Discrete-time multistate modeling for life course analysis

Christian Dudel[1,2], Daniel Schneider[1], Angelo Lorenti[1] & Mikko Myrskylä[1,3]

[1] *Max Planck Institute for Demographic Research, Rostock, Germany*
[2] *Federal Institute for Population Research, Wiesbaden, Germany*
[3] *University of Helsinki, Finland*

Keywords: Multistate modeling, life course analysis, discrete-time modeling, panel data, Markov assumption

Many social processes can be represented by individuals being in, and transitioning between, a finite number of states. Examples include transitions in the labor market, between marital states, and many more. Statistical modelling of such processes can be done using multistate models. Multistate models come in two different variants: discrete-time and continuous-time. While continuous-time multistate models have been extensively described in the literature, there is little guidance on their discrete-time counterparts. This is unfortunate since many data sources – regularly spaced longitudinal surveys in particular – naturally lend themselves to modelling in discrete time.

We provide new theoretical and practical insights into discrete-time multistate models (DTMMs) for panel data. From a theoretical perspective, we establish a set of theoretical estimands which can be studied using DTMMs and which are linked to life course theory and the concept of cumulative (dis)advantage. Moreover, we present novel technical results which counter one of the main criticisms of multistate modeling: multistate models usually rely on the Markov assumption, which implies that the studied process is memoryless. While this assumption likely is false for many potential applications, we show that several quantities can be consistently estimated using DTMMs even if the Markov assumption does not hold. Finally, related to the previous two points, we argue that DTMMs can provide adequate representations of population-level quantities, while being less well-suited for individual-level predictions.

From a practical perspective, we use data from SHARELIFE to show that DTMMs provide good representations of real-word data, and that finite sample bias is modest. An in-depth example is provided using data from the U.S. Health and Retirement Study. Finally, we briefly remark on a novel Stata package and existing R packages for the estimation of DTMMs.