Statistical modelling of time-varying covariates for survival data

Spreafico, M.

Conclusions

In this doctoral dissertation, various mathematical and statistical methods were introduced to represent time-varying processes from complex raw data, and model them within the context of time-to-event analysis using appropriate Cox-type survival models. All research topics were motivated by specific clinical questions related to gaining knowledge about personalised treatments for cardiological and oncological patients. The novelties of this work concern both the statistical and the clinical community. The main strengths lie in the methodological contributions of the individual chapters, as well as in the use of the developed time-varying approaches in a medical treatment context, which is not yet standard practice in the literature. The results obtained can thus be contextualised in both statistical and medical contexts.

Part I focused on the processes of drug consumption, re-hospitalization events and their effect on long-term survival in Heart Failure (HF) patients. First, we tackled the issue of adherence to polypharmacy, a direct consequence of the multi-morbidity condition that characterises HF (Chapter 1). On one hand, we described how evidence-based therapies were applied in a real world setting. On the other, we proposed a novel method for measuring adherence to polypharmacy and we investigated its effect on patients survival. Results from administrative data of *Friuli Venezia Giulia* region (Italy) showed that good patients' adherence to polypharmacy was associated with lower death rate. However, the target dose guidelines were not achieved and HF patients' adherence remained unsatisfactory. Although the Patient Adherence Indicator was developed as time-fixed covariate, this study represented a first step forward in the pharmacoepidemiology context for HF patients as few data on polypharmacy adherence exist in a real-world setting.
Since the time-fixed way adherence is usually computed in pharmacotherapy practice discards valuable information, we then proposed (Chapter 2) two novel time-varying covariates able to reflect the dynamics and the behaviour of drug consumption during therapy, i.e., the continuous cumulative months covered by drug assumption up to time t and the dichotomous adherence to the therapy at time t. To capture the interaction among time-varying and survival processes over time, (generalized) mixed effect models for longitudinal data were jointly modelled with Cox-type regression model for time-to-death. Results from administrative data of *Lombardia* region (Italy) showed that having a good adherence trend during time had a protective effect on patients' survival. With a dynamic study of adherence, it was possible to understand in real-time its effects on patient's health status directly monitoring the treatment. This ongoing analysis and quantification of drug consumption could help clinicians to better target therapies for their patients. Modelling the drug intake process as time-dependent covariates in a joint modelling framework is therefore an effective approach for drug utilization studies. This shows the importance of

developing methods to pharmacotherapy using a time-varying perspective.
For that reason, in Chapter 3 we developed a novel methodology based on marked point process theory and Functional Data Analysis (FDA) to reconstruct the compensators of suitable marked point processes for recurrent events intended as functional data. Functional Principal Component Analysis (FPCA) was then used to include the functional compensators into a Multivariate Linear Cox Regression Model (MFLCRM) for long-term survival. From the study on the administrative database of *Lombardia* region (Italy), we observed that the marked point process formulation was a natural way to represent the occurrence of re-hospitalisations or drug purchases over time. The use of FPCA made it possible to extract additional information contained in the functions, representing a powerful exploratory and modelling technique for highlighting trends and variations in the shape of processes over time. From a clinical point of view, this allowed us to model self-exciting behaviour for which the occurrence of events in the past increases the probability of a new event, including a large portion of information from administrative data to describe the patient's medical history. Furthermore, the proposed approach was able to include the information that HF patients usually consume different types of drugs at the same time, representing a novelty for clinical research in the direction of properly treating multimorbidity patients and polypharmacy.
From a pharmacoepidemiology perspective, it should be emphasised that using administrative databases has both strengths and limitations. On one hand, the analyses can be easily reproduced in all Italian regions where data on the drugs purchase are collected, thanks to the flexibility of the methods developed. On the other, the impossibility of ascertaining whether the patient was currently consuming the dispensed drug remains the major data-driven limitation of using drug purchases as a proxy for drug intake. In future research, it might be interesting to link, when possible, administrative data with information on prescribed daily doses in order to obtain a more realistic analysis of coverage periods and achieve more precise results. Nevertheless, we have developed a general analytical framework for processing and modelling information from administrative data sources in a fully innovative way for both health policy and research in clinical epidemiology. This is a great added value of our work.

Part II focused on chemotherapy treatment in osteosarcoma patients, considering the processes of dose modifications, biomarkers and toxicities evolution over time. First, we proposed a Functional covariate Cox Model (FunCM) combining FDA techniques to represent time-varying processes in terms of functional data and FPCA to include them into a MFLCRM (Chapter 4). This method was applied to data from the BO06 randomised controlled trial to study the effect of time-varying biomarkers and chemotherapy dose on overall survival. FunCM revealed differences between patients with different biomarkers and treatment evolution, even when randomized to the same regimen. The results based on this new method provided more information to the clinical community than the standard standard Intention-To-Treat (ITT) approach. Despite the introduction of relevant dynamic features related to the continuous nature of the processes, dose-intense profiles were not associated with survival. This suggested that considering only the achieved dose as a proxy for treatment was not sufficient. Several other aspects, such as latent accumulation of toxicity, needed to be taken into account.

For this reason, we proceeded focusing on the methodological aspects concerning a proper representation of the overall toxicity burden over time, still lacking in the medical literature. We first introduced a novel cycle-dependent longitudinal mean-max method for quantifying Multiple Overall Toxicity (MOTox) during therapy (Chapter 5). The developed MOTox score simultaneously included information on worst grade events, multiple lower grade chronic toxicities and the time dimension related to chemotherapy cycles. This is a flexible method to investigate the individual progression of overall toxicity in cancer patients compared to traditional indexes. The evolution of high MOTox over cycles was then analysed through the implementation of cycle-specific multivariable logistic regression models. Results for BO06 data showed that the highest impact on the risk of the re-occurrence of high-MOTox over cycles was observed for the last available toxic condition. This indicated that longitudinal methods should be considered in the analyses of cancer trials. For this reason, we then proposed (Chapter 6) a new taxonomy based on Latent Markov (LM) models and compositional data to model the evolution of a latent condition of interest (i.e., the Latent Overall Toxicity, LOTox) on the basis of interval-based categorical observations (i.e., the nominal toxicity grades registered over cycles according to the Common Terminology Criteria for Adverse Events, CTCAE). By assuming the existence of a LM chain for the LOTox condition of a patient, the proposed taxonomy identified sub-populations of patients characterized by a similar overall toxicity burden. Individual LOTox dynamics during treatment was then obtained by modelling the personalized longitudinal profiles representing the probability over time of being in a specific LOTox state or the relative risk with respect to a reference "good" toxic condition. Provided that toxicities are recorded according to the CTCAE scale or an analogous grading system, the developed approaches represent flexible methods to quantify the personal evolution of overall toxic risk during chemotherapy. In cooperation with medical staff, they might provide insights for the definition of new guidelines to reduce the impact of chemotherapy treatment in terms of toxic side effects, possibly improving patients quality of life.

In Chapter 7, we introduced Cox-type Marginal Structural Models (Cox MSMs) in combination with Inverse-Probability-of-Treatment Weighted (IPTW) estimators to model the causal effects of joint-exposure on survival outcome in presence of time-varying confounders. Suitable procedures were designed to mimic a randomized trial where joint-exposure – given by chemotherapy Received Dose Intensity (RDI) and histological response – was no longer confounded by toxicities or other confounders. Results from the randomised controlled trials BO03 and BO06 showed that Good Responders (GRs) had better survival than Poor Responders (PRs), but increased reductions in RDI created two opposite trends in the two groups. In particular, in PRs – for whom chemotherapy is less effective – the greater the reduction, the better the survival, meaning that an increase in RDI may be detrimental to survival due to the impact on the immune system. By focusing on a way of analysing chemotherapy data based on RDI rather than ITT, this study illustrated the key role played by toxicities during treatment and showed the detrimental effect of neglecting them.

The added value of Part II was the presentation of comprehensive analyses of complex chemotherapy data, with tutorial-like explanations of the difficulties encountered and the

problem-solving strategies employed. This required the fusion of statistical and medical expertise, showing that a close collaboration between statisticians and clinicians is fundamental. The contribution is thus not limited to statistical methodologies, but concerns the use of the developed methods to assess the potential of dynamic covariates in the context of clinical studies, where a time-fixed approach is usually preferred. The adoption of more refined approaches for managing chemotherapy data is therefore of great value and provides insights both for general guidelines and personalised chemotherapy treatments.

Motivated by these multiple and challenging medical problems, we have developed novel methodologies capable of extracting additional information from composite raw data to enrich advanced or traditional survival models for the clinical endpoints of interest. Weaknesses and limitations of the present work have been discussed on a case-by-case basis in the final remarks of each chapter, along with possible extensions and further developments. Overall, this work reflects the desire to create a comprehensive framework of methods whose intent is to go beyond the state of the art for clinical studies. By implementing a more valuable setting for dealing with complex data sources, this work has led to new insights that would not have been gained by traditional methods. The great added value of this thesis is therefore to have demonstrated the importance of going beyond current practice towards a more sophisticated and informative analytical framework. Clinicians and healthcare managers can benefit from identifying customised longitudinal and functional representations that reflect variability and differences between patients, as they can improve patient profiling and tailor their therapies. Moreover, thanks to their flexibility, the developed methods are not only suitable for the cardiological and oncological context under consideration. The procedures can in fact be extended and generalised to many different settings, adapting them to the different biological and clinical aspects of the specific application.

As a final remark, the importance of an interdisciplinary collaboration between statisticians and clinicians must be emphasised, as it can lead to great contributions for both fields. On one hand, interesting clinical research questions can provide inspiration for the development of new statistical methodologies. This has been demonstrated several times in this work. On the other, advanced statistical tools can help improve current clinical practice. Despite their potential relevance, elaborate methods are usually underused in the clinical setting due to their mathematical complexity. Close collaboration would therefore ensure that new methods are carefully introduced and explained by statisticians to the clinical community, so that the latter can properly benefit from their advantages. This cross-sectional cooperation can thus make the patient pathway through the healthcare system more efficient and effective, representing a significant step forward in the definition of new personalized monitoring tools.