# Longevity forecasting by socio–economic groups using compositional data analysis

**Søren Kjærgaard**[1,2], **Jim Oeppen**[1], **Marie–Pier Bergeron Boucher**[1] and **Rune Lindahl-Jacobsen**[1]

[1]Max–Planck Odense Center on the Biodemography of Aging, Odense, Denmark

[2]*skjaergaard@health.sdu.dk*

**Abstract**

Longevity differentials between sub–populations, as well as other classifications of mortality, present a significant challenge when planning both public and private pension systems. Examples include differentials between socio–economic groups, or mortality components such as causes of death. For socio–economic groups, modelling of longevity has often been done separately for each group, rarely coherently, and never using compositional techniques, even though the density of deaths across socio–economic groups are compositional data. Life table deaths only contain relative information, distributing a constant sum by age. This paper forecasts longevity in sub–populations using CoDa to account for the dependence between them and compares three alternative specifications of the mortality process.

Assume life table deaths in $n$ sub–populations are classified by year ($t$) and age ($x$). The first model follows Oeppen (2008) and stacks the sub–groups horizontally, with a unit sum for all $t$ rows, before applying a log–ratio transformation to the column–mean CoDa–centred $t \times (x \cdot n)$ matrix. The rank–1 Singular Value Decomposition (SVD) of this matrix estimates a fixed vector of $x \cdot n$ effects varying by age and sub–group, and a single temporal vector of length $t$ that can be forecast to estimate future rows of matrix. The second model stacks the sub–groups vertically, with unit sums for all $n \cdot t$ rows. After centreing and transformation, the rank–1 SVD estimates a single fixed age pattern and a vector containing $n$ group–specific time profiles, each of which has to be forecast. The third option is not to stack but introduce a third dimension along the $n$ sub–populations. In this three dimensional model each group's $t \times x$ death density slice is row–constrained summing to one and a new parameter vector is introduced controlling the sub–population dependency. Then the rank–1 SVD estimates one fixed vector for age effects, one vector for time that can be forecast, and one fixed population parameter vector (Bergeron Boucher *et al.*, 2016).

Each model offers an alternative approximation of the roles of time, age, and sub–group in mortality change. This paper compares their ability to estimate, and forecast, longevity across socio–economic groups. All three models will forecast the sub–populations in a coherent way and by using CoDa account for the constraints on life table deaths within the same year and across sub–populations.

## References

Oeppen, J. (2008). Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data. *In European Population Conference 2008, European Association for Population Studies*

Bergeron-Boucher, M., Simonacci V., Oeppen J., and Gallo M. (2016). Coherent modeling and forecasting of mortality patterns for subpopulations using multi-way analysis of compositions: An application to Canadian provinces *In Population Association of America 2016*