DEMOGRAPHIC RESEARCH

Research Article

Software for multistate analysis

Frans Willekens

Hein Putter

Table of Contents

Software for multistate analysis

Frans Willekens[1]

Hein Putter[2]

Abstract

BACKGROUND
The growing interest in pathways, the increased availability of life-history data, innovations in statistical and demographic techniques, and advances in software technology have stimulated the development of software packages for multistate modeling of life histories.

OBJECTIVE
In the paper we list and briefly discuss several software packages for multistate analysis of life-history data. The packages cover the estimation of multistate models (transition rates and transition probabilities), multistate life tables, multistate population projections, and microsimulation.

METHODS
Brief description of software packages in a historical and comparative perspective.

RESULTS
During the past 10 years the advances in multistate modeling software have been impressive. New computational tools accompany the development of new methods in statistics and demography. The statistical theory of counting processes is the preferred method for the estimation of multistate models and R is the preferred programming platform.

CONCLUSIONS
Innovations in method, data, and computer technology have removed the traditional barriers to multistate modeling of life histories and the computation of informative life-course indicators. The challenge ahead of us is to model and predict individual life histories.

[1] Max Planck Institute for Demographic Research, Germany. E-Mail: willekens@nidi.nl.
[2] University of Leiden, the Netherlands. E-Mail: H.Putter@lumc.nl.

1. Introduction

Software for the multistate analysis of event histories and population dynamics originates in biostatistics, demography, epidemiology, and econometrics. The increased availability of life-history data, and in particular longitudinal data, has stimulated the development of life-history models. New data and models called for new methods to estimate model parameters from empirical data and to derive easy-to-interpret life-history indicators. That innovation has been accompanied by software development. As software development is related to methodological innovation, this paper includes a brief discussion of methods. Some of the software products listed are quite old. They are included because they add a historical perspective and they illustrate the impact of technology on software development for multistate analysis.

Along the life course, individuals move between states. The sequence of states and sequence of transitions represent life histories. Some scientists approach a life history as a whole and aim at finding typical patterns. The approach is generally known as sequence analysis. Other scientists view a life history as a realization of a stochastic process and model that process to explain and predict life histories[3]. A common probability model used for this purpose is the continuous-time Markov process model. The Markov model and its extensions are implemented in a range of scientific software packages. That software is the subject of this paper. Sequence analysis software is beyond the scope of the paper.

Biostatisticians focus on the estimation of multistate models, that is, on the estimation of transition rates and transition probabilities from observations on transitions in a sample population. Statistical inference is the focus of interest and sampling may influence the appropriate estimation method (Keiding 2014). Demographers traditionally rely on census data and vital statistics. Here, sampling is not a real issue, but missing data and misreporting are. With an increased collection of micro-data in demographic studies, sampling has also become a more important issue in demography, and methods and models developed in biostatistics have been imported into demography. This history explains why statistical inference in the estimation of transition rates did not receive the attention in demography that it did in biostatistics. In demography the focus has traditionally been on using transition rates to compute summary indicators of lifespan and life history, and to study the impact of transition rates on the dynamics of cohorts and populations. The (multistate) life table combines transition rates and lifespan and life history indicators. They summarize how members of a synthetic birth cohort move between states, how they are distributed between states

---

[3] Abbott (2001) explains and discusses the complementarity of the two approaches.

at consecutive ages, and how long they spend in each state. The indicators are easier to interpret than transition rates or transition probabilities. The model underlying the life table is usually a continuous-time Markov model with rates that are fixed within age intervals and vary between age intervals. The medium-term consequences of transition rates are usually presented as population projections. The long-term consequences are described in terms of the characteristics of a stable population, that is, a population in dynamic (steady-state) equilibrium. The common model for multistate population projection is the multistate cohort-survival model, which distinguishes between birth cohorts and uses transition rates to estimate and project cohort life histories (cohort biographies). The cohort-survival model is essentially a Markov process model applied to birth cohorts.

Members of a birth cohort differ. To account for that heterogeneity, covariates are introduced and unobserved heterogeneity may be considered. To project individual life histories of members of a cohort, microsimulation may be used.

The paper consists of two sections, in addition to the introduction and the conclusion. In Section 2 we highlight a few core concepts in multistate event-history modeling. That brief account serves as a reference for a comparison of different software packages. An overview of software for estimating transition rates and transition probabilities follows, with an emphasis on transition rates. Section 3 covers software for multistate life-table analysis and population projection.

2. Statistical inference

In the early stages of event-history modeling, a single state and a single transition were considered (see e.g., Blossfeld and Rohwer 2002). Observation was assumed to start at the beginning of an episode and to end with the occurrence of the transition of interest, another transition, or interview date (censoring). Survival models are examples of single state models. The state is 'being alive' and the event is death. In other applications the state is 'married' and the event is marriage dissolution. In contrast, multistate models consider multiple states and transitions between states. When a state represents a stage of life, the multistate model is a multistage model. In a multistate model an individual who leaves a state may enter one of several other states. Since the possible destinations compete to be the destination state, the transition model is a competing risk model. Thus, multistate models may generally be viewed as a series of competing risk models. In a multistate model individuals may move freely between states and enter and leave states at any time. This flexibility poses particular problems for the estimation of transition rates and transition probabilities. Estimation methods should allow for observations that are left-truncated and right-censored. The theory of

counting processes, pioneered by Aalen (1975) in his dissertation, satisfies that requirement (for details, see Andersen et al. 1993; Andersen and Keiding 2002; Aalen et al. 2008; Beyersmann et al. 2012). Andersen and Gill (1982) suggested a format for counting process data that applies to multistate models: a period of observation (episode) is described by a starting time, an ending time, and a reason for ending (transition or censoring). This format is applied in several software packages. Often, the starting time and/or ending time of an episode may not be available. The information may be limited to the period in which the episode starts or ends. That is the case if individuals are observed at discrete intervals, as in panel surveys. The estimation of the parameters of a stochastic process in continuous time from observations in discrete time raises methodological issues. Most software packages reviewed in this paper deal with that issue in a similar way, by relying on a method developed by Kalbfleisch and Lawless (1985) for estimating a continuous-time Markov model from panel data. The method estimates transition rates from the transition probabilities of the discrete-time Markov chain embedded in the Markov process studied (embedded Markov chain). A similar approach was discussed by Singer and Spilerman (1976). Some methods go beyond Markov processes. They include non-parametric methods in illness-death models (Meira-Machado et al. 2006) and semi-Markov models based on Markov renewal processes (Wolf 1988; Rajulton and Lee 1988; Dabrowska et al. 1994; Spitoni et al. 2010). Software implementing these models is scarce.

It is useful to distinguish between transition probabilities and transition rates. Transition rates and transition probabilities are different parameters of the same underlying continuous-time Markov process model. A transition probability relates transitions during an interval to the population at risk at the beginning of the interval. If two states are considered the transition occurs or does not occur and the outcome variable is binary or dichotomous. When multiple states are considered the outcome is polytomous. Transition probability models include the logit, probit, and extreme value (log-log) models.

Transition rates relate the number of events during an observation period to the duration at risk during that period. The duration at risk is measured in person-days, person-months, or person-years. Transition rates, as defined here, are also known as occurrence-exposure rates because they relate the number of occurrences in a period to the exposure time in that period. Transition rate models, i.e., regression models that predict transition rates, are distinguished on the basis of the age- or duration-dependence they impose on the transition rate. Non-parametric models do not impose any duration dependence. The semi-parametric Cox model does not impose duration dependence either, but it determines that the effect of covariates on the transition rate does not vary with duration. The exponential transition rate model imposes a constant transition rate. If transition rates are assumed to be constant in the interval between two

ages and to vary between age intervals, the transition rates are piecewise constant. Different exponential models apply in different age intervals. Often the shape of duration-dependent transition rates is stated to follow a specific pattern, which can be mapped by a parametric function. In the Gompertz model the transition rate increases exponentially with duration and in the Weibull model the transition rate follows a power function. Recent multistate models account for unobserved heterogeneity (see e.g., Bijwaard in this Special Collection). This is typically done by including frailty terms in models of transition rates. Since there are many transition rates and typically few observed transitions for a given individual, inclusion of frailties in multistate models is a delicate balancing act between model fit and identifiability (Putter and van Houwelingen 2011).

Numerous software tools exist for the analysis of event history data. General-purpose statistical packages such as STATA, SAS, SPSS, BMDP, S-Plus, and R include procedures to analyze life history data and estimate transition rates and transition probabilities. Goldstein and Harrell (2005) list and briefly review several of these packages. In addition specialized packages exist. In the social sciences the first specialized package for multistate modeling was *RATE*, published by Tuma (1979). In the early days of transition rate and multistate modeling the package was widely used and its use was facilitated by the publication of the textbook by Tuma and Hannan (1984). Blossfeld et al. (1989) used *RATE* but they also used Fortran programs and SAS. Later, Blossfeld and Rohwer (2002) developed *TDA* (Transition Data Analysis), which includes a set of parametric hazard functions such as the Gompertz, Weibull, log-logistic, and sickle models. *TDA* considers single episodes and adopts the Andersen and Gill format for counting process data, although left truncation is not considered. More recently Blossfeld et al. (2007) published a book on event-history analysis with Stata. The text is the same as Blossfeld and Rohwer (2002) but Stata code replaces the TDA programs. Resources for the 1995/2002 and 2007 books are available from the authors' website at http://oldsite.soziologie-blossfeld.de/eha/ (Accessed 21 March 2014).

CTM (Continuous Time Models) was developed at the National Opinion Research Center at the University of Chicago under the direction of James Heckman (Yi et al. 1987). It is a program, written in Fortran, for the estimation of continuous-time multistate models in labor econometrics. *CTM* generalizes the theory of competing risks and allows for transitions between any number of states, time-varying covariates, person and state-specific unobserved heterogeneity, and arbitrary duration dependence of the transition rate. The theoretical basis for the model was published in Flinn and Heckman (1982), Heckman and Singer (1985), and other papers. For a description, see Steinberg (1991) and for an application see e.g., Heckman and Walker (1990). *aML* is a software for multilevel and multiprocess models (Lillard and Panis 2003). It is based on a method developed by Lillard (1993). It first became commercially available in 2000,

but now it is free (http://www.applied-ml.com/; accessed 21 March 2014). The package fits a range of multilevel models used in econometrics. The program was designed specifically for fitting multiprocess or simultaneous equation models to hierarchical data where one or more of the explanatory variables may be non-random or endogenous with respect to the outcome variable of interest. It considers processes based on unobserved heterogeneity. An extensive User's Guide and Reference Manual can be downloaded from the website. For recent applications of *aML*, visit the website of the Centre for Multilevel Modelling at Bristol University (http://www.bris.ac.uk/cmm) (accessed 21 March 2014).

In recent years the number of packages for the estimation of multistate models has increased rapidly, with R being the language of choice. R is a software environment for statistical computing and graphics. Functionality is contributed to the environment in the form of R packages, which are maintained in the Comprehensive R Archive Network (CRAN) (http://cran.r-project.org/). In January 2011, the Journal of Statistical Software devoted a special issue to R packages for competing risks and multistate models (Putter 2011). Most packages adopt the counting process perspective. Since February 2009 Allignol and Latouche have maintained an R Cran Task View on Survival Analysis that gives a comprehensive overview of R packages for the analysis of time-to-event data. The Task View includes multistate models (http://cran.r-project.org/web/views/Survival.html; accessed 28th March 2014). In the next paragraphs we review packages for multistate modeling. Unless otherwise stated, the packages are in R and are included in CRAN. The packages covered include:

- ◆ Packages for survival analysis that are used in multistate modeling (*survival*, *eha*, *Epi*),
- ◆ Packages for non-parametric multistate models (*mvna*, *etm*, *msSurv*, *p3state.msm*),
- ◆ Packages for semiparametric multistate models (*mstate*, *p3state.msm*), and
- ◆ Packages for parametric multistate modeling (*msm*).

The *survival* package, developed in S by Therneau and ported to R by Lumley, was first published in 2001 (see Therneau 1999, p. 17; Lumley 2004; 2007, p. 40). The methods implemented in *survival* are described in Therneau and Grambsch (2000). The package is among the most popular R packages, and functions of the *survival* package are used in a large number of other packages. A counting process perspective is adopted and the Andersen-Gill data format is used. The package includes a utility to convert data into the Andersen-Gill format. The result is a survival object, which is used in many of the functions of the *survival* package. The Cox model and extensions of the Cox model can be estimated from observations that are left-truncated and right-censored. As a

consequence, the package can be used to estimate multistate models, provided a semi-parametric Cox model gives an adequate description of the transition rates. The function `coxph` estimates the Cox model. The function `survreg` facilitates estimating parametric hazard functions such as Weibull models, although it does not accept left-truncated observations. Frailty terms can be added in the `coxph` and `survreg` functions.

The *eha* (event history analysis) package was developed by Broström and first published in 2003 (Broström 2102a). The emphasis is on Cox regression with parametric baseline hazard and other extensions. Unlike *survival*, *eha* can handle parametric survival models and accelerated failure-time models with left-truncated and right-censored data. The distributions included are: Weibull, lognormal, log-logistic, Gompertz, and extreme value. In his book Event History Analysis with R, Broström (2012b) describes the methods implemented in *eha*, with many illustrations. Mills (2011) covers *survival* and *eha* in an introduction to event-history analysis using R.

The *Epi* package (Carstensen et al. 2013) contains functions for drawing Lexis diagrams, fitting age-period-cohort models, and estimating multistate models. Carstensen and Plummer (2011) describe how to use these concepts in multistate modeling.

Some packages are for non-parametric estimation of transition rates. The transition rates are not restricted to following an imposed pattern, such as being constant or increasing exponentially. They may vary freely. Without any parametric assumption it is easier to estimate cumulative transition rates than to estimate transition rates. This is akin to estimating the cumulative distribution function, which is easier than estimating the density function (Aalen et al. 2008, p. 71). The non-parametric approach is to estimate the cumulative transition rate any time a transition occurs. The transition rate is the ratio of the event count at that time and the number of individuals at risk. The estimator is known as the Nelson-Aalen estimator (Aalen et al. 2008). The *mvna* package, developed by Allignol and first published in 2007, computes the nonparametric Nelson-Aalen estimator of the cumulative transition rates and its standard errors for multistate models (Allignol 2012). The method is described in Allignol et al. (2008, 2011). The package uses event data, contrary to other packages that use episode data in the Andersen-Gill format. For each event or transition the input data consist of the origin state, the destination state, and the date of transition. The package also allows the estimation of transition probabilities. The method used for this purpose is similar to the method employed to obtain the empirical survival function (Kaplan-Meier estimator). In 2008 Allignol published the *etm* package to estimate the empirical transition probability matrix of multistate models and its variance-covariance matrix (see also Allignol et al. 2011). The package implements the Aalen-Johansen

estimator of transition probabilities (Aalen et al. 2008). Additionally, the *etm* package provides state occupation probabilities, computed from the transition probabilities[4].

The *msSurv* package, published in 2011 by Ferguson et al. (2013), also generates nonparametric estimates of cumulative transition rates from left-truncated and right-censored data, using the theory of counting processes. Furthermore, it produces non-parametric estimates of continuous-time stochastic processes that are non-Markovian. In non-Markovian processes, transition rates depend on the history of the process; for instance, on the duration of stay in the state occupied before the transition. Bootstrapping is used to estimate the variance of the cumulative transition rates for data with dependent censoring. The package also provides non-parametric estimators of state occupation probabilities, state entry time distributions and state exit time distributions for interval-censored data. The method is described in Ferguson (2011) and in Ferguson et al. (2012).

Putter, de Wreede, and Fiocco (Putter et al. 2012; de Wreede et al. 2010, 2011) developed the *mstate* package. It was first published in 2009. A semiparametric model is used to estimate transition rates: it relies on functions from the *survival* package to estimate Cox models. Two features of *mstate* are of particular interest: model specification and prediction.

Model specification

Different transition rates in the multistate system can be estimated simultaneously while at the same time covariates may differ between transitions. The package generates transition-specific covariates and the expanded covariate data set is used in Cox models. The model specification allows testing for covariate effects on transition rates. The following questions can easily be answered using the *mstate* software. Does the effect of education on the transition rate vary between transitions? Does the fertility rate differ between people with the same characteristics but different living arrangements, e.g., when one group is married and the other cohabits?

Prediction

Using subject-specific transition rates and the Cox model, the package gives estimates of subject-specific transition probabilities and the associated variance-covariance matrices. Both are based on the Aalen-Johansen estimator. Like the *etm* package, *mstate*

---

[4] The package *etm* incorporates features offered by an earlier package, *changeLOS*, published in 2005 by Wangler and Beyersmann (2012). In 2012 the *changeLOS* package became deprecated. All the functionalities can now be found in the *etm* package maintained by Allignol.

also provides state probabilities: unlike the *etm* package these may be based on models including covariates.

The methods are described in Putter et al. (2007) and Putter (2014). Functions for simulating trajectories may be used to obtain transition probabilities in certain non-Markovian multi-state models (see Fiocco et al. 2008).

The *dynpred* package, a companion package to the book Dynamic Prediction in Clinical Survival Analysis (van Houwelingen and Putter 2011), provides software for alternative methods to multistate modeling to obtain transition probabilities, based on landmarking (van Houwelingen and Putter 2008).

The *msm* package, developed by Jackson (2011, 2014a) and first published in 2002, adopts the parametric method for estimating transition rates in multistate models. The package fits continuous-time Markov models assuming that the instantaneous transition rates or transition intensities are constant or piecewise constant. The data may be panel data (i.e., repeated observations) with observations collected at arbitrary observation times or event data (i.e., the dates of transitions are observed). In the case of panel data, transitions take place at unknown times between observation times. By default, *msm* estimates transition rates from information on state occupancies at different points in time (panel data). The package uses the maximum likelihood method developed by Kalbfleisch and Lawless (1985) for the analysis of panel data and the application of that theory by Wolf (1986) and Laditka and Wolf (1998). The method is closely related to the inverse method proposed by Singer and Spilerman (1976). Since subjects are observed at discrete points in time, the contribution to the likelihood of pairs of observations given the continuous-time model is the transition probability. The estimation method implies a particular relation between transition rates (model parameters) and transition probabilities. In *msm* it is assumed that the transition rates are constant or piecewise constant. The package *msm* also accommodates knowledge of exact times at transition. Note that *msm* does not derive the transition rates from event counts and exposures, since exposure times are missing. Thus, the method does not follow a counting process perspective (see the likelihood function in Jackson 2014b Section 1.4). The package supports a variety of observation schemes. A particular feature of *msm* is the estimation of hidden Markov models, which accommodates state misclassification. For an application, see Van den Hout et al. (2014). In addition, the package includes a microsimulation utility that simulates Markov processes with piecewise-constant rates that may depend on time-varying covariates described by a step function that remains constant in-between observation times.

Van den Hout (2013) uses transition rates estimated by the *msm* package to estimate expected state occupation times, i.e., expected sojourn times in each of the states. The package that transforms transition rates into occupation times is *ELECT*. The package is written in R but is not available in CRAN; it can be downloaded from

the author's personal website www.ucl.ac.uk/~ucakadl (accessed March 20, 2014). For some of the applications, see Van den Hout et al. (2009).

In 2007 Meira-Machado et al. published *tdc.msm* for the estimation of Cox models with time-dependent covariates. It is written in R but it is not part of CRAN. It is available from http://w3.math.uminho.pt/~lmachado/ (accessed March 20, 2014). For a description of the package, see Meira-Machado et al. (2009). In 2009 Meira-Machado and Roca-Pardinas (2011, 2012) published the package *p3state.msm* for the analysis of survival data from the progressive, i.e., irreversible, illness-death model. That package is included in CRAN. It provides non-parametric estimates of transition rates in multistate models with three states. Popular examples include the illness-death model (three states: healthy, ill, dead) and the progressive multistate model (transitions in one direction between stages of a process). The package fits semi-parametric Cox models with covariates (possibly time-varying described by step functions). The illness-death model is used in survival analysis to determine the influence of an intermediate event on survival. The intermediate event may be the onset of a chronic disease, a medical treatment (e.g., a therapy or transplantation), or another event. The occurrence of the intermediate event implies that the value of a time-varying covariate changes[5]. In the illness-death model, time-varying covariates are (re-)expressed as a multistate model with states based on the value of the covariate. A distinguishing feature of *p3state.msm* is its ability to obtain estimates for transition rates in non-Markovian models, with three states.

In 2011 Willekens published *Biograph*, a package in R designed to facilitate (1) the description and exploratory analysis of life-history data, (2) the estimation of multistate models, and (3) the computation of a multistate life table (Willekens 2012, 2014). *Biograph* tracks and documents life histories of individuals and cohorts or other groups of individuals. It counts transitions during specified periods, measures observed state occupation times (sojourn times), and computes various user-defined life history indicators. Being in a state exposes the individual to the risk of a particular transition. *Biograph* extracts from the data how many subjects with given attributes are at risk of a given transition at a given point in time and how long they are at risk during a given period. It computes occurrence-exposure rates from transition counts and exposure times. *Biograph* assumes that transition rates are constant during the specified periods. If the periods are age-intervals the transition rates are age-specific rates. They belong to the category of piecewise-constant rates. *Biograph* does not produce variances of the occurrence-exposure rates. It provides links to several packages for multistate modeling

---

[5] The effect of time-dependent covariates in the Cox model may also be estimated using the *survival* package (for details, see Therneau and Crowson 2014).

that generate variances of transition rates. A useful feature of *Biograph* is the visualization of life histories and state occupation diagrams.

A distinguishing feature of *Biograph* is the link it provides to several of the packages for multistate modeling listed above. Most packages rely on the Andersen-Gill data format for input data, but other packages demand different input data formats. For some of the packages listed, *Biograph* includes functions to transform the input data into the required format. That utility facilitates the application of the packages for multistate models. Input objects are produced for the following packages: *survival*, *eha* (same as *survival*), *mstate*, *mvna* and *etm*, and *msm*. *Biograph* also produces input data objects for *TraMiner*, a package for sequence analysis (Gabadinho et al. 2011).

The packages reviewed so far were developed in R. A few software tools for multistate modeling are implemented in other languages. They are listed in the remainder of this section. The *MKVPCI* package, developed by Alioum and Commenges (2001), has much in common with the *msm* package. It aims at estimating a multi-state continuous-time Markov model with piecewise constant transition rates from panel data when the number of intervals does not exceed three. To estimate the transition rate in an interval, an artificial time-dependent covariate is constructed. It has a value of one if an event occurs during that interval and zero otherwise. The model is a Cox model and the regression coefficient associated with a given transition in a given interval is the relative risk of making that transition. The relative risk is the ratio of the transition rate in that interval and the transition rate in the interval that is chosen as the reference interval. The authors refer to the model as a modified homogeneous Markov model. The package is an extension of the *MARKOV* package, developed by Marshall et al. (1995). The program provides maximum-likelihood estimates and standard errors. It is written in Fortran 77 and available at http://etudes.isped.u-bordeaux2.fr/ BIOSTATISTIQUE/MKVPCI/US-Biostats-MKVPCI.htm (accessed 21 March 2014). The software of the *MARKOV* package is now also available in the R package *SmoothHazard* (Touraine et al. 2013).

Schoumaker (2013) presents a Stata module for the estimation of transition rates. The module tracks transitions (events) and exposure times during specified periods. Transition rates are estimated from occurrences and exposures. The method consists of three steps. In the first step the observation period is split into unit intervals. For each respondent and unit interval, the number of transitions (of a given type) and the exposure time are determined by counting events and durations. The result is a person-period data file, showing for each respondent and unit interval the number of transitions and the exposure time. In the second step the person-period data are aggregated into a table of transition counts and exposure times. In the third step, transition rates and their variances are estimated using the Poisson regression model. To delineate unit intervals, age and calendar time are used. By controlling exposure time in an offset (a variable

whose coefficient is equal to one), the Poisson model becomes a log-linear model. The method was first proposed by Laird and Olivier (1981). A similar three-step approach is used by Cai et al. (2010) in the *SPACE* package (see below).

In 2000 Brouard and Lièvre published *IMaCh* (Interpolated Markov Chains), a package to estimate discrete-time Markov models from panel data. The latest version of *IMaCh* was published in July 2009 (Brouard and Lièvre 2009). The package is written in C. *IMaCh* estimates transition probabilities from panel data. The interval length between survey waves may vary between waves and subjects. The method is based on the theory for the estimation of Markov models from panel data (Kalbfleisch and Lawless 1985) and the application of that theory by Wolf (1986) and Laditka and Wolf (1998). To estimate transition probabilities, *IMaCh* focuses on the discrete-time Markov chain embedded in the Markov process (eMC approach). The eMC approach recognizes that panel data often hinder the detection of multiple events between follow-up interviews. To recover possible missing events, the eMC approach considers shorter transition periods (e.g., monthly, quarterly, etc.) that are 'embedded' within the longer interval between follow-ups. Wolf and Gill (2009) assessed the performance of the eMC method. They found that the estimates of transition probabilities are sensitive to the length of the intervals between waves. For example, the eMC method underestimates monthly transition probabilities. *IMaCh* estimates an illness-death model (3 states: healthy, disabled, and dead). It requires data that consist of month and year of interviews (for each panel wave) and the corresponding health state. The method is described by Lièvre et al. (2003). *IMaCh* uses multinomial logistic regression for estimating annual age-specific transition probabilities from discrete-time transitions over arbitrary time intervals. On the basis of the estimated transition probabilities, *IMaCh* computes multistate life table (MSLT) indicators as well as corresponding variance estimators. The latter are obtained by using the Delta method. One of the indicators computed is the expected duration in each state. Notice that the duration is estimated directly from transition probabilities and not from transition rates, which is the more common method. *IMaCh* was especially designed for estimating health expectancies and the package is used regularly for that purpose. Main users include members of the REVES network, an international network established in 1989 to promote the use of health expectancy as a population health indicator[6]. For a list of publications that use *IMaCh* see the list of references on health expectancy at www.revesbiblio.eu (search for "imach") (Accessed 21 March 2014).

All packages listed up to now rely on transition data. In such data, for each individual either the exact age or date of transition is known (event data), or the states

---

[6] http://reves.site.ined.fr/en/home/about_reves/ (Accessed 21 March 2014).

occupied at consecutive ages are known (panel data). When transition data are incomplete, other estimation methods should be used to obtain transition rates. A useful method is implemented in the *DisMod* software package for multistate modeling, developed in the field of chronic disease epidemiology (Barendregt et al. 2000, 2003). It was developed under contract with the World Health Organisation (WHO) for the Global Burden of Disease Study (Murray and Lopez 1997). The idea is to derive transition rates from available data and logical relations in the disease process. For instance, it is impossible to recover or die from a disease without having had the disease. In addition, it is assumed that the mortality due to a disease is not affected by the presence or absence of other diseases (no co-morbidity). In that case healthy and diseased people have the same rate of dying from other causes than the disease. The model is an illness-death model with three states and two competing risks of dying: the disease being studied and all other causes of death. The four transitions are: disease incidence, remission (recovery from the disease), case fatality (death caused by the disease), and 'all other causes' mortality. If the cause-specific mortality rate is not known, as is often the case, then the transition rates are estimated from the overall mortality rate, the disease prevalence, and the remission or recovery rate, or case fatality and the relative risk of total mortality. Since prevalence of a disease, total mortality, and relative risk of dying from a disease are given, the disease-specific mortality may be derived. *DisMod* produces multistate life tables, yielding expected years with and without the disease. Confidence intervals of the indicators are obtained by parametric bootstrapping, which draws samples from probability distributions of the disease input variables previously specified. The current version (*DisMod II*) has a graphical user interface. It is available from the website of the WHO http://www.who.int/healthinfo/global_burden_disease/tools_software/en/ (accessed March 21, 2014). The method developed by Barendregt et al. was recently applied by Majer et al. (2012) in a study of health expectancy forecasting. The data consist of overall mortality, disease prevalence, and mortality hazard ratio. *DisMod* is being developed further. Recently it was redesigned at the Institute for Health Metrics and Evaluation, University of Washington, for the Global Burden of Disease Study 2010. The tool that results is referred to as a Bayesian meta-regression tool and the package as *DisMod-MR*. The MR tool allows users to specify different types of expert priors to estimate the proportions of major causes of death in the total number of deaths (see supplement to Vos et al. 2012). The Global Burden of Disease Study 2010 (GBD 2010), and its successor GBD 2013, are systematic efforts to describe the global distribution and causes of a wide array of major diseases, injuries, and health risk factors. The results of GBD 2010 were published in the Lancet in a series of seven articles, with comments from WHO Director-General Margaret Chan and World Bank President Jim

Yong Kim (see http://www.thelancet.com/themed/global-burden-of-disease; accessed 21 March 2014).

3. Multistate life tables and multistate projections

Multistate life tables present life history indicators that summarize how members of a synthetic birth cohort move between states, how distributions between states change with age, and how much time is spent in each state. The indicators include transition probabilities, state probabilities, and expected state occupation times. Indicators are easier to interpret than transition rates or transition probabilities. The stochastic process underlying the life table is usually a continuous-time Markov process. Multistate projections reveal the impact of transition rates on population dynamics. Multistate life tables and multistate projections rely on transition rates or transition probabilities, estimated from census data, vital statistics, survey data, or follow-up studies. Census data and vital statistics typically produce aggregate, tabulated data and transition rates are estimated as occurrence-exposure rates, which relate event counts during an interval to estimates of exposure times during the same interval. Exposure time is often not observed: in that case it is approximated by the product of mid-period population and period length.

This section has two parts. Multistate life-table software is presented first. Projection software is presented next. Most packages were developed in demography. Unless stated otherwise, the packages reviewed do not provide variances of the life-table indicators or uncertainty measures of projections.

3.1 Multistate life tables

Many software tools produce multistate life tables (MSLTs). Some were developed quite a long time ago and others are more recent. Older software packages often served as the basis for more recent ones. Palloni (2001) gives a brief overview of software tools for multistate life tables. Older software packages produce MSLTs from transition rates based on vital statistics and census data. They include *SPA* (Spatial Population Analysis) by Willekens and Rogers (1978) and *LIFEINDEC* by Willekens (1979). The packages were developed at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria, under the leadership of Andrei Rogers. The method is described in Willekens and Rogers (1978). The programs are in Fortran 77 and the source code is included in the related publications. The *SPA* program was extended by Rogers (1995) into the *SPACE* program (not to be confused with the *SPACE* program

developed by Cai et al. 2010). The *SPACE* program was included (on diskette) in Rogers (1995). Schoen (1988) published the *IDLT* and *IFLT* programs for the multistate life table. The programs derive MSLTs from event counts and exposures, generally approximated by the product of the mid-period populations and the period length. They are in Fortran 77 and the code is included in the book. *LIFEHIST* is a package, also in Fortran (initially Fortran 77 and later Fortran 90), developed by Rajulton (1985, 1992, 2001). The package implements the Markov process model and the discrete-time semi-Markov model. Tiemeyer and Ulmer (1991) developed a MSLT program in C++. Zeng (1991) developed a multistate life table program for family dynamics (*Famy*), initially in Fortran and later ported to C and C++. *Famy* simulates changes in nuclear (two generation) and extended (three generation) families. The program is based on the *FAMTAB* program (Bongaarts 1987). *Famy* and *FAMTAB* are no longer maintained and are therefore not discussed further. *Famy* had been replaced by the projection program *ProFamy* (see below).

lxpct_2, developed by Weden (2005), implements the multistate life table in Stata. It calculates multistate life tables from age-specific transition probabilities. The source code is available. The procedure was applied by Diehr et al. (2007a, 2007b) using data from the US Cardiovascular Health Study, to determine the number of sick persons in a cohort and to estimate health needs.

Lynch and Brown (2005a, 2005b) use a Bayesian perspective on MSLTs and develop a Gibbs sampler for multistate life tables (*GSMLT* v.90) that estimates transition probabilities using a constrained discrete-time multinomial probit model. To generate feasible multistate life tables, a large number of samples are drawn from the posterior distribution of the transition probabilities (by covariate profile). The software consists of two programs. *MSTATEHAZARD.C* is for multivariate (2 states in the current version) hazard model estimation and *MSTATETABLES.C* is for multistate life-table construction. The second program uses the output from *MSTATEHAZARD.C*. By applying multistate life-table calculations to samples of the distribution of the hazard model parameters, a series of life tables arise. The program was rewritten in R (Lynch and Brown 2010); the program can be obtained from Scott Lynch at Princeton University.

Cai et al. (2010) published the *SPACE* package (Stochastic Population Analysis for Complex Events), written in SAS. The package is available at www.cdc.gov/nchs/data_access/space.htm (accessed 21 March 2014). The aim of *SPACE* is to derive consistent interval estimates of life-table indicators from complex sample surveys. *SPACE* uses panel data and allows at most one event between two successive observations. The authors recognize that the assumption of at most one transition between two successive observations ("the traditional event history approach", Cai et al. 2010, p. 132) is a controversial one. The assumption introduces a bias (Wolf and Gill

2009). The package uses a multinomial logistic regression model to estimate transition probabilities from the data. Transition probabilities are age-specific. Two types of MSLT models can be fitted in the SPACE program: one follows a first-order Markov chain where the transition probabilities are conditional only on current age and status, and another follows a time-inhomogeneous semi-Markov process (SMP) model where the transition probabilities depend on current state, duration in the current state, and age. The theory of the semi-Markov process model is presented in Cai et al. (2006).

A distinct feature of *SPACE* is the use of microsimulation as the primary means of computation of MSLT functions. *SPACE* generates individual life histories from the transition probabilities (micro-simulation in discrete time). Life-table measures, such as expected sojourn times in each state and life expectancy, are computed from the simulated life histories. Bootstrapping is applied to estimate the variances of the differences between life-table measures for different groups. The method involves taking a sample with replacement from the initial sample, taking into account the sample design (e.g., stratification and sample weights). The method allows for consistent variance estimates from complex survey samples that include stratification and multi-stage clustering. *SPACE* is the first package for multistate modeling that uses a virtual population (simulated life histories) to derive indicators of population health. Since individual life histories are used, many more indicators can be computed than in conventional approaches. Cai et al. (2010) compare *SPACE* with *IMaCh* and *GSMLT*.

SPACE also estimates one-year transition rates using a log-linear model. The authors convert observation intervals of uneven lengths to observation intervals of one year to facilitate the computation of annual transition parameters. If the length of the interval between interviews is two or more years, then a corresponding number of annual intervals are created. The events are assumed to occur randomly in one of these annual intervals. This approach results in the events occurring, on average, in the middle of the observation interval. Notice that the assumption of at most one transition during the observation unit leads to an underestimation of transition rates. The method produces person-year data that can be used in the log-linear model (with offset), a method first proposed by Laird and Olivier (1981).

3.2 Multistate projections

Multistate population projection models describe how populations change as a result of mortality rates, fertility rates, and rates of transition between states. A population is usually stratified by birth cohort and the projection model is known as the cohort component model. Fertility and mortality rates vary by state occupied and they may vary in time and may depend on covariates. Most multistate projection models describe

the population dynamics at the aggregate (population) level: they project the number of people in each state. Some models generate individual life histories. They are microsimulation models. In microsimulation, population figures are aggregations of individual life histories. Most models covered in this section are population models, sometimes also referred to as macro-simulation models. Three microsimulation models are discussed: *Modgen*, *Moses,* and *MicMac*. The recently published package *MicSim*, which is based on *MicMac*, is also included.

The first published software for multistate projection was developed in the 1970s at IIASA. The work was developed further at the Netherlands Interdisciplinary Demographic Institute (NIDI) in The Hague. That resulted in *MUDEA*, *LIPRO,* and *Famy*. *MUDEA* was designed for regional population projection and was implemented at the National Physical Planning Agency in the Netherlands. It was written in Fortran 77. For the method, see Willekens and Drewe (1984) and Willekens (1995). For a description of the MUDEA project, see Drewe et al. (1991) and for an overview of applications, see Muhidin (2002, Chapter 6). *MUDEA* is no longer available.

Van Imhoff and Keilman developed *LIPRO* (Van Imhoff and Keilman 1992). *LIPRO* was originally developed for multistate household projection but has been designed as a general-purpose multistate projection model. It is user-friendly and has been applied in several European countries. The most recent version is 4.0, released in 2005 (Van Imhoff 2005). *LIPRO* contains a multistate life table module. Most computer programs developed in the 1990s or earlier implemented the so-called linear model for estimating multistate life tables. The linear model assumes uniform distribution of transitions within age intervals, and hence a piecewise linear survival function. An alternative assumption is that the transition rate is constant within age intervals, resulting in piecewise exponential survival functions (van Imhoff 1990). *LIPRO* implements both the linear and the exponential model. *LIPRO* is available at the NIDI website: http://www.nidi.knaw.nl/en/research/al/270101 (accessed 21 March 2014).

ProFamy (Zeng et al. 1997, 1998) is a multistate model for families and households. In *LIPRO* and *ProFamy* the individual is the unit of analysis. *LIPRO* considers transitions between household types. In *ProFamy* transitions between household types are derived from demographic events such as leaving the parental home, cohabitation, marriage, childbearing, and divorce. As a consequence, *ProFamy* uses conventional demographic data that are readily available from surveys, vital statistics, and censuses. Because of its reliance on demographic rates, *ProFamy* provides a tool to assess which of the demographic changes in marriage, divorce, fertility, and so on, affect family households most. Zeng et al. (2006) extended the *ProFamy* model by including race. For a description of the method implemented in *ProFamy*, see Zeng et al. (2013). Zeng et al. (2014) is the most extensive coverage of *ProFamy*. The book includes a detailed description of the method, household

projections for the United States and China, a documentation of the software, and a user's manual.

In *LIPRO* and *ProFamy*, several transitions in the family and household life cycle involve more than one individual. For example, cohabitation, marriage, and separation involve two people. When a child leaves the parental home the household characteristics of other members need to be updated. This poses consistency problems. *LIPRO* and *ProFamy* incorporate algorithms to ensure consistency.

At IIASA, Scherbov et al. (1986, 1988) developed *DIALOG / DIAL* for multistate population projection.

In the 1970s at the University of Leeds, Rees developed a system of multistate demographic accounts (Rees and Wilson 1977; Rees 1981) and applied the method in spatial demography (multiregional models). A population account is a two- or multi-dimensional table that integrates data on population change in a systematic and consistent way, even if the data originate from different sources. It integrates information on population stocks and flows, and demographic events such as births, deaths, and migration that cause population change (see e.g., Willekens 2011a). Accounting methods are methods that produce internal consistency. The method goes back to sir Richard Stone, who in the 1940s developed national income accounts and socio-demographic accounts. Rees developed computer programs that implement the accounting method in population projection. More recently, Wilson and Rees (Wilson 2001; Wilson and Rees 2003) used the multiregional demographic accounting method to develop UKPOP, a population projection model for the United Kingdom. Wohland et al. (2010, 2012) used a similar accounting approach to develop a bi-regional model for ethnic population projection for local areas in the UK (16 ethnic groups and 352 local areas). A bi-regional model considers migration between a given region and the rest of the country. The bi-regional model gives results that are not much different from those produced by a full multiregional model. The model is coded in R, but is not made available. Wilson (2011) uses a similar accounting and bi-regional approach in The New South Wales Demographic Simulation System (NEWDSS). NEWDSS consists of a pre-processor for preparation of input data, a processor to project the population, and a post-processor for tabulation, graphic presentation, and analysis of results. The processor is in Fortran 95. Excel is used for data preparation and output processing.

Kupiszewski and Kupiszewska (2010, 2011) developed *MULTIPOLES* for multiregional population projection. MULTIPOLES is a cohort-component, hierarchical, multiregional, supranational model of population dynamics. It is the first hierarchical multistate model, with three levels: subnational, national, and supranational (a group of countries, such as the European Union). Migration flows are assumed to be consistent at all levels. For the first and second level, migration flows are by origin and destination. Migration with the rest of the world is in terms of net migration between a

country (of the multi-country system) and the rest of the world. MULTIPOLES is based on the demographic accounting method developed by Rees (Rees and Wilson 1977). Migration is measured as a change of address and not, as mostly done in the approach of Rees, by comparing places of residence at two points in time. The accounts are movement accounts and not transition accounts. For a description of MULTIPOLES and applications to Europe, see Kupiszewski (2013).

In recent years multistate projection models have been extended to microsimulation models and microsimulation models are currently being extended to agent-based models. The use of microsimulation in population projection is not new (Van Imhoff and Post 1998). What is new is the extension of the multistate cohort-component model to accommodate microsimulation, with the cohort-component model yielding expected values of population size and composition, and microsimulation producing probability distributions around these expected values. *SPACE* (Cai et al. 2010) integrates microsimulation in multistate life tables (see above). In the remainder of this section, three utilities for microsimulation are reviewed: *Modgen*, *MicMac,* and *Moses*. *Modgen* is a generic language for dynamic microsimulation. It is used in several applied microsimulation models. The *MicMac* model integrates the multistate cohort component model and microsimulation. The *Moses* model extends microsimulation models by transcending transition probabilities and transition rates and incorporating in microsimulation models theoretical insights on demographic behavior. *Moses* is a microsimulation model that incorporates important features of agent-based modeling (ABM).

Modgen (Model generator) is a generic microsimulation programming language developed by Steve Gribble at Statistics Canada[7]. It is written in C++. It supports the creation, maintenance, and documentation of dynamic microsimulation models. Several types of model can be accommodated in continuous or discrete time, and with interacting or non-interacting populations. The best-known microsimulation models of Statistics Canada are coded in *Modgen*. They include (i) *LifePaths*, which generates individual life histories, including health trajectories, employment trajectories, and earnings trajectories, (ii) the population health model *Pohem*, and (iii) the population projection microsimulation model *Demosim* (originally called *Popsim*). *Pohem* and *Demosim* are variants of the *LifePaths* model. They include the same main assumptions on the shape and the dependence structure of transition rates. *Demosim* projects the population of Canada by visible minority group, religious denomination, mother tongue, place of birth, generation status, period of immigration, level of education, and

---

[7] *Modgen*'s home page: http://www.statcan.gc.ca/microsimulation/modgen/modgen-eng.htm (accessed 21 March 2014).

participation in the labor force[8]. Spielauer (2009) developed *Riskpaths* as a teaching tool. It is based on *Modgen* and was developed in the context of the European Doctoral School of Demography (EDSD). Wolfson and Rowe (2014) use Modgen to simulate health trajectories in continuous time to quantify the relative importance of selected health determinants. All software products based on *Modgen* are listed on the *Modgen* website and reviewed by Décarie et al. (2012). Statistical methods implemented in the *Modgen* language are not fully documented, which is a weakness. The documentation presented in this section is based on the description given in the *Modgen* Developers Guide (Statistics Canada 2009), on the description of LifePaths by Rowe and Moore (2009), the Statistics Canada website[9] and Spielauer (2006, 2009). *Modgen* is the main modeling technology at Statistics Canada.

Actors are the unit of analysis in *Modgen*. Actors are individuals or organizations. Two types of individuals are distinguished: person and child. An individual is defined by attributes, such as sex, marital status, and health status. A combination of attributes is a state. Individuals move between states. In the continuous-time model the waiting time to a transition is determined by transition rates. The transition rate that applies to an individual is a function of the individual's characteristics. The user supplies the hazard function (called 'event function'). At any age, an individual has several attributes and each attribute may change in more than one direction. To determine which attribute changes and when, the latent failure time model of competing risks is used: a latent waiting time is determined for all possible events (event queue) and the event with the lowest waiting time is selected[10, 11]. This method is common in discrete-event simulation models (DEVS), developed in system theory. Whenever an individual experiences an event or an explanatory factor changes, new waiting times are determined. Rowe and Moore (2009) describe microsimulations with transition rates that vary every month. As a result, random waiting times are generated at least every

---

[8] For information on *Demosim*: http://www.statcan.gc.ca/microsimulation/demosim/demosim-eng.htm (accessed 21 March 2014).
[9] In particular the report "What's new in 2011 at Statistics Canada – microsimulation related Population, Health and LifeCourse." http://sociology.uwo.ca/cluster/en/documents/StatisticsCanada, Microsimulation2011.pdf (Accessed 21 March 2014).
[10] *Modgen* allows for two or more attributes changing at the same time and hence two events occurring at that time (ties between events). A user-provided priority code determines which attribute changes first. If the priority code is absent, events are executed in alphabetical order of event name.
[11] For an alternative approach, see Beyersmann et al. (2009).

month. A transition does not occur unless the waiting time generated is less than one month[12]. When an event occurs to an individual, the time is updated for all individuals.

The user writes an application in *Modgen* language and *Modgen* translates the code into C++ using the Visual Studio 2008 C++ compiler. *Modgen* requires the Visual Studio 2008 environment to be installed on the computer of use, which may be a disadvantage for users. Parameters are submitted to *Modgen* in a text file.

The output of *Modgen* consists of aggregate results, stored in Microsoft Access files, Excel files, or text files, and individual life histories, stored in Microsoft Access (.mdb) files. The individual life histories are of particular interest because they can be displayed and they provide the flexibility to compute various population indicators. *Modgen* does not store entire life histories because that would take too much space. Before running *Modgen* users must specify what life history information they want to track and save in the *Modgen* database. The database facilitates "on-the-fly tabulation" of user-specified summary statistics in user-specified tables. The track command controls the actors and states to be saved. It acts as a filter. For instance, if one wants to track married males between ages 20 and 50, information during marriage episodes between ages 20 and 50 is written into the database. Information on other episodes (e.g., divorce) is omitted.

A stand-alone software product, the BioBrowser (*Modgen* Biography Browser), was developed to examine changes in personal attributes during the life course and to display life histories graphically. BioBrowser also allows displaying cross-sectional information on individuals at a given reference time. Top-down menus allow the user to select graphic displays. The BioBrowser accepts the *Modgen* database as an input file. Another stand-alone product, the *Modgen* Web interface, enables *Modgen* models to be viewed and run through a web browser.

MicMac consists of a multistate cohort-component model (*MacCore*) and a microsimulation model (*MicCore*) (Zinn and Gampe 2011). *MacCore* implements the multistate model described in Willekens (2005, 2007). It generates cohort biographies from age-and-period-specific transition rates. *MicCore* generates individual life histories from age- and period-specific transition rates. In *MicCore* all individuals are simulated simultaneously in continuous calendar time (Zinn et al. 2009; Willekens 2009). Individuals can enter the population by birth and immigration and leave by death and emigration (absorbing states). Individuals move freely between states. For individuals in the base population, life histories are generated starting in the base year.

---

[12] A similar method is used by Willekens (2009) but with episodes of one year. Random waiting times do not need to be generated more than every month if prospective cumulative transition rates are used (Zinn 2010, Willekens 2011b).

For immigrants simulation starts at the date of immigration, and for newborns it starts at the date of birth. *MicCore* offers the possibility that transition rates for immigrants differ from those applied to native members of the population. Simulation steps are discrete, from one event to the next. Time-to-event is measured in continuous time. The same approach is used in the *LifePaths* model. The approach is known as discrete-event simulation (DES or DEVS), which is the common method in the field of system theory (Zinn 2010; Zinn et al. 2010). The DEVS method is consistent with the theory of competing risks. To determine the transition that occurs and the waiting time to that transition, latent waiting times are computed for all possible transitions and the transition with the lowest waiting time is selected. Ages at transition are stored in milliseconds (a feature of the Java-based modeling and simulation framework JAMES II used in *MicMac*, see below). Calendar dates of transitions are derived by combining age and simulated dates of birth. If *MicCore* is used for projection, the states individuals occupy at given calendar dates are obtained from the individual life histories. Aggregation of individual data at a given calendar date gives population size and composition at that date. The microsimulation results are very similar to the cohort-component projections if the number of simulated life histories is sufficiently large (Zinn 2010; Willekens 2011b).

MicMac has three parts: a pre-processor to input the base population and the transition rates, a processor producing cohort projections (*MacCore*) and individual life histories (*MicCore*), and a post-processor to process the results. The pre-processor and post-processor are stand-alone products. They are written in R. *MacCore* and *MicCore* read the base population and transition rates from input text files. The projections (*MacCore*) and the simulated life paths (*MicCore)* are stored in two text files: a population file, which shows the dates of transitions, and a file of newborns generated by the microsimulation. Ogurtsova et al. (2009) describe the pre-processor. Gampe and Zinn (2011) describe the post-processor. For an application, see Ogurtsova et al. (2010). The processor is extremely fast. Millions of life histories are generated in seconds. *MicMac* uses the JAMES II modeling and simulation framework, developed since 2003 at the Computer Science Department of the University of Rostock (www.jamesii.org; Accessed 21 March 2014). JAMES II is a generic technology for building specialized modeling and simulation software products (Himmelspach 2012). It is written in Java and is available as open source. Among many other things, JAMES II provides the functionality to implement and execute DEVS (Discrete Event Simulation) models. It is based on a 'Plug'n simulate' concept (Himmelspach and Uhrmacher 2007). This concept enables the developer to integrate novel components (as plug-ins) into an existing framework. *MicCore* comprises two JAMES II plug-ins: a modeling plug-in and a simulation plug-in. It is written in Java. *MacCore* and *MicCore* are parts of an executable Java program called MicMacCore.jar. *MicMacCore* features a graphical user

interface and can be run without further ado. It only requires the Java Virtual Machine to be installed on the computer of use. For a description of *MicCore*, see Zinn et al. (2013). Recently, Zinn (2014a) implemented a version of *MicCore* that is entirely in R. The package is called *MicSim* and is available on CRAN. Its runtimes are not as good as the runtimes of *MicCore* because *MicSim* is entirely written in R and does not use routines implemented in Java, C++, or Python. To cope with that issue the package includes a function (*micSimParallel*) for parallel computing, using more than one CPU. For a description of *MicSim*, see Zinn (2014b).

Moses (Birkin et al. 2009; Wu and Birkin 2013a, 2013b; Wu et al. 2011) is a spatial microsimulation model of the UK population. It considers six demographic processes: aging, mortality, fertility, household formation, health change, and migration. The spatial unit is a census ward (about 11,000 in the UK). It generates individual life histories from empirical transition probabilities. The simulation operates in discrete time intervals of one year, i.e., for each individual, attributes are updated once a year. The microsimulation runs from the base year 2001 until 2031. *Moses* aims at combining the strengths of dynamic spatial microsimulation models and agent-based modeling. The Agent Based Model (ABM) represents individual decision-making units as interacting agents with built-in intelligence. Such intelligence will then guide agents to make decisions and take actions during their interactions with other agents and the environment they live in, according to their individual attributes and rules. The simulated output then becomes the base-population for the next year's simulation so that the impact of events and conditions during the previous year can be captured within the model. Each component of change is constructed separately but individual components can also interact with each other during the simulation. For instance, in many cases household formation leads to migration. Ties between events and between individuals are handled in the agent-based model. In the ABM, individuals are agents with an agency (capability to decide and act). They can stop following 'external rules' imposed by the microsimulation and adopt unique built-in 'internal rules' that depend on individual characteristics and preferences. For instance, individuals aged 19 to 30 in full-time education are modeled as "university student agents". In the migration process they only stay in a city for a certain period of time according to their higher education programs, then leave upon graduation. The model is applied to the urban area of Leeds in the north of England. The policy applications are health care, transportation planning, and public finance. For a technical description of *Moses*, see Townend et al. (2009).

4. Conclusion

Software for multistate analysis evolved rapidly for three reasons. The first reason is the increased availability of life-history data from retrospective life-history surveys and prospective panel surveys and follow-up studies. Data on transitions and populations at risk constitute the raw data for the estimation of transition rates and transition probabilities. The second reason is increased computer power, shared software environments, and open source policies. Over the years R emerged as the lingua franca of statistical computing. That trend is also manifest in software for multistate modeling. In the early stages of software development various computer languages were used, with Fortran as the dominant language. Today the majority of packages for multistate modeling are written in R and authors make the source code publicly available on the Comprehensive R Archive Network (CRAN). The object-orientation of R and its focus on functions performing particular tasks enhance reuse and sharing of software. The graphics capabilities of R stimulate visualization of state sequences and event histories. The third reason is the statistical theory of counting processes. The theory offers a powerful method for the estimation of transition rates and transition probabilities: it allows for multiple state entries and state exits during a period of observation.

Multistate modeling involves the estimation of transition rates and transition probabilities. Life-history indicators such as state occupation probabilities and expected state occupation times (sojourn times) are derived from these rates and probabilities. Covariates and unobserved heterogeneity can be incorporated using standard statistical techniques. Multistate modeling also involves the prediction of transition rates and the assessment of the consequences of changing rates for the remaining life paths of individuals, groups of individuals (e.g., cohorts), and the population. The cohort-survival projection model and microsimulation are combined to respond to the growing interest in population forecasts that account for population heterogeneity. Cai et al. (2010) were the first to take full advantage of the flexibility of subject-specific life histories (generated by microsimulation) over cohort biographies (generated by multistate life table). Statistics Canada was the first to switch entirely to microsimulation for population forecasting. Microsimulation enables the use of detailed information and updating information regularly, as the *Moses* model shows. The prediction of subject-specific life histories as a technique for incorporating idiosyncratic and up-to-date information on subjects is emerging as a central issue in a range of fields from medical statistics (Van Houwelingen and Putter 2011) to demography and public health (Cai et al. 2010) and geography (Birkin et al. 2009). The concern and the methodology of predictive modeling and fine-grained population forecasting seem to be converging. If that observation is true, then the software is likely to converge across disciplines too.

5. Acknowledgements

References

Aalen, O.O. (1975). Statistical inference for a family of counting processes. [PhD thesis]. Berkeley: University of California.

Aalen, O.O., Borgan, Ø., and Gjessing, H.K. (2008). *Survival and event history analysis. A process point of view*. New York: Springer. doi:10.1007/978-0-387-68560-1.

Allignol, A. (2012). *The mvna package. Nelson-Aalen estimator of the cumulative hazard in multistate models*. CRAN repository.

Allignol, A., Beyersmann, J., and Schumacher, M. (2008). mvna: An R package for the Nelson-Aalen estimator in multistate models. *R Newsletter* 8(2): 48–50.

Abbott, A. (2001). *Time matters. On theory and method*. Chicago: The University of Chicago Press.

Allignol, A., Schumacher, M., and Beyersmann, J. (2011). Empirical transition matrix of multistate models: the etm package. *Journal of Statistical Software* 38(4): 15.

Alioum, A. and Commenges, D. (2001). MKVPCI: a computer program for Markov models with piecewise constant intensities and covariates. *Computer Methods and Programs in Biomedicine* 64(2): 109–119. doi:10.1016/S0169-2607(00)00094-8.

Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N. (1993). *Statistical models based on counting processes*. New York: Springer. doi:10.1007/978-1-4612-4348-9.

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes. A large sample study. *Annals of Statistics* 10: 1100–1120. doi:10.1214/aos/1176345976.

Andersen, P.K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research* 11: 91–115. doi:10.1191/0962280202SM276ra.

Barendregt, J.J., Baan, C.A., and L. Bonneux, L. (2000). An indirect estimate of the incidence of non-insulin-dependent diabetes mellitus. *Epidemiology* 11: 274–279. doi:10.1097/00001648-200005000-00008.

Barendregt, J.J., van Oortmarssen, G.J., Vos, T., and Murray, C.J.L. (2003). A generic model for the assessment of disease epidemiology: the computational basis of DisModII. *Population Health Metrics* 1(4): 1–8.

Beyersmann, J., Latouche, A., Buchholz, A., and Schumacher, M. (2009). Simulating competing risk data in survival analysis. *Statistics in Medicine* 28(6): 956–971. doi:10.1002/sim.3516.

Beyersmann, J., Schumacher, M., and Allignol, A. (2012). *Competing risks and multistate models with R*. New York: Springer. doi:10.1007/978-1-4614-2035-4.

Bijwaard, G. (2014). Multistate event history analysis with frailty. *Demographic Research* 30(58): 1591–1620. doi:10.4054/DemRes.2014.30.58.

Birkin, M., Wu, B., and Rees, Ph. (2009). Moses: dynamic spatial microsimulation with demographic interactions. In: Zaidl, A., Harding, A., and Williamson, P. (eds.). *New frontiers in microsimulation modelling*. Farnham (Surrey): Ashgate: 53–78.

Blossfeld, H.P., Hamerle, A., and Mayer, K.U. (1989). *Event history analysis. Statistical theory and application in the social sciences*. Hillsdale, N.J.: Lawrence Erlbaum.

Blossfeld, H.P. and Rohwer, G. (2002). *Techniques of event history modeling. New approaches to causal analysis*. Mahwah, N.J.: Lawrence Erlbaum.

Blossfeld, H.P., Golsh, K., and Rohwer, G. (2007). *Event history analysis with Stata*. Mahwah, N.J.: Lawrence Erlbaum.

Bongaarts, J. (1987). The projection of family composition over the life course with family status life tables. In: Bongaarts, J., Burch, T.K., and Wachter, K.W. (eds.): *Family demography. Methods and their application*. Oxford: Clarendon Press: 189–212.

Broström, G. (2012a). *The eha package*. CRAN repository.

Broström, G. (2012b). *Event history analysis with R*. Chapman and Hall.

Brouard, N. and Lièvre, A. (2009). IMaCh: a maximum-likelihood computer program using interpolation of Markov chains [electronic resource]. http://euroreves.ined.fr/imach/.

Cai, L., Schenker, N., and Lubitz, J. (2006). Analysis of functional status transitions by using a semi-Markov process model in the presence of left-censored spells. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 55(4): 477–491. doi:10.1111/j.1467-9876.2006.00548.x.

Cai, L., Hayward, M., Saito, Y., Lubitz, J., Hagedorn, A., and Crimmins, E. (2010). Estimation of multi-state life table functions and their variances using the SPACE program. *Demographic Research* 22(6): 129–158. doi:10.4054/DemRes.2010.22.6.

Carstensen, B. and Plummer, M. (2011). Using Lexis objects for multi-state models in R. *Journal of Statistical Software* 38(6): 1–18.

Carstensen, B., Plummer, M., Laara, E., and Hills, M. (2013). Epi: A package for statistical analysis in epidemiology. R package version 1.1.49 [electronic resource]. http://CRAN.R-project.org/package=Epi.

Dabrowska, D.M., Sun, G.W., and Horowitz, M.M. (1994). Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. *Journal of the American Statistical Association* 89(427): 867–877. doi:10.1080/01621459.1994.10476819.

Décarie, Y., Boissoneault, M., and Légaré, J. (2012). An inventory of Canadian microsimulation models. Hamilton, Ontario: McMaster University, Social and Economic Dimensions of an Aging Population (SEDAP), SEDAP Research Paper No. 298.

de Wreede, L.C., Fiocco, M., and Putter, H. (2010). The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine* 99: 261–274. doi:10.1016/j.cmpb.2010.01.001.

de Wreede, L.C., Fiocco, M., and Putter, H. (2011). mstate: An R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software* 38: 7.

Diehr, P., Derleth, A., Newman, A.B., and Cai, L. (2007a). The number of sick persons in a cohort. *Research on Aging* 29(6): 555–575. doi:10.1177/0164027507305727.

Diehr, P., Derleth, A., Newman, A.B., and Cai, L. (2007b). The effect of different public health interventions on longevityy, morbidity, and years of healthy life. *BMC Public Health* 7: 52. doi:10.1186/1471-2458-7-52.

Drewe, P., Eichperger, L., Keilman, N. , Van Poppel, F., and Willekens, F. (1991). Multiregional population projection in the Netherlands. In: Pumain, D. (ed.). *Spatial analysis and population dynamics*. Montrouge and London: John Libbey: 59–74.

Ferguson, N. (2011). Methods and software for nonparametric estimation in multistate models. [PhD dissertation]. Louisville, Kentucky: University of Louisville. http://digital.library.louisville.edu/cdm/singleitem/collection/etd/id/2232/rec/1.

Ferguson, N., Datta, S., and Brock, G. (2012). msSurv: An R Package for Nonparametric Estimation of Multistate Models. *Journal of Statistical Software* 50(14): 1–24.

Ferguson, N., Datta, S., and Brock, G. (2013). *msSurv: nonparametric estimation of multistate models*. CRAN repository.

Fiocco, M., Putter, H., and van Houwelingen, H.C. (2008). Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistics in Medicine* 27: 4340–4358. doi:10.1002/sim.3305.

Flinn, C. and Heckman, J. (1982). New methods for analyzing structural models of labor force dynamis. *Journal of Econometrics* 18: 115–168. doi:10.1016/0304-4076(82)90097-5.

Gabadinho, A., Ritschard, G., Müller, N.S., and Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software* 40(4).

Gampe, J. and Zinn, S. (2011). The MicMac post-processor. Manual. Rostock: Max Planck Institute for Demographic Research.

Goldstein, R. and Harrell, F. (2005). Survival analysis, software. In: Armitage, P. and Colton, T. (eds.). *Encyclopedia of Biostatistics*. Chichester: Wiley.

Heckman, J. and Singer, B. (1985). Social science duration analysis. In: Heckman, J. and Singer, B. (eds.). *Longitudinal analysis of labor market data*. Cambridge: Cambridge University Press: 39–110. doi:10.1017/CCOL0521304539.

Heckman, J.J. and Walker, J.R. (1990). The relationship between wages and income and the timing and spacing of births: evidence from Swedish longitudinal data. *Econometrica* 58(6): 1411–1441. doi:10.2307/2938322.

Himmelspach, J. (2012). Tutorial on building M&S software based on reuse. In: Laroque, C., Himmelspach, J., Pasupathy, R., Rose, O., and Uhrmacher, A.M. (eds.). *Proceedings of the 2012 Winter Simulation Conference. Article 167*.

Himmelspach, J. and Uhrmacher, A.M. (2007). Plug'n simulate. In: ANSS '07. *Proceedings of the 40th Annual Simulation Symposium.* Washington, DC: 137–143. doi:10.1109/ANSS.2007.34.

Jackson, C. (2011). Multi-state models for panel data: the msm package for R. *Journal of Statistical Software* 38(8): 28.

Jackson, C. (2014a) *msm: Multi-state Markov and hidden Markov models in continuous time*. CRAN repository.

Jackson, C. (2014b) *Multi-state modelling with R: the msm package. Vignette distributed with the msm package*. CRAN repository.

Kalbfleisch, J. and Lawless, J.F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* 80: 863–871. doi:10.1080/01621459.1985.10478195.

Keiding, N. (2014). Event history analysis. *Annual Reviews of Statistics and its Application* 1: 333–360.

Kupiszewski, M. (ed.) (2013). *International migration and the future of populations and labour in Europe*. New York: Springer. doi:10.1007/978-90-481-8948-9.

Kupiszewski, M. and Kupiszewska, D. (2010). Demographic and Migratory Flows affecting European Regions and Cities. Multilevel scenario model. DEMIFER project report (Deliverable 4). Luxembourg: European Observation Network for Territorial Development and Cohesion (European Spatial Planning Observatory Network, ESPON). www.espon.eu/main/Menu_Projects/Menu_AppliedResearch/demifer.html.

Kupiszewski, M. and Kupiszewska, D. (2011). MULTIPOLES: A revised multiregional model for improved capture of international migration. In: Stillwell, J. and Clarke, M. (eds). *Population dynamics and projection methods*. New York: Springer: 41–60. doi:10.1007/978-90-481-8930-4_3.

Laditka, S. and Wolf, D.A. (1998). New methods for modeling and calculating active life expectancy. *Journal of Aging and Health* 10: 214–241. doi:10.1177/089826439801000206.

Laird, N. and Olivier, D. (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association* 76(374): 231–240. doi:10.1080/01621459.1981.10477634.

Lièvre A., Brouard, N., and Heathcote, Ch. (2003). Estimating health expectancies from cross-longitudinal surveys. *Mathematical Population Studies* 10(4): 211–248. doi:10.1080/713644739.

Lillard, L.A. (1993). Simultaneous equations for hazards: marriage duration and fertility timing. *Journal of Econometrics* 56(1/2): 189–217. doi:10.1016/0304-4076(93)90106-F.

Lillard, L.A. and Panis, C.W.A. (2003). *Multilevel multiprocess modeling. aML Version 2. User's guide and reference manual.* Los Angeles: EconWare. http://www.applied-ml.com/.

Lumley, T. (2004). The survival package. *The newsletter of the R project* 4/1(June 2004): 26–28. http://cran.r-project.org/doc/Rnews/Rnews_2004-1.pdf.

Lumley, T. (2007). *The survival package*. CRAN repository.

Lynch, S.M. and Brown, J.S. (2005a). Gibbs sampler for multistate life table software (GSMLT v.90). http://www.ined.fr/fichier/t_telechargement/25787/telechargement_fichier_fr_lynch_brown_gsmltmanual.pdf.

Lynch, S.M. and Brown, J.S. (2005b) A new approach to estimating life tables with covariates and constructing interval estimates of life table quantiles. *Sociological Methodology* 35: 177–225. doi:10.1111/j.0081-1750.2005.00168.x.

Lynch, S.M. and Brown, J.S. (2010). Obtaining multistate life table distributions for highly refined subpopulations from cross-sectional data: a Bayesian extension of Sullivan's method. *Demography* 47(4): 1053–1077. doi:10.1007/BF03213739.

Majer, I.M., Stevens, R., Nusselder, W.J., Mackenbach, J.P., and van Baal, P.H.M. (2012). Modeling and forecasting health expectancy: theoretical framework and application. *Demography* 50(2): 673–697. doi:10.1007/s13524-012-0156-2.

Marshall, G., Guo, W., and Jones, R.H. (1995). MARKOV: A computer program for multi-state Markov models with covariables. *Computer Methods and Programs in Biomedicine* 47(2): 147–156. doi:10.1016/0169-2607(95)01641-6.

Meira-Machado, L., Una-Álvarez, J., and Cadarso-Suárez, C. (2006). Nonparametric estimation of transition probabilities in a non-Markov illness-death model. *Lifetime Data Analysis* 12: 325–344. doi:10.1007/s10985-006-9009-x.

Meira-Machado, L.M., Cadarso-Suárez, C., and de Una-Alvarez, J. (2007). tdc.msm: an R library for the analysis of multi-state survival data. *Computer Methods and Programs in Biomedicine* 86: 131–140. doi:10.1016/j.cmpb.2007.01.010.

Meira-Machado, L.M. and Roca-Pardinas, J. (2011). p3state.sms: analyzing survival data from an illness-death model. *Journal of Statistical Software* 38(3): 1–18.

Meira-Machado, L.M. and Roca-Pardinas, J. (2012). *Package "p3state.msm"*. CRAN repository.

Meira-Machado, L.M., de Uña-Álvarez, J., Cadarso-Suárez, C., and Andersen, P.K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research* 18(2): 195–222. doi:10.1177/0962280208092301.

Mills, M. (2011). *Introducing survival and event history analysis*. London: Sage.

Muhidin, S. (2002). *The population of Indonesia: regional demographic scenarios using a multiregional method and multiple data sources*. Amsterdam: Rozenberg Publishers. http://irs.ub.rug.nl/ppn/238631389.

Murray, C.J. and Lopez, A.D. (1997). Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study. *Lancet* 349: 1347–1352. doi:10.1016/S0140-6736(96)07494-6.

Ogurtsova, K., Gampe, J., and Zinn, S. (2009). The MicMac pre-processor. Manual. Rostock: Max Planck Institute for Demographic Research.

Ogurtsova, E., Gampe, J., and Zinn, S. (2010). *Practical population forecasting by microsimulation: application of the MicMac software*. Paper presented at the Work session on demographic projections, Lisbon, 28-30 April 2010. http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-10-009.

Palloni, A. (2001). Increment-decrement life tables. In: Preston, S.H., Heuveline, P., and Guillot, M. (eds.). *Demography. Measuring and modelling population processes*. Oxford: Blackwell: 256–272.

Putter, H. (2011). Special issue about competing risks and multistate models. *Journal of Statistical Software* 38(1).

Putter, H. (2014). Tutorial in biostatistics: competing risks and multi-state models. Analysis using the mstate package [electronic resource]. http://cran.r-project.org/web/packages/mstate/vignettes/Tutorial.pdf.

Putter, H., Fiocco, M., and Geskus, R.B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine* 26: 2389–2430. doi:10.1002/sim.2712.

Putter, H. and van Houwelingen, H.C. (2011). Frailties in multi-state models: Are they identifiable? Do we need them? *Statistical Methods in Medical Research*. doi:10.1177/0962280211424665.

Putter, H., de Wreede, L., and Fiocco, M. (2012). *Package 'mstate'*. CRAN repository.

Rajulton, F. (1985). Heterogeneous marital behavior in Belgium, 1970 and 1977: an application of the semi-Markov model to period data. *Mathematical Biosciences* 73: 197–225. doi:10.1016/0025-5564(85)90012-4.

Rajulton, F. (1992). Life history analysis: guidelines for using the program LIFEHIST (PC version). Population Studies Centre Discussion Paper No. 92–95. University of Western Ontario, Population Studies Centre.

Rajulton, F. (2001). Analysis of life histories: a state-space approach. *Canadian Journal of Population* 28(2): 341–359.

Rajulton, F. and Lee, H.Y. (1988). A semi-Markovian approach to using event history data in multi-regional demography. *Mathematical Population Studies* 1(3): 289–315. doi:10.1080/08898488809525279.

Rees, P. (1981). Accounts based models for multiregional population analysis: methods, program and users' manual. Working Paper 295. Leeds, UK: School of Geography, University of Leeds.

Rees, P. and Wilson, A. (1977). *Spatial Population Analysis*. London: Edward Arnold.

Rogers, A. (1995). *Multiregional demography. Principles, methods and extensions*. Chichester: Wiley.

Rowe, G.T. and Moore, K.D. (2009). Simulating employment careers in the LifePaths model: validation across multiple time scales. In: Zaidi, A., Harding, A., and Williamson, P. (eds.). *New frontiers in microsimulation modelling*. Surrey: Ashgate: 333–349.

Scherbov, S., Yashin, A., and Grechucha, V. (1986). Dialog system for modeling multidimensional demographic processes. Working Paper WP-86-29. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Scherbov, S. and Grechucha, V. (1988). DIAL – A system for modelling multidimensional demographic processes. Working Paper WP-88-36. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://www.iiasa.ac.at/publication/more_WP-88-036.php.

Schoen, R. (1988). *Modeling multigroup populations*. New York: Plenum Press. doi:10.1007/978-1-4899-2055-3.

Schoumaker, B. (2013). A Stata module for computing fertility rates and TFRs from birth histories: tfr2. *Demographic Research* 28(38): 1093–1144. doi:10.4054/DemRes.2013.28.38.

Singer, B. and Spilerman, S. (1976). The representation of social processes by Markov models. *American Journal of Sociology* 82: 1–54. doi:10.1086/226269.

Spielauer, M. (2006). The "Life Course" model, a competing risk cohort microsimulation model: source code and basic concepts of the generic microsimulation programming language Modgen. MPIDR Working Paper WP 2006-046. Rostock: Max Planck Institute for Demographic Research. http://www.demogr.mpg.de/papers/working/wp-2006-046.pdf.

Spielauer, M. (2009). General characteristics of ModGen applications: exploring the model Riskpaths. Statistics Canada, Modeling Division. http://www.statcan.gc.ca/microsimulation/pdf/chap3-eng.pdf.

Spitoni, C., Verduijn, M., and Putter, H. (2010). Estimation and asymptotic theory for transition probabilities in Markov renewal multi-state models. *International Journal of Biostatistics* 8(1): Article 23.

Statistics, Canada (2009). Modgen 10.1.0 Developer's Guide [electronic resource]. http://www.statcan.gc.ca/microsimulation/modgen/doc/devguide/dev-guide-eng.htm.

Steinberg, D. (1991). Competing risk multi-spell survival analysis: the CTM procedure. SAS Institute Inc. Proceedings of the Sixteenth Annual SAS Users Group International Conference, Sugi-91-267, SAS Institute, Cary, NC. http://www.sascommunity.org/sugi/SUGI91/Sugi-91-267%20Steinberg.pdf.

Therneau, T.M. (1999). A package for survival analysis in S [electronic resource]. http://www.mayo.edu/research/documents/tr53pdf/DOC-10027379.

Therneau, T.M. and Crowson, C. (2014). Using time dependent covariates and time dependent coefficients in the Cox model [electronic resource]. http://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf.

Therneau, T.M. and Grambsch, P.M. (2000). *Modeling survival data: extending the Cox model*. New York: Springer. doi:10.1007/978-1-4757-3294-8.

Tiemeyer, P. and Ulmer, G. (1991). MSLT. A program for the computation of multistate life tables. Working Paper 91–134. Center for Demography and Ecology, University of Wisconsin.

Touraine, C., Joly, P., and Gerds, T.A. (2013). *SmoothHazard: Fitting illness-death model for interval-censored data. R package version 1.0.9*. CRAN Repository.

Townend, P., Xu, J., Birkin, N., Turner, A., and Wu, B. (2009). MoSeS: Modelling and simulation for e-social science. *Philosophical Transactions of the Royal Society A (Mathematical, Physical and Engineering Sciences)* 367. doi:10.1098/rsta.2009.0041.

Tuma, N.B. (1979). *Invoking RATE*. Menlo Park, CA: SRI International.

Tuma, N.B. and Hannan, M.T. (1984). *Social dynamics: models and methods*. Orlando, FL: Academic Press.

Van den Hout, A. (2013). ELECT: estimation of life expectancies using continuous-time multi-state survival models [electronic article]. http://www.ucl.ac.uk/~ucakadl/ELECT_Manual_13_02_2013.pdf.

Van den Hout A., Jagger, C., and Matthews, F.E. (2009). Estimating life expectancy in health and ill health using a hidden Markov model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 58: 449–465. doi:10.1111/j.1467-9876.2008.00659.x.

Van den Hout A., Ogurtsova, E., Gampe, J., and Matthews, F.E. (2014). Investigating healthy life expectancy using a multi-state model in the presence of missing data and misclassification. *Demographic Research* 30(42): 1219–1244. doi:10.4054/DemRes.2014.30.42.

Van Houwelingen, H.C. and Putter, H. (2008). Dynamic prediction by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Analysis* 14: 447–463. doi:10.1007/s10985-008-9099-8.

Van Houwelingen, H.C. and Putter, H. (2011). *Dynamic prediction in clinical survival analysis*. Boca Raton: Chapman and Hall.

Van Imhoff, E. (1990). The exponential multidimensional demographic projection model. *Mathematical Population Studies* 2(3): 171–182. doi:10.1080/08898489009525305.

Van Imhoff, E. (2005). LIPRO 4.0. Tutorial. The Hague: NIDI. http://www.nidi.knaw.nl/shared/content/output/lipro/LIPRO%204.0%20Tutorial.pdf.

Van Imhoff, E. and Keilman, N. (1992). LIPRO 2.0: an application of a dynamic demographic projection model to household structure in the Netherlands. NIDI/CBGS Publications nr. 23. Amsterdam/Lisse: Swets & Zeitlinger. http://www.nidi.knaw.nl/en/research/al/270101.

Van Imhoff, E. and Post, W. (1998). Microsimulation methods for population projection. *Population. An English Selection* 10: 97–138.

Vos, T. et al. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380(9859): 2163–2196. doi:10.1016/S0140-6736 (12)61729-2.

Wangler, M. and Beyersmann, J. (2012). *changeLos: change in LOS*. CRAN Repository.

Weden, M. (2005). LXPCT_2: Stata module to calculate multistate life expectancies [electronic resource]. http://ideas.repec.org/c/boc/bocode/s453001.html (Accessed 21 March 2014).

Willekens, F. (1979). Computer program for increment-decrement (multistate) life table analysis: a user's manual to LIFEINDEC. Working Paper WP-79–102. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://depot.knaw.nl/2495/.

Willekens, F.J. (1995*). MUDEA. Version 2.0. Manual and tutorial*. Faculty of Spatial Sciences, University of Groningen.

Willekens, F.J. (2005). Biographic forecasting: bridging the micro-macro gap in population forecasting. *New Zealand Population Review* 31(1): 77–124.

Willekens, F.J. (2007). Multistate model for biographic analysis and projection. Working Paper 2007/1. The Hague: NIDI. http://www.nidi.nl/shared/content/ output/papers/nidi-wp-2007-01.pdf.

Willekens, F.J. (2009). Continuous-time microsimulation in longitudinal analysis. In: Zaidi, A., Harding, A., and Williamson, P. (eds.). *New frontiers in microsimulation modelling*. Surrey: Ashgate: 413–436.

Willekens, F. (2011a). Population accounts. In: Stillwell, J. and Clarke, M. (eds.). *Population dynamics and projection methods. Understanding Population Trends and Processes.* Dordrecht, Netherlands: Springer: 29–40.

Willekens, F. (2011b). Microsimulation in population projection. *Cahiers Québécois de Démographie* 40(2): 267–297 (in French). http://www.erudit.org/revue/cqd/2011/v40/n2/1011542ar.pdf.

Willekens, F. (2012). *Biograph: explore life histories*. CRAN repository.

Willekens, F. (2014). *Multistate analysis of life histories with R*. New York: Springer (in print).

Willekens, F.J. and Drewe, P. (1984). A Multiregional model for regional demographic projection. In: Ter Heide, H. and F. Willekens, F. (eds.). *Demographic Research and spatial policy*. London: Academic Press: 309–334.

Willekens, F. and Rogers, A. (1978). Spatial population analysis: methods and computer programs. Research Report RR-78-18. Laxenburg, Austria: International Institute for Applied Systems Analysis. http://www.iiasa.ac.at/publication/more_RR-78-018.php.

Wilson, T. (2001). A new subnational population projection model for the United Kingdom. [PhD Thesis]. University of Leeds.

Wilson, T. (2011). Modelling with NEWDSS: producing state, regional and local area population projections for New South Wales. In: Stillwell, J. and Clarke, M. (eds.). *Population dynamics and projection methods. Understanding population trends and processes*. Dordrecht: Springer: 61–97. doi:10.1007/978-90-481-8930-4_4.

Wilson, T. and Rees, Ph. (2003). Why Scotland needs more than just a new migration policy. *Scottisch Geographical Journal* 119(3): 191–208.

Wohland, P., Rees, Ph., Norman, P., Boden, P., and Jasinska, M. (2010). Ethnic population projections for the UK and local areas, 2001-2051. Working Paper 10/2. School of Geography, University of Leeds. http://www.geog.leeds.ac.uk/fileadmin/documents/research/csap/10-02.pdf.

Wohland, P., Rees, Ph., Norman, P., and Boden, P. (2012). Ethnic population projections for the UK, 2001-2051. *Journal of Population Research* 29(1): 45–89. doi:10.1007/s12546-011-9076-z.

Wolf, D.A. (1986). Simulation methods for analyzing continuous-time event history models. *Sociological Methodology* 16: 283–308. doi:10.2307/270926.

Wolf, D.A. (1988). The multi-state life table with duration dependence. *Mathematical Population Studies* 1(3): 217–245. doi:10.1080/08898488809525276.

Wolf, D.A. and Gill, T.M. (2009). Modeling transition rates using panel current-status data: How serious is the bias? *Demography* 46(2): 371–386. doi:10.1353/dem.0.0057.

Wolfson, M. and Rowe, G. (2014). HeathPaths: using functional health trajectories to quantify the relative importance of selected health determinants. *Demographic Research* (this volume).

Wu, B. and Birkin, M. (2013a). Moses: planning for the next generation. *International Review for Spatial Planning and Sustainable Development* 1(1): 19–30. doi:10.14246/irspsd.1.1_17.

Wu,B. and M. Birkin (2013b). Moses: a dynamic spatial microsimulation model for demographic planning. In: Tanton, R. and Edwards, K.L. (eds.). *Spatial microsimulation: a reference guide for users*. Springer: 171–194.

Wu, B., Birkin, M., and Rees, Ph. (2011). A dynamic MSM with agent elements for spatial demographic forecasting. *Social Science Computer Review* 29: 145–160. doi:10.1177/0894439310370113.

Yi, K.M., Honore, B., and Walker, J. (1987). *CTM: a user's guide*. Chicago: NORC (National Opinion Research Center).

Zeng, Y. (1991). *Family dynamics in China. A life table analysis*. Madison, Wisconsin: The University of Wisconsin Press.

Zeng, Y., Vaupel, J.W., and Wang, Z. (1997). A Multidimensional model for projecting family households – with an illustrative numerical application. *Mathematical Population Studies* 6(3): 187–216. doi:10.1080/08898489709525432.

Zeng, Y., Vaupel, J.W., and Wang, Z. (1998). Household projection using conventional demographic data. *Population and Development Review* 24(Supplementary Issue: Frontiers of Population Forecasting): 59–87.

Zeng, Y., Land, K., Wang, Z., and Gu, D. (2006). U.S. family household momentum and dynamics: an extension and application of the ProFamy method. *Population Research and Policy Review* 25:1–41. doi:10.1007/s11113-006-7034-9.

Zeng, Y., Land, K., Wang, Z., and Gu, D. (2013). Household and living arrangement projections at the subnational level: an extended cohort-component approach. *Demography* 50(3): 827–852. doi:10.1007/s13524-012-0171-3.

Zeng, Y., Land, K., Wang, Z., and Gu, D. (2014). *Household and living arrangement projections. The extended cohort-component method and applications to the U.S. and China*. New York: Springer.

Zinn, S. (2010). A continuous-time microsimulation and first steps towards a multi-level approach in demography. [PhD dissertation]. Faculty of Information Science and Electrotechnics, University of Rostock. http://rosdok.uni-rostock.de/file/rosdok_derivate_0000004766/Dissertation_Zinn_2011.pdf.

Zinn, S. (2014a). *MicSim: performing continuous-time microsimulation*. CRAN repository.

Zinn, S. (2014b). *Continuous-time microsimulation for population projections: the MicSim package for R*. Submitted.

Zinn, S. and Gampe, J. (2011). The MicMacCore. Version 1.0. Manual. Rostock: Max Planck Institute for Demographic Research.

Zinn, S., Gampe, J., Himmelspach, J., and Uhrmacher, A.M. (2009). Mic-Core: tool for microsimulation. In: Rossetti, M.D., Hill, R.R., Johansson, B., Dunkin, A., and Ingalls, R.G. (eds.). *Proceedings of the 2009 Winter Simulation Conference*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.302.6913&rep=rep1&type=pdf.

Zinn, S., Gampe, J., Himmelspach, J., and Uhrmacher, A.M. (2010). A DEVS model for demographic microsimulation. *Proceedings of the Spring Simulation Multiconference 2010*: Article 146. doi:10.1145/1878537.1878689.

Zinn, S., Gampe, J., Himmelspach, J., and Uhrmacher, A.M. (2013). Building Mic-Core, a specialized M&S software to simulate multi-state demographic micro models, based on JAMES II, a general M&S framework. *Journal of Artificial Societies and Social Simulation* 16(3): 5.