The resurrection of time as a continuous concept in biostatistics, demography and epidemiology

Bendix Carstensen
Senior Statistician, Steno Diabetes Center Copenhagen, Gentofte, Denmark
b@bxc.dk   http://BendixCarstensen.com

In classical demography the actuarial estimator of the survival function has been used for centuries, typically with one year time intervals. This estimator was also the default in medical science and epidemiology till the advent of computers, when the Kaplan-Meier estimator and later the Cox-model became the *de facto* standard for analysis of survival data and more generally, for follow-up data from cohort studies. The reporting of survival curves in the medical literature is therefore almost exclusively as the well-known noisy step curves, which is then left to eye-ball smoothing by the reader.

In epidemiology there has also been a tradition in analysis of occurrence rates by tabulation of events and risk time in five- or ten-year intervals, and fitting models with a separate parameter in each interval. In some circumstances such as Age-Period-Cohort models based on data classified in Lexis triangles this approach gives distinctly illogical results [1].

Both of these approaches are essentially based on models where the effect of time is taken to be different in different intervals, and the intervals taken as being exchangeable — the inherent ordering of time-intervals are ignored in the model. For the very broad time-intervals the usual problems associated with categorization applies [2]. Only in the reporting of the effects is the ordering of time-intervals reintroduced, either by connecting the estimates from broad time-categories, or hiding the ragged nature of the estimates by only showing cumulative effects.

I will advocate the use of models using the quantitative nature of time in modeling of occurrence rates, by imposing restrictions on the time-effects reflecting this. This can be seen as a combination of the very fine subdivision of time used in the non-parametric modeling and assumptions about continuous, smooth effects of time. Besides the ability to show the time-effects directly on the rate-scale, this approach also includes the possibility to accommodate multiple time scales such as age and duration of disease simultaneously. It also readily extends to estimation of transition rates in multistate models [3]. Parametric estimates of transition intensities also allow derivation of demographic quantities such as residual life time and years of life lost.

An essential prerequisite for practical handling of these models is a proper way of data representation. I shall describe the philosophy of the `Lexis` machinery for representation of multistate data on multiple time scales implemented in the `Epi` package for **R**[4, 5]. Furthermore I will show how otherwise intractable quantities from large complex multistate models can be handled by simulation [6], even if models also involve time since entry to intermediate states as timescale. Examples will be drawn from the clinical literature, and I shall indicate directions for further work on devising proper measures of uncertainty of such quantities.

References

[1] B Carstensen. Age-Period-Cohort models for the Lexis diagram. *Statistics in Medicine*, 26(15):3018–3045, July 2007.

[2] Frank Harrell. Problems caused by categorizing continuous variables. `http://biostat.mc.vanderbilt.edu/wiki/Main/CatContinuous`, 2015.

[3] S. Iacobelli and B. Carstensen. Multiple time scales in multi-state models. *Stat Med*, 32(30):5315–5327, Dec 2013.

[4] Martyn Plummer and Bendix Carstensen. Lexis: An R class for epidemiological studies with long-term follow-up. *Journal of Statistical Software*, 38(5):1–12, 1 2011.

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

[6] Bendix Carstensen. Simulation of multistate models with multiple timescales: `simLexis` in the `Epi` package. `http://BendixCarstensen.com/Epi/simLexis.pdf` or `Epi` vignette, 2015.

About the author

Bendix Carstensen is a senior statistician in the department of Clinical Epidemiology at Steno Diabetes Center Copenhagen, and is associated with the department of Biostatistics at the University of Copenhagen and the department of Medical Epidemiology and Biostatistics at Karolinska Institutet in Stockholm.

He holds a degree in mathematical statistics from the University of Copenhagen (1983) and has 30 years of experience in practical biostatistics and epidemiology. He spent 11 years at the Danish Cancer Registry, 4 years at the Danish Zoonosis Center and has currently been 18 years at the Steno Diabetes Center Copenhagen.

He has a keen interest in demographic methods and the application in practical epidemiology (mostly diabetes), and is the maintainer and co-author of the `Epi` package for **R**. He is also the main author and maintainer of the `MethComp` package for analysis of method comparison studies. He has over 25 years of international teaching experience in practical statistics and epidemiology - of which a lot is available at his website, `http://BendixCarstensen.com`.