case where $j = i$.

Notice that each row in (4) sums to 0 and that the diagonal elements are a sum of the transition rates in the same row. In the absence of an absorbing state, the elements of the last row and column of matrix $\mathbf{M}(x, n)$ should be excluded. Alternatively, according to Rogers and Ledent (1976) and Ledent and Zeng (2010), matrix $\mathbf{M}(x, n)$ could be transposed and have death rates included in the main diagonal instead of in the last row and column.

Alternatively, under the assumption that rates are **constant** in age intervals, the solution, first used by DuPasquier (1912) and later applied by Krishnamoorthy (1979) to investigate marital statuses, can be expressed as

$$\mathbf{l}(x + n) = \mathbf{l}(x) \exp \{-n\mathbf{M}(x, n)\} \quad (5)$$

where

$$\exp \{-n\mathbf{M}(x, n)\} = \mathbf{I} - n\mathbf{M}(x, n) + \frac{n^2}{2!}\mathbf{M}^2(x, n) - \frac{n^3}{3!}\mathbf{M}^3(x, n) \dots \quad (6)$$

If the rates in $\mathbf{M}(x, n)$ are small, the values on the left of (6) will usually stabilize after the second or third element in the infinite power series.⁶

If instead of rates one enters survivorship proportions $_n S_x^{ij}$, the option **proportion**, in the `mslt` command, must be requested to convert such values to rates before calculating the remaining functions of the multistate table. Under the linear assumption, following Schoen (1988a) and Rees and Wilson (1977), this conversion is done by solving two matricial equations,

$$\mathbf{M}(x + n) = \frac{2}{n} \{\mathbf{I} + \mathbf{\Pi}(x, n)\}^{-1} \{\mathbf{I} - \mathbf{\Pi}(x, n)\}$$

5. Absorbing states are conditions that once entered cannot be left. They have only entrance flows, such as educational achievement, death, unrecoverable illnesses, and nonsingle marital status. Once individuals enter into these conditions they cannot go back to their previous states.

6. Age-specific convergence is achieved when the maximum relative difference (see `help mreldif()`) between positive and negative terms on the right side of (6) is less than or equal to $1e-10$.

in which

$$\mathbf{\Pi}(x, n) \cong \frac{1}{2} \{ \mathbf{S}(x, n) + \mathbf{S}(x - n, n) \}$$

where $\mathbf{S}(x, n)$ has element ${}_nS_x^{ij}$ in the i th row and j th column, with restricted values between 0 and 1. The structure of square matrix $\mathbf{\Pi}(x, n)$ is not the same as $\mathbf{M}(x, n)$, so **mslt** invokes different Mata functions when the option **proportion** is inserted in the command line. In particular, in the presence of an absorbing state, the $(k+1)$ by $(k+1)$ square matrix of observed transition probabilities is

$$\mathbf{\Pi}(x, n) = \begin{pmatrix} {}_n\pi_x^{11} & {}_n\pi_x^{12} & \dots & \pi^{1,k+1} \\ {}_n\pi_x^{21} & {}_n\pi_x^{22} & \dots & \pi^{2,k+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & 1 \end{pmatrix} \quad (7)$$

Each row in (7) sums to 1, and in the absence of an absorbing state, the elements in the last row and column should be excluded. As an alternative compact form, if one added the probabilities of moving to the absorbing state to the main diagonal, matrix $\mathbf{\Pi}(x, n)$ could have as its elements interstate transition probabilities arranged in transposed order (Rogers and Ledent 1976).

Under the **constant** risks assumption, the conversion of probabilities to rates is approximate by the matrix logarithm solution offered by Hall (2015):

$$\begin{aligned} \mathbf{M}(x, n) &= -\frac{1}{n} \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\{ \mathbf{\Pi}(x, n) - \mathbf{I} \}^k}{k} \\ &= -\frac{1}{n} \left[\{ \mathbf{\Pi}(x, n) - \mathbf{I} \} - \frac{\{ \mathbf{\Pi}(x, n) - \mathbf{I} \}^2}{2} + \frac{\{ \mathbf{\Pi}(x, n) - \mathbf{I} \}^3}{3} \dots \right] \end{aligned}$$

Once matrices $\mathbf{M}(x, n)$ and vectors $\mathbf{l}(x)$ are obtained through linear or exponential solutions, the other functions of the multistate life table are straightforward. Continuing with matrix notation, under the linear assumption and following (3), we give the number of transitions from state i to j between ages x and $x+n$ by the off-diagonal elements of

$$\begin{aligned} \mathbf{D}_1(x, n) &= \text{diag} \{ \mathbf{l}(x) \} \left\{ \mathbf{I} - \frac{n}{2} \mathbf{M}(x, n) \right\} \left\{ \mathbf{I} + \frac{n}{2} \mathbf{M}(x, n) \right\}^{-1} \\ &= \begin{pmatrix} l_{x+n}^{11} & {}_nd_x^{12} & \dots & {}_nd_x^{1,k+1} \\ {}_nd_x^{21} & l_{x+n}^{22} & \dots & {}_nd_x^{2,k+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & l_{x+n}^{k+1,k+1} \end{pmatrix} \end{aligned} \quad (8)$$

where the diagonal elements l_{x+n}^{ii} represent the number of individuals in the cohort who are in state i at age $x+n$ after accounting for the outflows from state i to j presented in the same row (but without considering the inflows from j to i).⁷ Under the **constant** forces assumption, following a slight variation of (5), the analogous matrix of $\mathbf{D}_1(x, n)$ is

$$\mathbf{D}_c(x, n) = \text{diag} \{ \mathbf{l}(x) \} \exp \{ -n\mathbf{M}(x, n) \} \quad (9)$$

which, in longhand form, has elements structured in the same way as in $\mathbf{D}_1(x, n)$. However, the value of the elements in $\mathbf{D}_c(x, n)$ differ to reflect the assumption of constant risks of the event within age intervals.

In the absence of an absorbing state, matrices $\mathbf{D}_1(x, n)$ and $\mathbf{D}_c(x, n)$ do not have their last rows and columns. After executing (8) or (9) for all ages, **mslt** removes the main diagonal and columns with zero decrements (that is, in which $i = j$ or in which i is an absorbing state) and provides a rearranged array $\mathbf{d}(x, n)$ for the number of transitions from state i to state j between ages x and $x + n$. This matrix is saved in Mata as

$$\mathbf{d}(x, n) = \begin{pmatrix} {}_n d_x^{ij} & \cdots & {}_n d_x^{k, k+1} \\ {}_n d_{x+n}^{ij} & \cdots & {}_n d_{x+n}^{k, k+1} \\ {}_n d_{x+2n}^{ij} & \cdots & {}_n d_{x+2n}^{k, k+1} \\ \vdots & & \vdots \\ {}_n d_{x+\lambda n}^{ij} & \cdots & {}_n d_{x+\lambda n}^{k, k+1} \end{pmatrix}$$

where λ represents the number of age groups in the dataset and, as before, k accounts for the number of nonabsorbing states and $k + 1$ destinations (including death). Columns of $\mathbf{d}(x, n)$ are indexed by states of origin i and destination j ; rows identify age intervals of length n ; the first age interval starts at age x and the last at age $x + \lambda n$.

The probabilities that an individual in state i at exact ages x will occupy state j at ages $x + n$ can then be obtained as

$${}_n q_x^{ij} = \frac{{}_n d_x^{ij}}{l_x^i}$$

Under the linear assumption, the number of person-years lived in state i between exact ages x and $x + n$ can be approximated by the arithmetic mean of the number of survivors times the length of the age interval:

$$\mathbf{L}(x, n) = \frac{n}{2} \{ \mathbf{l}(x) + \mathbf{l}(x + n) \} \quad (10)$$

7. In a case with four states and $n = 1$, as in the example in section 4.2, the number of individuals in the cohort who are in state 1 at exact age 11 is equal to $l_{11}^{11} = l_{10}^{11} - {}_1 d_{10}^{12} - {}_1 d_{10}^{13} - {}_1 d_{10}^{14}$. Notice that l_{11}^{11} differs from l_{11}^1 for not including the transitions from states 2 and 3 into state 1 between exact ages 10 and 11. Only the outflows from states 1 to 2, 1 to 3, and 1 to 4 are accounted for in l_{11}^{11} .

And under the exponential assumption, we obtain the number of person-years by solving the expression⁸

$$\mathbf{L}(x, n) = n\mathbf{l}(x) \left\{ \frac{n}{2!}\mathbf{M}(x, n) + \frac{n^2}{3!}\mathbf{M}^2(x, n) - \frac{n^3}{4!}\mathbf{M}^3(x, n) \dots \right\}$$

In the presence of an absorbing state and a last open-ended age group (that is, not constrained by a length of n , as in the case of all other age groups), the vector of person-years in the last time interval is⁹

$$\mathbf{L}(x, \infty) = \mathbf{l}(x, \infty) \left(\begin{array}{cccc} \infty M_x^{1d} + \sum \infty M_x^{1j} & -\infty M_x^{21} & \dots & -\infty M_x^{k1} \\ -\infty M_x^{12} & \infty M_x^{2d} + \sum \infty M_x^{2j} & \dots & -\infty M_x^{k2} \\ \vdots & \vdots & \ddots & \vdots \\ -\infty M_x^{1k} & -\infty M_x^{2k} & \dots & \infty M_x^{kd} + \sum \infty M_x^{kj} \end{array} \right)^{-1} \quad (11)$$

where each diagonal element contains the total rate of exit of state k —rates of mortality ∞M_x^{kd} plus mobilities $\sum \infty M_x^{kj}$ to state j —while the off-diagonal elements represent rates of moving from state j to k preceded by a minus sign.¹⁰

After we use Baum’s (2006) neat function to put matrices upside down, the number of person-years lived in state i above age x is

$$T_x^i = \sum_{y \geq x}^{\infty} L_y^i \quad (12)$$

Unconditional (or population-based) life expectancy at age x in state i is

$$e_x^i = \frac{T_x^i}{\sum l_x^i} \quad (13)$$

Notice that the denominator in (13) includes all survivors at age x , regardless of state of occupancy at that same age (Guillot 2015). The unconditional life expectancy refers to the number of years to be lived in state i after age x . It is “the average number of person-years lived in state i at and above exact age x [T_x^i] by all persons alive at exact age x [$\sum l_x^i$]” (Schoen 1988a, 83).

8. In fact, regardless of the assumption made (linear or exponential), when there is an absorbing state, $\mathbf{L}(x, n) = \{\mathbf{l}(x) - \mathbf{l}(x + n)\}\mathbf{M}(x, n)^{-1}$, with $\mathbf{M}(x, n)$ having its elements structured according to Rogers and Ledent (1976) [see (11)].

9. If the last age interval ends in $x + n$ instead of $x + \infty$, the option **censored** must be declared when **mslt** is invoked to avoid the default implementation of (11), which would lead to an overestimation of the last row or age group in matrix $\mathbf{L}(x, n)$ and a consequential inflation of derived life expectancies. With an absorbing state and censored data, $\mathbf{L}(x, n) = \{\mathbf{l}(x) - \mathbf{l}(x + n)\}\mathbf{M}(x, n)^{-1}$. Without an absorbing state and censored data, the calculation of $\mathbf{L}(x, n)$ in the last age group follows (10).

10. This approach uses the square matrix of mortality and mobility suggested by Rogers and Ledent (1976) and Krishnamoorthy (1979). This is a slightly modified longhand version of the equation suggested by Palloni (2001, 270) and Schoen (1988a, 70).

Alternatively, to calculate **conditional** (or status-based) life expectancies at age x in state j , for individuals who are in state i at age x , we use

$$e_x^{ij} = \frac{T_x^{ij}}{l_x^i} \quad (14)$$

The conditional life expectancy “reflects the average number of future years lived in state j [T_x^{ij}] by the closed group of persons in state i at exact age x [l_x^i]” (Schoen 1988a, 84). Conditional life expectancies refer to the lived experience of a single cohort born at a given age and in just one state (l_x^i, T_x^{ij}), while unconditional life expectancies combine the experience of cohorts born in all states ($\sum l_x^i, T_x^i$).¹¹

In addition, when **summary** is invoked, and depending on whether **death** is also specified, **mslt** produces the following summary measures (SM) (Schoen 1988a, 95):

Proportion of life spent in state i :

$$SM_1^i = \frac{T_0^i}{\sum T_0^i} \quad (15)$$

Probability of dying in state i ,

$$SM_2^i = \frac{\sum n d_x^{i\delta}}{\sum l_0^i}$$

where δ represents the absorbing state.

Average duration of state i :

$$SM_3^i = \frac{T_0^i}{\sum n d_x^{ji}} \quad (16)$$

Mean age at transfer from state i to j :

$$SM_4^{ij} = \frac{\sum (x + \frac{n}{2}) d_x^{ij}}{\sum n d_x^{ij}}$$

Mean age of persons in state i :

$$SM_5^i = \frac{\sum (x + \frac{n}{2}) L_x^i}{T_0^i} \quad (17)$$

For convenience, these quantities are saved in Mata and, for the sake of presentation, also as Stata matrices. When the inputs of **mslt**—occurrence rates or proportion

11. According to Guillot (2015, 112), unconditional life expectancies are weighted averages of conditional life expectancies, with the weights being the proportion of survivors in states 1 to k at age x .

of survivors—are not available, one can still obtain them by aggregating individual estimates, produced by event history (Cleves, Gould, and Marchenko 2016) or probability models (Long and Freese 2014), into population-level rates by age. The advantage of these microlevel approaches is that they are able to incorporate individual heterogeneities into the transitions; that is, individuals in state i at age x must not assume the same probability of experiencing a transition (Guillot 2015).

4 Extended examples

To illustrate and validate the results of `mslt`, I use information describing children's experience of cohabitation, marriage, or union disruption (Palloni 2001). The original data source is the National Survey of Family Growth, conducted in 1995 with 10,847 women between 15 and 44 years of age (Bumpass and Lu 2000). The survey contains retrospective information on union and fertility histories during the previous five years, which allows the calculation of transition rates between marital statuses. Figure 1 portrays, diagrammatically, the state-space of marital transitions.

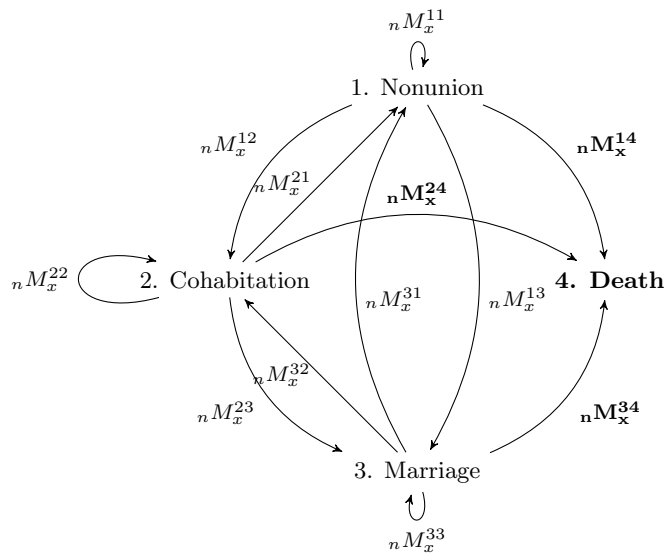


Figure 1. Multistate-space representation of children's transitions between their mothers' marital experience

In this representation, the stages of union formation and disruption are sequentially numbered, and the transition rates between them are indexed, so that the first superscript corresponds to the state of origin and the second to the state of destination. For example, ${}_nM_x^{13}$ represents the rate at which children who live with mothers who are not in a union experience a change between exact ages x and $x + n$ and begin to live with a married female parent. Transitions to the absorbing state (Death) are in boldface,

and widowhood, separation, and divorce states are not considered in this application. In the following section, I demonstrate the use of `mslt` in the absence of an absorbing state. Afterward, I simulate a situation in which death is included.

4.1 No absorbing state

This first example neglects the presence of death as an absorbing state. In such situation, the appropriate structure for a dataset containing transition rates¹² of children 0 to 14 years old according to their mother's marital status (1 = nonunion, 2 = cohabitation, 3 = marriage) is from the National Survey of Family Growth, *apud* Palloni (2001, 262).¹³ The suffixes in the variable names represent, respectively, the marital states of origin i and destination j : 1 is for mothers not in unions; 2 indicates cohabitation; and 3 indexes marriage.

```
. use mslt
. drop m14 m24 m34
. list, noobs separator(15)
```

| age | m12 | m13 | m21 | m23 | m31 | m32 |
|-----|-------|-------|-------|-------|-------|-------|
| 0 | .0777 | .0421 | .0968 | .146 | .0121 | .0086 |
| 1 | .0858 | .0405 | .0984 | .1411 | .0211 | .0082 |
| 2 | .1068 | .035 | .0759 | .1468 | .0196 | .0069 |
| 3 | .1054 | .0354 | .0829 | .1639 | .021 | .0045 |
| 4 | .0832 | .0475 | .0656 | .1282 | .0216 | .0084 |
| 5 | .0939 | .0497 | .0555 | .1433 | .0214 | .0076 |
| 6 | .0617 | .0469 | .0506 | .1229 | .0251 | .0022 |
| 7 | .0808 | .058 | .0471 | .1326 | .0201 | .0078 |
| 8 | .0507 | .0305 | .0655 | .1387 | .0196 | .0027 |
| 9 | .0621 | .0375 | .0815 | .143 | .0215 | .0031 |
| 10 | .0854 | .0411 | .0508 | .137 | .0201 | .0049 |
| 11 | .0435 | .0343 | .0855 | .1149 | .0186 | .0032 |
| 12 | .0656 | .0521 | .088 | .0896 | .026 | .0043 |
| 13 | .0427 | .0313 | .0812 | .1307 | .0204 | .0071 |
| 14 | .0837 | .0314 | .0851 | .0712 | .026 | .0066 |

The first column must always define the lower limit of exposure time (usually age) in a given age group. The remaining columns are observed transition rates from state i to state j in age group x to $x + 1$, ${}_1M_x^{ij}$. For example, the rate of a child aged 0 who lived with a mother not in a union to live with a cohabiting mom is, by age 1, 0.0777. Also notice that the rates of permanence in the same state (that is, `m11`, `m22`, and `m33`) are purposefully omitted and should not be part of the working data.

12. If instead of rates one had survivorship proportions, then the layout of the input data would have to be the same.

13. I benefited from Jann's (2016) excellent `texdoc` program to import Stata outputs into L^AT_EX.

Once a dataset with this structure and variables is entered into Stata, we are ready to explore the full potential of `mslt`.¹⁴

Unconditional (population-based) life expectancies

To estimate the unconditional duration of time lived in a particular state (13) by all individuals who were in a given age x , `mslt` requires the number of births in each state i , l_0^i . For example, by setting a radix of $l_0^i = 1000$ for every i state (even though it likely varies according to the number of registered births in each condition) and after assuming that transition rates are **constant** within age intervals, we obtain multistate life-table functions (option `matrix`) and key `summary` measures by typing

```
. mslt, l0(1000 1000 1000) censored constant matrix summary
li_x[15,3]: Individuals in the cohort who are in state i at age x
      state_1    state_2    state_3
0  1000.0000  1000.0000  1000.0000
1   983.9356   861.1597  1154.9047
2   966.0091   761.0255  1272.9655
3   914.4258   707.1952  1378.3790
4   874.4246   641.6943  1483.8812
5   836.4060   604.9001  1558.6939
6   786.9084   575.8688  1637.2228
7   772.1041   531.6504  1696.2454
8   727.4395   511.9164  1760.6442
9   734.6720   455.2725  1810.0555
10  735.9921   409.7317  1854.2762
11  703.3387   403.9149  1892.7464
12  716.3216   364.0969  1919.5815
13  713.8585   355.4149  1930.7267
14  727.7257   327.7023  1944.5720
```

14. If instead of transition rates by age we had individual observations in the rows and states of occupancy at times x (`m_x`) and $x + n$ (`m_xn`) in the columns, we would have to aggregate the data by time. To generate `m11`, `m12`, and `m13` by **age**, for example, we could use Jann's (2005b) `estpost` command in the following code snippet:

```
generate m=.          //Generates a blank variable
forvalues i=1/3 { //Indicates one's states of origin `i` and destination `j`
    forvalues j=1/3 {
        replace m=`i` `j` if m_x==`i` & m_xn==`j`
    }
}
estpost: tabulate age m if m<=13, row notot quietly //Tabulates transitions by age
matrix state1=e(b)
matrix m1=state1[1..1,1..15]`, state1[1..1,16..31]`, state1[1..1,32..47]`
```

The columns of matrix `m1`, from left to right, correspond to variables `m11`, `m12`, and `m13`. The other transitions could be obtained in a similar fashion with minor adjustments to the “if `m`” condition in the seventh line of this code.

Li_x[15,3]: Person-years lived in state i between age x and x+n

| | state_1 | state_2 | state_3 |
|----|----------|----------|-----------|
| 0 | 992.7694 | 927.7672 | 1079.4634 |
| 1 | 975.3960 | 809.1440 | 1215.4601 |
| 2 | 939.7762 | 733.5101 | 1326.7137 |
| 3 | 894.2234 | 673.4104 | 1432.3662 |
| 4 | 855.0677 | 622.9143 | 1522.0180 |
| 5 | 811.0594 | 590.2409 | 1598.6998 |
| 6 | 779.4352 | 553.1859 | 1667.3789 |
| 7 | 749.2250 | 521.7466 | 1729.0284 |
| 8 | 731.3328 | 482.5897 | 1786.0775 |
| 9 | 735.5727 | 431.6328 | 1832.7946 |
| 10 | 719.2816 | 406.9487 | 1873.7698 |
| 11 | 710.1561 | 383.2873 | 1906.5566 |
| 12 | 715.1053 | 359.6369 | 1925.2577 |
| 13 | 721.0414 | 341.0122 | 1937.9465 |
| 14 | 725.5465 | 338.0102 | 1936.4433 |

Ti_x[15,3]: Person-years lived in state i above age x

| | state_1 | state_2 | state_3 |
|----|------------|-----------|------------|
| 0 | 12054.9886 | 8175.0370 | 24769.9743 |
| 1 | 11062.2193 | 7247.2698 | 23690.5109 |
| 2 | 10086.8233 | 6438.1258 | 22475.0509 |
| 3 | 9147.0471 | 5704.6158 | 21148.3372 |
| 4 | 8252.8237 | 5031.2053 | 19715.9710 |
| 5 | 7397.7560 | 4408.2910 | 18193.9530 |
| 6 | 6586.6966 | 3818.0502 | 16595.2532 |
| 7 | 5807.2614 | 3264.8643 | 14927.8743 |
| 8 | 5058.0364 | 2743.1178 | 13198.8459 |
| 9 | 4326.7035 | 2260.5280 | 11412.7684 |
| 10 | 3591.1308 | 1828.8953 | 9579.9739 |
| 11 | 2871.8493 | 1421.9466 | 7706.2041 |
| 12 | 2161.6932 | 1038.6593 | 5799.6475 |
| 13 | 1446.5878 | 679.0224 | 3874.3898 |
| 14 | 725.5465 | 338.0102 | 1936.4433 |

di_x[15,6]: Individuals moving from state i to state j between ages x and x+n

| | 12 | 13 | 21 | 23 | 31 | 32 |
|----|---------|---------|---------|----------|---------|---------|
| 0 | 65.1182 | 44.3189 | 81.7053 | 130.2331 | 11.6675 | 7.9799 |
| 1 | 70.6161 | 42.1733 | 71.8625 | 108.1147 | 23.0003 | 9.2268 |
| 2 | 86.2497 | 37.8060 | 49.1940 | 99.8369 | 23.2785 | 8.9509 |
| 3 | 79.6387 | 36.7484 | 49.4665 | 102.4381 | 26.9193 | 6.7650 |
| 4 | 62.1008 | 42.5404 | 36.6328 | 74.6198 | 29.9899 | 12.3576 |
| 5 | 66.4119 | 43.1773 | 29.1491 | 78.2759 | 30.9425 | 11.9818 |
| 6 | 42.2419 | 37.2103 | 26.1316 | 64.7363 | 38.5164 | 4.4077 |
| 7 | 53.4157 | 44.9514 | 22.0095 | 64.3182 | 31.6929 | 13.1780 |
| 8 | 32.0470 | 23.3965 | 29.7396 | 63.9952 | 32.9364 | 5.0440 |
| 9 | 38.9026 | 28.8319 | 32.2320 | 58.2562 | 36.8226 | 6.0448 |
| 10 | 53.8412 | 31.9304 | 18.3122 | 50.9544 | 34.8060 | 9.6086 |
| 11 | 26.6951 | 24.5652 | 30.4838 | 42.1542 | 33.7593 | 6.1250 |
| 12 | 40.6700 | 36.6025 | 28.0625 | 30.2150 | 46.7469 | 8.9255 |
| 13 | 26.5365 | 23.0538 | 25.4840 | 41.7127 | 37.9736 | 12.9476 |
| 14 | 53.3286 | 23.2205 | 24.6592 | 21.6686 | 47.5314 | 13.6151 |

```

qi_x[15,6]: Prob. that an individual in state i at age x will be in state j
> at age x+n
      12      13      21      23      31      32
0 0.0651 0.0443 0.0817 0.1302 0.0117 0.0080
1 0.0718 0.0429 0.0834 0.1255 0.0199 0.0080
2 0.0893 0.0391 0.0646 0.1312 0.0183 0.0070
3 0.0871 0.0402 0.0699 0.1449 0.0195 0.0049
4 0.0710 0.0486 0.0571 0.1163 0.0202 0.0083
5 0.0794 0.0516 0.0482 0.1294 0.0199 0.0077
6 0.0537 0.0473 0.0454 0.1124 0.0235 0.0027
7 0.0692 0.0582 0.0414 0.1210 0.0187 0.0078
8 0.0441 0.0322 0.0581 0.1250 0.0187 0.0029
9 0.0530 0.0392 0.0708 0.1280 0.0203 0.0033
10 0.0732 0.0434 0.0447 0.1244 0.0188 0.0052
11 0.0380 0.0349 0.0755 0.1044 0.0178 0.0032
12 0.0568 0.0511 0.0771 0.0830 0.0244 0.0046
13 0.0372 0.0323 0.0717 0.1174 0.0197 0.0067
14 0.0733 0.0319 0.0752 0.0661 0.0244 0.0070

ei_x[15,3]: Population-based (unconditional) life expectancy at age x in state i
      state_1 state_2 state_3
0 4.0183 2.7250 8.2567
1 3.6874 2.4158 7.8968
2 3.3623 2.1460 7.4917
3 3.0490 1.9015 7.0494
4 2.7509 1.6771 6.5720
5 2.4659 1.4694 6.0647
6 2.1956 1.2727 5.5318
7 1.9358 1.0883 4.9760
8 1.6860 0.9144 4.3996
9 1.4422 0.7535 3.8043
10 1.1970 0.6096 3.1933
11 0.9573 0.4740 2.5687
12 0.7206 0.3462 1.9332
13 0.4822 0.2263 1.2915
14 0.2418 0.1127 0.6455

Measure_1[1,3]: Proportion of life spent in state i
      state_1 state_2 state_3
Proportion 0.2679 0.1817 0.5504

Measure_3[1,3]: Average duration of state i (Assuming flows from all
> non-absorbing states)
      state_1 state_2 state_3
Duration 11.5723 8.7436 15.9595

Measure_4[1,6]: Mean age at transfer from state i to state j
      12      13      21      23      31      32
Mean age 6.4012 6.7479 5.8709 5.6974 8.5404 7.7887

Measure_5[1,3]: Mean age of persons in state i
      state_1 state_2 state_3
Mean age 7.0137 6.1541 8.1809

```

In this particular example, because the last age group is not open ended (ending in $x+n$ instead of $x+\infty$), the option `censored` must be specified to avoid the assumption that everybody must die in the last age group (see footnote 9).

Quantities $(l_{x,n}^{ij}, {}_n d_x^{ij}, e_x^i)$ are analogous to those reported by Palloni (2001); hence, they have the same interpretation. Matrix **li_x** shows the number of individuals at each exact age and state. The quantities can increase or decrease, depending on the flows $({}_n d_x^{ij})$ from states i to j in specific age groups.

Matrices **di_x** and **qi_x** show exits and transition probabilities between all nonabsorbing states.¹⁵ According to matrix **di_x**, for example, between ages 0 and 1, ${}_1 d_0^{12} = 65$ individuals left state 1 (nonunion) for state 2 (cohabitation), and ${}_1 d_0^{13} = 44$ left state 1 for state 3 (marriage). During the same period of one year, ${}_1 d_0^{21} = 82$ individuals moved from state 2 to 1, and ${}_1 d_0^{31} = 12$ moved from state 3 to 1. As a result of these 109 exits and 93 entrances, there were 984 children living in state 1 at exact age 1, l_1^1 . Note that $l_1^1 + l_1^2 + l_1^3 = l_1 = 3000$, the total number of cohort members at age $x = 1$. In this example, because there is no mortality or any other absorbing state, this number equals to the sum of birth cohorts at all ages.

Another matrix of interest is **qi_x**, which shows six transition probabilities between three different states. The first column, ${}_1 q_x^{12}$, shows the probability for children who lived with a mother not in a union (state 1) at age x to live with a cohabiting mother (state 2) at age $x + 1$. At age 10, for example, a child in state 1 has three possibilities: a) to live with a cohabiting mother, with probability ${}_1 q_{10}^{12} = 7.32\%$; b) to live with a married mother, ${}_1 q_{10}^{13} = 4.34\%$; or to remain in the current living arrangement with a mother not in a union, ${}_1 q_{10}^{11} = 88.34\%$. Needless to say, these three probabilities add up to 100%, and a similar interpretation applies to the other columns or states.¹⁶

Matrix **ei_x**, which summarizes the mean duration of state occupancy above age x , shows that in the first 15 years of their lives, children could expect to experience 4.02 years in **state_1** (nonunion), 2.73 years in **state_2** (cohabitation), and 8.26 years in **state_3** (married). They will spend, therefore, most of the initial stages of their lives living in a family with married mothers. For ease of visualization, life expectancies are automatically plotted in a graph (see figure 2), which could be further edited and optionally saved by the user.

15. The quantities in **di_x** are the result of multiple flows involving one or more transitions between x and $x + 1$. They are affected by flows into state j but also out of it. Thus, ${}_1 d_x^{ij}$ is not a measure of pure decrements, except when j is an absorbing state. Also note that it excludes individuals who start in state 1, move to 2, and then exit 2 before attaining age $x + 1$. It also excludes children with multiple transitions, whose mothers may have moved from state 1 to 3, 3 to 1, 1 to 2, and then out of 2 from age 0 to just before reaching age 1. In most cases, however, time intervals are too short to allow more than one transition. As a result, "[...] the likelihood that an individual will experience multiple events in a single time interval is remote" (Palloni 2001, 266).

16. If death is present as an absorbing state, the values of ${}_{\infty} q_x^{ij}$ in the last age group must be adjusted accordingly. Probabilities of transitioning to nonabsorbing states must be equal to 0, and the probability of moving to the death state must be equal to 1 in the last open-ended age group.

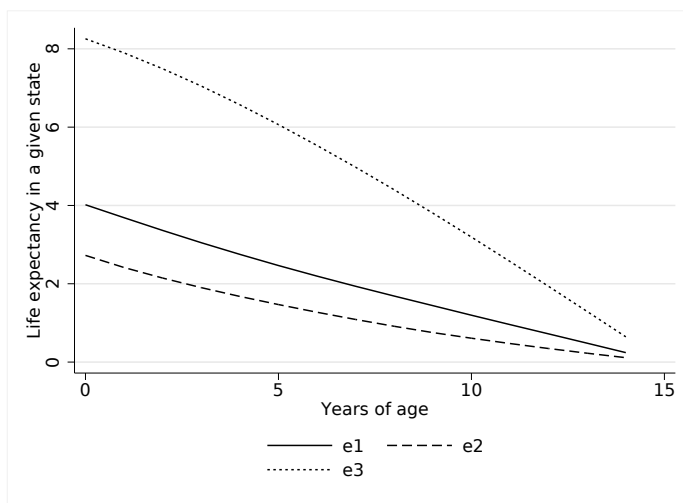


Figure 2. Unconditional life expectancies by age

The other two matrices, Li_x and Ti_x , account for person-years lived in state i . ${}_1L_0^1 = 992.77$, for example, represents the number of person-years at risk of experiencing a transition from state 1 to state j between exact ages 0 and 1. And $T_0^1 = 12,054.99$ corresponds to the cumulative sum of ${}_1L_{14}^1$ to ${}_1L_0^1$, as generally stated in (12). The meanings of ${}_nL_x^i$ and T_x^i , per se, are not intuitive for nondemographers unfamiliar with the concept of person-years. They are, however, fundamental steps to calculate important quantities of multistate life tables, shown in (13), (14), (15), (16), (17).

Finally, an added virtue of `mslt` is to produce key measures of interest, which show that about half (55.04%) of children's first 15 years will be spent with married mothers; that cohabitations have an average total (and not necessarily consecutive) duration of 8.7 years; that children move from the nonunion to the married state when they are about 6.75 years old; and that those living with married mothers are, on average, 8.18 years old.

Conditional (status-based) life expectancies

Conditional life expectancies are the predicted number of years to be lived in state j by those who are in state i at exact age x [see (14)]. They differ from unconditional estimates by considering the lived experience of a single cohort, born in just one (conditional) state. The strategy to obtain the quantities T_x^{ij} , required to estimate conditional life expectancies, is to recalculate the model with all persons born just in the (conditional) state and age of interest. By setting the number of “births” in the other cohorts and states to zero and requesting the option `conditional`, we specify that `mslt` produce status-based life expectancies at every age. For instance, by recalling `mslt` with 1,000 “births” at every age only in the nonunion state, we obtain the number of expected years lived in every state by those who were in `state_1` at exact age x :

```
. mslt, 10(1000 0 0) censored constant conditional
ei_x[15,3]: Status-based (conditional) life expectancy at age x in state i
      state_1 state_2 state_3
0    7.8650  2.4663  4.6686
1    7.5233  2.3341  4.1426
2    7.2012  2.1670  3.6318
3    6.9440  1.9141  3.1420
4    6.6759  1.6363  2.6878
5    6.3569  1.4234  2.2197
6    6.0858  1.1659  1.7483
7    5.6197  1.0068  1.3734
8    5.2619  0.7985  0.9396
9    4.6051  0.6665  0.7284
10   3.9785  0.5095  0.5120
11   3.3908  0.2973  0.3119
12   2.6029  0.2034  0.1937
13   1.8479  0.0882  0.0639
14   0.9474  0.0366  0.0160
```

Conditional life expectancy at birth (7.86 years)—for children who were born and were living with mothers out of wedlock—is about twice larger than previously estimated unconditional life expectancy (4.02 years) in `state_1` at that same age—for children born in any state. Therefore, the average number of future years lived in `state_1` by a closed group of persons in `state_1` at age x is almost two times larger than the number of years lived regardless of the state of origin. If we had instead set `10(0 1000 0)` or `10(0 0 1000)`, we would have obtained, respectively, conditional life expectancies for states 2 and 3.

To calculate conditional life expectancies, `mslt` generates a set of l_x^i , L_x^{ij} , and T_x^{ij} matrices for each age group and nonabsorbing state. Thus, in this particular example, 45 sets are generated (one for each of the 15 age groups in each of the 3 states). For instance, conditional life expectancies at ages 4 and 5 were respectively calculated from state-specific rows in Mata matrices ($T_4^{1j} = T5[5,.]$, $l_4^1 = 15[5,1] = 1000$; and $T_5^{1j} = T6[6,.]$, $l_5^1 = 16[6,1] = 1000$):

```
. quietly mslt, 10(1000 0 0) censored constant conditional
. mata T5[5,.]:/15[5,1]
      1          2          3
1  6.675944544  1.636253547  2.687801909
. mata T6[6,.]:/16[6,1]
      1          2          3
1  6.356854301  1.423423288  2.219722412
```

In this example, the starting radix ($l_x^1 = 1000$) is entered 15 times in state 1, once for each age group, to obtain the appropriate values of T_x^{ij} for every starting age of interest. Only the initial radix (that is, 1,000) was used in the denominator of (14). Life expectancies conditional on those who were in state 2 (that is, `mslt, 10(0 1000 0) conditional`) would use the same initial value in the denominator of (14), but the

number of person-years lived in state j above age x , the numerator T_x^{ij} , would change accordingly.

The options `matrix` and `summary` are not available when `conditional` is specified because it generates several (135 in this case) state-specific Mata matrices to derive status-based life expectancies at ages x and not only one set, as in the case of population-based (unconditional) life expectancies.

4.2 Incorporating an absorbing state (exempli gratia, death)

`mslt` is capable of handling any number of nonabsorbing states and, at most, one absorbing state. We now turn, therefore, to an example with four states, in which “4.death” is a possibility. Variables `m12`, `m13`, `m21`, `m23`, `m31`, and `m32` are the same as in section 4.1, but the data now also include transitions from nonunion, cohabitation and marriage statuses to death (that is, `m14`, `m24`, and `m34`). Note that the data from the National Survey of Family Growth now include randomly generated death rates. Also note that the suffixes in the variable names represent, respectively, the marital states of origin i and destination j : 1 is for mothers not in unions; 2 indicates cohabitation; 3 indexes marriage; and 4 represents the death state.

```
. use mslt, clear
. list, noobs separator(15)
```

| age | m12 | m13 | m14 | m21 | m23 | m24 | m31 | m32 | m34 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0 | .0777 | .0421 | .008 | .0968 | .146 | .0218 | .0121 | .0086 | .0149 |
| 1 | .0858 | .0405 | .0053 | .0984 | .1411 | .0137 | .0211 | .0082 | .0007 |
| 2 | .1068 | .035 | .0014 | .0759 | .1468 | .0067 | .0196 | .0069 | .021 |
| 3 | .1054 | .0354 | .0356 | .0829 | .1639 | .0835 | .021 | .0045 | .0061 |
| 4 | .0832 | .0475 | .0054 | .0656 | .1282 | .1937 | .0216 | .0084 | .0202 |
| 5 | .0939 | .0497 | .0551 | .0555 | .1433 | .0042 | .0214 | .0076 | .0146 |
| 6 | .0617 | .0469 | .018 | .0506 | .1229 | .0134 | .0251 | .0022 | .0098 |
| 7 | .0808 | .058 | .0275 | .0471 | .1326 | .0081 | .0201 | .0078 | .0152 |
| 8 | .0507 | .0305 | .0074 | .0655 | .1387 | .2037 | .0196 | .0027 | .0154 |
| 9 | .0621 | .0375 | .0326 | .0815 | .143 | .0605 | .0215 | .0031 | .0013 |
| 10 | .0854 | .0411 | .1036 | .0508 | .137 | .1211 | .0201 | .0049 | .0037 |
| 11 | .0435 | .0343 | .0003 | .0855 | .1149 | .1841 | .0186 | .0032 | .0198 |
| 12 | .0656 | .0521 | .005 | .088 | .0896 | .0167 | .026 | .0043 | .0289 |
| 13 | .0427 | .0313 | .0025 | .0812 | .1307 | .0139 | .0204 | .0071 | .0095 |
| 14 | .0837 | .0314 | .01 | .0851 | .0712 | .0099 | .026 | .0066 | .0185 |

The mortality rate for a child aged 0, who lived with a mother not in a union (`m14`), is 0.008. Notice that a) transitions to the absorbing state (`m14`, `m24`, `m34`) should be the last column within each nonabsorbing state of analysis; b) there are, by definition, no decrements from the absorbing state (that is, no `m41`, `m42`, and `m43` variables); and c) the last age group is censored, instead of open ended. With an absorbing state, the option `death` must be declared, and the last value in `10(numlist)` must be zero. Once again, assuming `constant` force of mortality, unconditional life expectancies and `summary` measures are obtained by entering the following command line:

```

. mslt, 10(1000 1000 1000 0) death censored constant summary
ei_x[15,3]: Population-based (unconditional) life expectancy at age x in state i
      state_1 state_2 state_3
0   3.4564   2.1479   7.1961
1   3.1733   1.8692   6.9406
2   2.8635   1.6120   6.5731
3   2.5781   1.3867   6.2040
4   2.3554   1.2098   5.9171
5   2.1716   1.0794   5.6738
6   1.9438   0.9344   5.2520
7   1.7061   0.7818   4.7385
8   1.4854   0.6370   4.2203
9   1.2957   0.5231   3.7580
10  1.0686   0.4178   3.1707
11  0.8766   0.3284   2.6172
12  0.6776   0.2465   2.0008
13  0.4618   0.1682   1.3499
14  0.2325   0.0862   0.6735

Measure_1[1,4]: Proportion of life spent in state i
      state_1 state_2 state_3 state_4
Proportion  0.2304  0.1432  0.4797  0.1466

Measure_2[1,3]: Probability of dying in state i
      state_1 state_2 state_3
Probability  0.0743  0.1148  0.0961

Measure_3[1,3]: Average duration of state i (Assuming flows from all
> non-absorbing states)
      state_1 state_2 state_3
Duration 12.1275  8.0572 17.0240

Measure_4[1,9]: Mean age at transfer from state i to state j
      12      13      14      21      23      24      31      32      34
Mean age 5.8651 6.2394 7.1838 4.8637 4.8835 6.4688 8.2024 7.3921 8.3758

Measure_5[1,4]: Mean age of persons in state i
      state_1 state_2 state_3 state_4
Mean age  6.4638  5.2862  7.8327 10.2018

```

Unconditional life expectancies at birth for `state_1`, `state_2`, and `state_3` are now 3.46, 2.15, and 7.20 years. And conditional values could be obtained as before, by recalling `mslt` with the option `conditional` and zeros in specific states of `10(numlist)`.

`Measure_1` shows that individuals, respectively, spend 23%, 14%, and 48% of their lives in states 1, 2, and 3 (the remaining 15% accounts for those who died). When there is an absorbing state and the option `summary` is requested, `mslt` reports the probability of dying in state i , `Measure_2`. In our example, it shows that the probability of dying in `state_1` is 7.43% and in `state_2` is 11.48%.

Multistate life-table functions are saved as Mata and as Stata matrices and can be listed accordingly. To list ${}_1d_x^{ij}$, the number of transitions from state i to state j between exact ages x and $x + 1$, for example, one could invoke

```
. mata d
      (output omitted)
. matrix list di_x, format(%9.3f)
di_x[15,9]
      12      13      14      21      23      24      31      32      34
0   64.165   43.795   8.588   80.509  127.888   20.520   11.533   7.837  14.777
1   69.354   41.670   5.372   69.698  105.111   10.454   22.586   9.025   0.908
2   84.603   36.823   1.920   47.393   95.235    5.492   22.633   8.684  25.726
3   73.731   35.223   32.808   44.946   94.532   49.695   25.381   6.240   8.693
4   53.279   39.478   10.201   29.659   60.027   92.713   28.235  10.670  29.061
5   60.237   39.027   39.515   21.226   58.146    2.786   27.772  11.005  21.508
6   36.556   32.263   12.121   19.781   49.205    5.850   34.448   3.949  14.630
7   45.197   37.920   17.347   16.611   48.837    3.643   27.679  11.613  22.959
8   24.226   19.247    7.139   20.799   44.592   66.763   28.429   3.986  23.761
9   30.717   23.413   19.580   19.761   36.298   15.766   30.913   5.008   2.649
10  38.219   23.998   56.713   10.243   29.988   27.341   27.898   7.626   7.683
11  17.491   17.353    2.054   16.191   22.182   36.089   27.837   4.643  31.037
12  28.691   25.659    3.034   13.883   14.771    3.024   37.184   7.065  43.752
13  18.603   16.192    1.427   13.241   21.597    2.442   29.434   9.982  14.154
14  37.072   16.074    5.150   13.222   11.569    1.800   35.901  10.286  27.110
```

One of the values in matrix `di_x` shows, for example, that from exact ages 10 to 11, 57 children moved from states 1 to 4, ${}_1d_{10}^{14} = 56.713$ deaths. This number is about eight times higher than the quantity of children who made transitions from state 3 to state 2 at the same age interval, ${}_1d_{10}^{32} = 7.626$.

Mata and Stata commands report the same values, but if the option `matrix` had been called with `mslt`, a neat title for each multistate function would have been displayed. Graphs of Mata matrices can be produced with function `mm_plot()`, available in Jann's (2005a) `moremata`, or by converting Stata matrices to variables using `svmat` (see [M-4] **Matrix**) for manipulation with `twoway` plots (see [G-2] **graph**).

5 The mslt command

5.1 Syntax

```
mslt, 10(numlist) [proportion constant death censored conditional matrix
summary]
```

5.2 Description

The most important quantities in increment–decrement tables are conditional and unconditional life expectancies, which summarize the average duration of time spent in a particular state i above a given age x . `mslt` calculates the functions $({}_x^i l_x, {}_n L_x^i, T_x^i, {}_n d_x^{ij}, {}_n q_x^{ij}, e_x^i, e_x^{ij})$ of increment–decrement life tables following the methods and procedures described in Rogers and Ledent (1976), Schoen (1988a), and Palloni (2001). By specifying the `summary` option, `mslt` also calculates mean ages, proportions of life, and probabilities of dying in a given state. The inputs for `mslt` are

1. transition rates (or survivorship proportions) between various states by age; and
2. the initial number of individuals in each cohort and state.

An adequate dataset, before entered into Stata, should have the same structure and order of variables described in the examples of sections 4.1 and 4.2.

By default, **mslt** assumes that rates are the working input and that the underlying risks of transition are linear within age intervals. Users can, however, optionally enter survivorship proportions (that is, **proportion**) and assume **constant** rates in each age interval. Doing so will calculate the number of survivors in the cohort in state i and age x , function l_x^i , using the exponential method.

5.3 Options

10(numlist) lists the initial number of individuals in each cohort and state. Traditionally, these radices are set as multiples of hundreds (1,000, 10,000, etc.), but the user is allowed to enter other values. A suitable option is to consider the actual distribution of births by state, as observed in the populations under investigation. In a situation with three nonabsorbing and one absorbing state, four values must be provided, and the last one of them is necessarily zero because people cannot be born in an absorbing state. The *numlist* in such a situation could be, for example, **10(1000 1000 1000 0)**. **10()** is required.

proportion declares that the input data refer to survivorship proportions instead of rates. When one enters this option, **mslt** converts proportions to probabilities and then to rates under the linear assumption. If the option **constant** is simultaneously informed, the conversion is made under the constant forces assumption.

constant implements the exponential solution, under the constant forces assumption, to calculate the number of individuals in state i at exact age x , l_x^i . It should be declared whenever this function is nonlinear (exponential) with age or the underlying risks are relatively stable. Otherwise, “the exponential method produces somewhat inaccurate values in cases where the transition probabilities are increasing or decreasing rapidly” (Schoen 1988a, 75–76).

death declares the presence of an absorbing state. This option should be included whenever there is a state to which people are able to enter but not leave. Examples of absorbing states are death, schooling, and incurable diseases.

censored declares that the last age interval is censored (that is, ending in $x+n$), instead of open ended (ending in $x+\infty$). To warrant an accurate calculation of the number of person-years, one must specify this option whenever the last age group is not open ended.

conditional calculates life expectancies conditional on state of occupancy at age x and must be properly combined with births in just one state in `10(numlist)`. By default, **mslt** estimates unconditional life expectancies.

matrix tells **mslt** to display matrices on the screen containing the key functions (l_x^i , L_x^i , T_x^i , ${}_n d_x^{ij}$, ${}_n q_x^{ij}$, e_x^i) of increment–decrement life tables for each state i and at exact ages x . This option cannot be specified with the option **conditional**.

summary displays summary measures for multistate life tables. It reports the proportion of life spent in state i (**Measure_1**), the average age of persons in state i (**Measure_5**), the average duration of state i (**Measure_3**)—assuming the presence of flows between all nonabsorbing states—and the mean age at transfer from state i to state j (**Measure_4**). Finally, if **death** is simultaneously requested, **summary** shows the probability of dying in state i (**Measure_2**). This option is not available when status-based life expectancies (that is, **conditional**) are requested.

5.4 Output

As a minimum, **mslt** displays a table and a graph of life expectancies by state and time. With option **matrix**, it shows matrices of multistate functions on the screen, and with option **summary**, it displays summary measures of multistate life tables.

5.5 Stored results

Multistate life-table functions are stored in Mata (without the “**i_x**” suffix) and as Stata matrices under the following names:

| | |
|------------------|--|
| li_x | individuals in the cohort who are in state i at age x |
| Li_x | person-years lived in state i between age x and $x + n$ |
| Ti_x | person-years lived in state i above age x |
| di_x | individuals moving from state i to state j between ages x and $x + n$ |
| qi_x | probability that an individual in state i at age x will be in state j at age $x + n$ |
| ei_x | life expectancy at age x in state i |
| Measure_1 | proportion of life spent in state i vector |
| Measure_2 | probability of dying in state i vector (if death is included) |
| Measure_3 | average duration of state i vector |
| Measure_4 | mean age at transfer from state i to state j vector |
| Measure_5 | mean age of persons in state i vector |

6 Acknowledgments

This article was drafted while I was a visiting scholar in the Department of Sociology at the University of British Columbia, Vancouver, Canada. During this sabbatical year, I received financial support from the CAPES Foundation, process PVEX - 88881.170751/2018-01. The production of **mslt** benefited from discussions in Statalist and *The Stata Blog*. These places remain as the most valuable sources of brief and precise answers to various challenging technical questions.

I thank anonymous reviewers for their insightful comments and Dr. Guy Stecklov for giving me feedback on a preliminary draft of this article. I am most grateful and in debt to Dr. Jacques Ledent, one of the pioneers of multistate methods, who has enormously contributed to the field. Dr. Ledent kindly tried `mslt` and provided thoughtful comments in the elaboration of this article. The views expressed in this article, as usual, are those of the author. Remaining errors are hence my sole responsibility.

7 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 20-3
. net install st0615      (to install program files, if available)
. net get st0615          (to install ancillary files, if available)
```

8 References

- Alm, G. 1959. Connection between maturity, size, and age in fishes. Technical report, Drottningholm Institution Freshwater Research.
- Baum, C. F. 2006. Stata tip 37: And the last shall be first. *Stata Journal* 6: 588–589. <http://doi.org/10.1177/1536867X0600600411>.
- Bumpass, L., and H.-H. Lu. 2000. Trends in cohabitation and implications for children's family contexts in the United States. *Population Studies* 54: 29–41. <https://doi.org/10.1080/713779060>.
- Caswell, H. 2001. *Matrix Population Models: Construction, Analysis, and Interpretation*. 2nd ed. Sunderland, MA: Sinauer Associates.
- Cleves, M., W. W. Gould, and Y. V. Marchenko. 2016. *An Introduction to Survival Analysis Using Stata*. Rev. 3rd ed. College Station, TX: Stata Press.
- DuPasquier, L. G. 1912. Mathematische theorie der invaliditätsversicherung. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker* 7: 1–17.
- Guillot, M. 2015. Multistate Transition Models in Demography. In *International Encyclopedia of the Social and Behavioral Sciences*, ed. J. D. Wright, 2nd ed., 109–115. Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.31019-4>.
- Hall, B. C. 2015. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. 2nd ed. Switzerland: Springer.
- Hughes, T. P., and J. B. C. Jackson. 1985. Population dynamics and life histories of foliaceous corals. *Ecological Monographs* 55: 141–166. <https://doi.org/10.2307/1942555>.

- Jackson, J. B. C. 1985. Distribution and ecology of clonal and aclonal benthic invertebrates. In *Population Biology and Evolution of Clonal Organisms*, ed. J. B. C. Jackson, L. W. Buss, and R. E. Cook, 297–356. New Haven, London: Yale University Press. <https://doi.org/10.2307/j.ctt2250w9n.12>.
- Jann, B. 2005a. moremata: Stata module (Mata) to provide various functions. Statistical Software Components S455001, Department of Economics, Boston College. <https://ideas.repec.org/c/boc/bocode/s455001.html>.
- . 2005b. Making regression tables from stored estimates. *Stata Journal* 5: 288–308. <https://doi.org/10.1177/1536867X0500500302>.
- . 2016. Creating L^AT_EX documents from within Stata using texdoc. *Stata Journal* 16: 245–263. <https://doi.org/10.1177/1536867X1601600201>.
- Jiménez, J. A., A. E. Lugo, and G. Cintrón. 1985. Tree mortality and mangrove forests. *Biotropica* 17: 177–185.
- Krishnamoorthy, S. 1979. Classical approach to increment-decrement life tables: An application to the study of the marital status of United States females, 1970. *Mathematical Biosciences* 44: 139–154. [https://doi.org/10.1016/0025-5564\(79\)90033-6](https://doi.org/10.1016/0025-5564(79)90033-6).
- Land, K. C., J. M. Guralnik, and D. G. Blazer. 1994. Estimating increment-decrement life tables with multiple covariates from panel data: The case of active life expectancy. *Demography* 31: 297–319. <https://doi.org/10.2307/2061887>.
- Ledent, J. 1980. Multistate life tables: Movement versus transition perspectives. *Environment and Planning A: Economy and Space* 12: 533–562. <https://doi.org/10.1068/a120533>.
- Ledent, J., and Y. Zeng. 2010. Multistate Demography. In *Encyclopedia of Life Support Systems*, vol. 2, ed. Y. Zeng, 139–163. Oxford: EOLSS Publishers/Unesco.
- Lefkovich, L. P. 1965. The study of population growth in organisms grouped by stages. *Biometrics* 21: 1–18. <https://doi.org/10.2307/2528348>.
- Long, J. S., and J. Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. 3rd ed. College Station, TX: Stata Press.
- Lutz, W., and A. Goujon. 2001. The world's changing human capital stock: Multi-state population projections by educational attainment. *Population and Development Review* 27: 323–339. <https://doi.org/10.1111/j.1728-4457.2001.00323.x>.
- Palloni, A. 2001. Increment-decrement life tables. In *Demography: Measuring and Modeling Population Processes*, ed. S. H. Preston, P. Heuveline, and M. Guillot, 256–272. Oxford: Blackwell.
- Rees, P. H., and A. G. Wilson. 1977. *Spatial Population Analysis*. London: Arnold.
- Rogers, A. 1973. The multiregional life table. *Journal of Mathematical Sociology* 3: 127–137. <https://doi.org/10.1080/0022250X.1973.9989827>.

- . 1975. *Introduction to Multiregional Mathematical Demography*. New York: Wiley.
- Rogers, A., and J. Ledent. 1976. Increment-decrement life tables: A comment. *Demography* 13: 287–290. <https://doi.org/10.2307/2060807>.
- Schoen, R. 1975. Constructing increment-decrement life tables. *Demography* 12: 313–324. <https://doi.org/10.2307/2060768>.
- . 1988a. *Modeling Multigroup Populations*. New York: Plenum Press.
- . 1988b. Practical uses of multistate population models. *Annual Review of Sociology* 14: 341–361. <https://doi.org/10.1146/annurev.so.14.080188.002013>.
- Schoen, R., and K. C. Land. 1979. A general algorithm for estimating a Markov-generated increment-decrement life table with applications to marital-status patterns. *Journal of the American Statistical Association* 74: 761–776. <https://doi.org/10.1080/01621459.1979.10481029>.
- Smith, S. J. 1982. *Tables of working life: The increment-decrement model*, Bulletin 2135. Washington, DC: U.S. Dept. of Labor, Bureau of Labor Statistics. <http://catalog.hathitrust.org/Record/011408963>.
- Somerton, D. A., and R. A. MacIntosh. 1983. The size at sexual maturity of blue king crab, *Paralithodes platypus*, in Alaska. *Fishery Bulletin* 81: 621–628.
- Usher, M. B. 1966. A matrix approach to the management of renewable resources, with special reference to selection forests. *Journal of Applied Ecology* 3: 355–367. <https://doi.org/10.2307/2401258>.
- Willekens, F., and H. Putter. 2014. Software for multistate analysis. *Demographic Research* 31: 381–420. <http://doi.org/10.4054/DemRes.2014.31.14>.
- Willekens, F. J. 1980. Multistate analysis: Tables of working life. *Environment and Planning A: Economy and Space* 12: 563–588. <https://doi.org/10.1068/a120563>.
- Yang, Y., and M. Walij. 2010. Increment-decrement life table estimates of happy life expectancy for the U.S. population. *Population Research and Policy Review* 29: 775–795. <https://doi.org/10.1007/s11113-009-9162-5>.
- Zon, R. 1915. Seed production of western white pine. Technical report, Bulletin No. 210, Department of Agriculture.

About the author

Jerônimo O. Muniz is an associate professor and researcher in the Department of Sociology at the Federal University of Minas Gerais, Brazil. His research interests include formal demography, forecast methods, social stratification, and sociology of race.