Statistical Modelling 2019; **19(3)**: 223–224

Guest Editorial: Statistical Modelling in Demographics

Statistics and demography are cognate disciplines. Not only do they share historical roots of systematically collecting data on people and their economic activities for the state, ultimately leading to the term 'statistics', but also important methodological achievements are regarded as cornerstones in both domains. The concept of a life table, as published by John Graunt in the *Natural and Political Observations upon the Bills of Mortality* in 1662, undoubtedly is the most important example. Not surprisingly, a considerable number of scholars have contributed to both disciplines with their research since then. With the development of 'modern statistics' during the 20th century the historically close ties loosened but never broke up. This special issue collects five papers that demonstrate the ongoing cross-fertilization of the two disciplines.

Demography studies the size and composition of populations (often, but not necessarily of humans) and how they change over time. This, more specifically, requires the analysis of the core processes—fertility, mortality and migration—that jointly determine population dynamics. But also other life course developments, such as educational careers, health trajectories or labour market careers, which influence population dynamics indirectly, are considered in demographic studies.

The papers in this issue take specific questions and demonstrate how careful statistical modelling can contribute to a better understanding of demographic issues.

The contribution by Schmertmann and Hauer is a paramount example of how demographic and statistical modelling can successfully work hand in hand. The total fertility rate (TFR) is one of the key measures in the analysis of fertility. It quantifies the lifetime average number of children per woman should the current age-specific fertility rates persist. To obtain the TFR, age-specific fertility rates (number of children born to women in the fertile ages, usually between age 15 and 50) are needed, which requires accurate registration of vital events. In case such information is not available or inadequate, so called indirect methods have a long tradition in demography. In particular, formal relationships allow to approximate the TFR from information about the population's age–sex structure. Schmertmann and Hauer embed the formal relationship in a Bayesian model, use demographic priors and estimate the TFR, including its uncertainty distribution, in this setting.

Bremhorst, Kreyenfeld and Lambert deal with a different question in fertility analysis. A better understanding of the factors that influence the quantum (the number of children) and the timing of childbearing (when do couples have children) requires statistical models that can address both aspects of fertility. In the medical statistical literature such models are called cure (or immune) survival models, since they allow to separately describe the probability to not experience an event (here: childbirth) and the hazard of the event in the 'susceptible' group. Covariate effects often are nonlinear, however, and flexible regression models can provide insights in the nature of some effects. In their paper, Bremhorst, Kreyenfeld and Lambert propose a novel approach to a so-called promotion time model that includes a flexible effect of continuous covariates as well as a nonparametric specification of the baseline hazard function for the susceptible group.

 10.1177/1471082X19852227

Estimation is performed via Bayesian P-splines and the model is employed to study fertility progression in Germany and the effect of age at first parenthood, which is an important factor of family size when family formation and childbearing tends to be postponed.

Mortality, that is the age-specific risk of death, varies with age but also over calendar time. So-called mortality surfaces collect death rates over an age–time grid so that both components can be studied concurrently. For humans at adult ages the increase in mortality is described remarkably well by an exponential hazard, the so-called Gompertz model. The parameters of the model, however, vary with time, country or sex. Lagona and Barbi in their contribution suggest a hidden Markov model to capture the time-varying parameters in mortality surfaces. The derived estimates allow a segmentation of the surface that sheds further light on the dynamics of mortality change.

Understanding mortality also requires understanding of the causes of death and their changing importance over time. Medical progress and public health efforts, together with a changing age composition toward the elderly, constantly modify the prevalence of the different causes of death. Internationally accepted categorizations, such as the International Classification of Diseases and Related Health Problems (ICD) by the World Health Organization (WHO), help to make meaningful comparisons, however, such classifications regularly undergo revisions. As a consequence, time series of mortality by cause exhibit disruptions that hamper the study of time trends. Camarda provides an approach to construct coherent series of mortality by cause of death. He circumvents the problem that an exact double-classification of causes in two coding regimes mostly is not available by estimating the transition coefficients which quantify the proportions of deaths in one scheme that would have been classified in the different categories of another scheme. Requiring smoothly joining time-series in the years of the revision the coefficients can be estimated by penalized likelihood and, after re-classification, consistent time-series by cause are available for trend analysis.

Between birth and death the life course unfolds, and an important aspect that has repercussions in many other domains of life is work. Van den Hout and Tan offer a multi-state model to flexibly describe employment histories. The transition rates are modelled by the generalized gamma distribution, and a latent class specification allows different (latent) subgroups to proceed with different intensities through their employment career. The model is applied to data from the German Life History Study.

This special issue originated from some casual discussions at the 31st International Workshop on Statistical Modelling in Rennes, and we are indebted to the contributors to this issue who were so positive about this endeavour. We also would like to express our thanks to the reviewers of the papers who helped us to assure quality. And finally our thanks go to Brian Marx, our always-available partner of the journal's editorial board.

Jutta Gampe
Carlo Giovanni Camarda

Guest Editors