Statistical modelling of time-varying covariates for survival data
Spreafico, M.

Summary

In clinical research, *time-varying covariates* are of great interest since they represent the way dynamic patterns evolve, affecting patient's health status and disease progression. In literature such data are usually considered using a piecewise-constant or fixed-baseline approach, losing the potential information content they may provide if the association between time-varying and *time-to-event* data is properly captured. Due to the complexity of the phenomena, an advanced analytical framework is required to adequately model disease evolution and characterise its relationship to the dynamic nature of time-dependent features. These aspects are rarely addressed and may provide new insights for medical research, representing a challenge of both clinical relevance and statistical interest.

This dissertation focuses on developing mathematical and statistical methods to properly represent time-varying processes and model them within the context of time-to-event analysis by means of appropriate Cox-type survival models. The main purpose is to enrich the knowledge available for modelling survival with relevant features related to the time-varying covariates of interest. The efforts of this work address the complexities of both (i) developing adequate dynamic characterizations of the processes under study (i.e., the *representation* problem) and (ii) identifying and quantifying the association between time-varying processes and patient survival (i.e., the time-to-event *modelling* problem). In both cases, the main issue is dealing with complex data sources while taking into account the nature of the processes and their particular aspects (such as temporal dynamics, categorical levels or recurring events) and managing the complex trade-off between clinical interpretability and mathematical formulation.

Depending on the context of the study, different approaches are proposed. In terms of *representations*, on one hand complex data integration and functional data analysis are exploited to propose new longitudinal or functional features capable of incorporating trends and variations in the evolution of processes over time. On the other, novel statistical methodologies are introduced to extract additional information in terms of longitudinal or functional data from different data sources. Stochastic process theory, latent Markov models and compositional data analysis are exploited, among others, to address the research questions. In terms of time-to-event *modelling*, advanced versions of Cox-type regression are proposed to include the dynamic covariates in the survival analysis. Subject to the type of study and data, joint models, functional survival models exploiting dimensionality reduction techniques or marginal structural models are employed.

By solving the aforementioned statistical complexities, this work is not only impacting the community of researchers in mathematics and statistics, but it aims at providing useful tools to support doctors and clinicians in their work. All research topics are motivated by specific clinical questions aimed at gaining insights into personalised treatments for cardiological and oncological patients. Part I of this thesis focuses on the study of the processes of drug consumption, adherence to medications and re-hospitalizations in heart failure patients. Part II deals with the investigation of the time-varying aspects of chemotherapy treatment, such as dose modifications, biomarkers and accumulation of toxicities, in patients affected by osteosarcoma.

In conclusion, the development of novel tailored methodologies capable of enhancing time-to-event modelling with time-varying characteristics may represent a significant step forward in the definition of new customized and flexible monitoring tools, which could then be applied to the study of different pathologies characterised by complex data sources.