# *Multistate Analysis of Life Histories with R*

by Frans Willekens
Dordrecht, Heidelberg, London, New York: Springer Press 2014
ISBN: 978-3-319-08382-7
Softcover, $59.99, 308 pp.

Reviewed by David A. Swanson
Department of Sociology, University of California Riverside

This inexpensive book is part of the Springer series "Use R!" that is under the general editorship of Robert Gentleman, Kurt Hornik, and Giovanni Parmigiani. There are currently ten books in this series, each of which discusses an application of the free statistical software package R (www.r-project.org). The series is intended to have books focus on either a particular subject area, such as demography, or a particular statistical topic, such as longitudinal data analysis. In the case of *Multistate Analysis of Life Histories with R*, we get both in that the subject area of demography is combined with the statistical topic of longitudinal data analysis. As such, the book is largely built around a description of population change as a set of life histories, viewed as the realization of stochastic processes. Consequently, it emphasizes the "biograph" module of R. However, it also discusses in some depth six other modules in R that support life history analysis, and links them to statistical models and demographic concepts.

The book consists of 9 chapters, a preface, 3 appendices (called "annexes" in the book), a name index, a subject index, 81 tables, 25 sidebars (called "boxes" in the book), and 45 figures, many of which are in full or partial colour. The first chapter consists of a six-page introduction that provides a concise overview of the book along with a useful context. The second chapter discusses real and synthetic life histories, while Chapter 3 offers a description of R and its biograph module. Chapters 4 and 5 continue the general introduction by discussing exploratory data analysis and the visualization of life histories, respectively. Chapter 6 covers six modules in R that support life history analysis: (1) survival; (2) eha; (3) mvna; (4) etm; (5) mstate; and (6) msm. It includes discussions of the functions within each module and provides examples of their use. Chapter 7 discusses the multistate life table, and Chapter 8 provides example applications of the life history approach, using R with data from the Netherlands Family and Fertility Survey. In the last chapter (9), Willekens gives a retrospective summary of the book that is concise but comprehensive.

Annex (Appendix) A shows how to use the biograph module. Annex B is a list of biograph functions and data, and Annex C consists of a list of biograph functions and the R functions on which they are based.

As a demographer who was trained in and still largely uses data aggregated by others (e.g., Statistics Canada, the US Census Bureau), it has taken me a while to start thinking about micro-level data. My initial introduction was via population forecasting done through micro-simulation, something pioneered at Statistics Canada (www.statcan.gc.ca/microsimulation/demosim/demosim-eng.htm). For me, this book by Willekens was a real-eye-opener in extending my understanding of using micro-level data and the tools for doing so. I greatly appreciated not only his clear descriptions and the many examples used to illustrate theoretical and technical concepts, but his skillful combinations of demographic and statistical approaches.

For those not familiar with R, the book requires learning a new software application. Given that R is freely available, widely used, and well-supported, my advice is that it is worth learning. In addition, Willekens and Hein Putter (2014) have an article in the journal *Demographic Research* that reviews software packages designed

for life history analysis. In it, one can read comparisons that include R. As clearly illustrated in many of the illustrations, the colour graphics generated by R provide excellent visualizations of concepts and data. Willekens states in the Preface that the book is aimed at describing, explaining, and predicting life histories. He makes it a point that his use of the word *prediction* is not aimed at forecasting. Rather, he uses "prediction" in the same manner that I would use the word "estimation." One could, however, use the concepts and tools that Willekens describes to generate forecasts, not just estimates. As such, I find that book provides a very nice addition to the micro-simulation approach to generating population forecasts, one that could handle a wide range of ascribed and achieved characteristics. In this regard, it would be an interesting exercise to consider achieved variables in regard to taking a simulation approach to a stable population.

This book is a very nice addition to the many other important methodological contributions made by Frans Willekens and I recommend that it be read by all demographers, regardless of orientation: social, formal, academic, or applied.

## Reference

Willekens, F., and H. Putter. 2014. Software for multistate analysis. *Demographic Research* 31(14):381–420.